An Overview of Animal Testing Issues

Table of Contents

New Large-Scale Animal Testing Programs

- **High Production Volume Chemical Testing Program**
- **Endocrine Disruptor Screening Program**
- **Voluntary Children Chemical Evaluation Program**

Concluding Remarks

References

Glossary

Tables

Figures

Introduction

Many animal activists are disappointed with the pace at which animal testing is being *reduced*, *refined*, and especially, *replaced* (the "Three Rs" of the alternatives approach). Considerable progress has been made, however, in the science of alternatives and in the attitudes among toxicologists and corporate decision-makers, as well as in developing government-based mechanisms for the regulatory acceptance of alternative methods.[1] While there are scientific, financial, and regulatory obstacles to replacing animal testing, a comprehensive review of the history of technical and policy developments suggests that there will come a time when animals no longer are used in harmful testing of any kind. The safety testing of chemicals and consumer products probably accounts for only about 10 to 20 percent of the use of animals in laboratories, or approximately 2 to 4 million animals in the United States. The remaining 80 to 90 percent, or 16 -to 18 million animals, are used in basic and applied research, education, and other arenas. Yet the use of animals in safety testing figures prominently in the animal research controversy. It raises issues such as the ethics and humaneness of deliberately poisoning animals, the propriety of harming animals for the sake of marketing a new cosmetic or household product, the applicability of animal data to humans, and the possibility of sparing millions of animals by developing alternatives to a handful of widely used procedures.

History of Safety Testing

The safety testing of chemicals and products is a relatively recent development in history. While prominent people in antiquity may have had food tasters at their side to protect them from being poisoned, and coal miners in the 19th century used canaries to warn them of pockets of dangerous gases, the widespread testing of drugs, chemicals, consumer products, and foods has been going on for less than seventy years, essentially a feature of industrialized, consumer-driven society.

Safety testing grew out of other forms of testing. Initially, tests were developed to standardize new batches of powerful drugs like digitalis and insulin that were prepared from natural products and that varied in potency from batch to batch. These are called *biologicals*, therapeutically active chemicals that are similar or identical to compounds normally found in the body. Testing for potency was soon extended from biologicals to vaccines against infectious diseases. Gradually, the approaches developed for potency testing came to be applied to the safety assessment of chemicals and products that had been implicated in cases of human poisoning. For example, in 1937, 107 Americans died as a result of poisoning when a sulfanilamide preparation was mixed with the toxic diethylene glycol as a solvent (Parascandola, 1991). Shortly after the Elixir of Sulfanilamide tragedy, Congress passed the Food, Drug and Cosmetic Act of 1938, requiring the safety testing of drugs.

Potency testing and safety testing are not the only forms of testing. In 1962, following the thalidomide tragedy in which many infants were born with deformed or missing arms and legs, Congress required that, prior to marketing new drugs, companies should test not only for safety but also for efficacy, i.e., that drugs do what they claim to do.

Types of Safety Testing

Safety testing is a multifaceted process. It includes testing chemicals to see if they do or do not have adverse effects; for example, can a chemical cause birth defects, cancer, organ failure, or some other problem? This is known as "hazard identification" or "hazard assessment." There is also the nuanced issue of whether a chemical is likely to be a problem in the real world given practical issues such as how often people are expected to come into contact with it and at what dose. This is known as "risk assessment."
Today, a variety of safety tests is conducted on a wide range of chemicals and products, including drugs, vaccines, cosmetics, household cleaners, pesticides, industrial chemicals, foodstuffs, and packing materials. The most thorough testing is reserved for drugs and products that will be used in or on foodstuffs. For these agents, tests lasting less than a month (*acute*), a month to three months (*subchronic*), and more than three months (*chronic*) are performed to determine general toxicity (e.g., organ damage), eye and skin irritancy, and potential to cause mutations (*mutagenicity*), cancer (*carcinogenicity*), reproductive problems, and fetal malformations (*teratogenicity*). (See 1 for a listing of specific toxicity tests.) The cost of a full-scale battery of tests runs to several million dollars and takes three to four years to complete. Other agents such as cosmetics are not subjected to the same comprehensive battery of tests but would still call for information on, for example, general oral toxicity, eye and skin irritancy, *phototoxicity* (toxicity triggered by exposure to ultraviolet light), and perhaps mutagenicity.

---

[1] In principle, the best way around the problem of biological variation is to develop an understanding of basic toxicity mechanisms and then develop alternative tests that are specifically based on those mechanisms. In practice, developing an understanding of these mechanisms may be years away, even for most of the common toxicity endpoints.

Problems with Animal Tests

The use of animals in safety tests raises a variety of humane and technical issues. Anyone concerned about the welfare of animals should be given pause by the practice of exposing animals to chemicals that can potentially cause eye and skin irritation, developmental abnormalities, cancer, and death. The suffering associated with toxicity testing can be severe, yet pain- or distress-relieving drugs typically are not administered for a variety of reasons including the concern that these drugs might alter the toxicity profile of the chemical being tested. What are some of the problems associated with animal testing?

Validation

Many of the animal tests in use today were developed decades ago, when the science of toxicology was in its infancy. Some animal tests have been evaluated to determine how well they predict human hazard but none has been formally validated (assessed in multiple laboratories to see if they reliably give the correct answers). Some proponents of animal testing argue that these tests have proven themselves over time; critics would like to see the hard data that supports this claim.

Extrapolating from Animals to Humans

Although it is likely that the animal tests in use today do identify the most toxic chemicals, they are by no means perfect predictors of human hazard. The discrepancies between the results obtained in animals and the likely toxicity to humans is typically played down in literature supporting animal testing and played up by animal protectionists.

The problem of extrapolating the results of animal tests to humans is exacerbated by the subjective nature of many animal tests. In the Draize Test, for example, technicians visually judge the degree of irritation to the rabbits' cornea, iris, and conjunctiva and apply a somewhat arbitrary weighting formula to come up with an overall score. Extrapolation to humans is also confounded by substantial variability associated with many animal tests. This variability stems not only from the subjective nature of the scoring, but also from inherent biological variability among animals or, indeed, with the same individual animals over time. There is also a problem in extrapolating from alternative in vitro (test tube) tests to human experience; however, one can use human tissue in in vitro tests and such tests usually show less variability than do animal tests. Traditionally, it has been argued that because most test animals are mammals and humans are mammals, tests on animals provide adequate warning of danger to humans. But when rat and mouse carcinogens are compared, the tests in rats agree with the tests in mice only two thirds of the time. In making a decision on whether to regulate such a chemical as a carcinogen, how should one decide? Are humans more like mice or like rats? Fortunately, for very toxic chemicals, the mouse and rat tests give the same results more than 90 percent of the time, so the animal tests are probably alerting us to the danger of the worst chemicals (Gold, 1998).

Practical Problems

Animal tests not only lack formal validation and generate uncertainties associated with their extrapolation to humans. They also have practical problems. Some take years to complete and/or are very expensive. For example, the standard rodent bioassay for assessing carcinogenicity takes two years to conduct and costs more than a million dollars.

Obstacles to Replacing Animal Tests

Those who seek to replace a particular animal test with an alternative face formidable obstacles, even though the alternative may perform as well as, or better than, the animal test.

Lack of a True "Gold Standard"

New alternative tests should be compared to the pre-existing animal tests that they seek to replace, in order to see which test performs better. Ideally, the alternative and animal tests should be assessed according to an independent standard. Predicting hazards to humans is the aim of most animal testing, so, in principle, human data should be the standard against which the performance of animal and alternative tests are compared. However, good quality human data are often lacking because of appropriate ethical limitations on human testing. Human data from limited clinical testing, as well as from occupational exposure, accidental exposure, and suicides, are often unreliable or of limited precision and availability. Consequently, the animal test itself is typically used as the default standard against which the alternative test is judged. In other words, the lack of a true gold standard means that in vitro tests as judged according to how well they accord with animal data, not with human data. This "stacks the deck" in favor of the animal tests.

Biological Variability

Those seeking to compare the performance of alternative and animal tests also have to overcome the problem of biological variability. Both types of tests, being based on biological systems, are likely to have inherent variability. In eye irritancy testing, for example, it has been shown that animal test data have a relatively high coefficient of variation (standard deviation divided by the mean) of approximately 0.5 (the range is from 0 to 1.0). The alternative, in vitro, tests have a lower coefficient of variation, around 0.2 (an advantage of many cell culture tests). If one then plots a comparison of a thousand test agents that are assumed to behave exactly the same in both the animal and the alternative tests, the natural variance of the test systems will produce a graph of points that are widely scattered (Bruner et al., 1996). It will appear as if the alternative test has only a 70 to 80 percent correlation with the animal test.

Consequently, no matter how hard one tries to develop a perfect replacement for an animal test, the inherent biological variability of the animal tests means that perfect correlations (or even 90 percent correlations) are impossible to achieve. It is perhaps not surprising that most of the more promising alternative tests demonstrate no more than a 70 percent correlation with the existing animal tests. Thus, alternative tests nearly always

seem to be inferior to animal tests, but the comparison is clouded by this biological variability (as well as by the lack of a true gold standard for fairly comparing the animal and alternative tests). Ironically, a shortcoming of animal testing (its variability) makes those tests seem to be superior to alternative tests.

Regulatory Practices

Another obstacle to replacing animal tests is that many of these procedures are encouraged, if not required, by national and international laws, regulations, and guidelines. In the United States, few laws actually mandate specific types of animal testing, but some regulations (which implement laws) do specify animal testing. The most common situation, however, is that animal testing is encouraged by a regulatory environment that has historically relied on such testing, thereby developing expectations and biases among regulators as well as corporate toxicologists. Regulators tend to be cautious about switching from methods that seem to be tried and true; anyone who makes such a switch and then runs into problems is inviting unwelcome scrutiny. For their part, corporate toxicologists and decision-makers worry about corporate liability issues if consumers claim injury following use of a company product that had been tested using alternative techniques; judges will want to know why the company moved away from customary testing practices. Notwithstanding these impediments, regulatory practices in the United States and Europe are slowly beginning to embrace alternative methods (see section on Success Stories).

Clearly, any alternative that has been successfully validated is a strong candidate for regulatory acceptance. The key additional feature that needs to be demonstrated to regulatory authorities is the alternative's relevance to their regulatory needs. Does a given agency have use for a test that assesses the endpoint (e.g., skin corrosivity) in question? Does the alternative test adequately characterize that endpoint? To increase the likelihood of gaining regulatory acceptance, developers of alternative tests should involve, or at least consult with, the relevant regulatory authorities throughout all phases of test development and validation.

Validation

Validation of a new alternative method typically involves testing a wide array of chemicals in multiple laboratories. The results are compared to either preexisting or newly generated data from the corresponding animal test. Unfortunately, there usually is no independent way of comparing the animal test and its alternative (see above). New tests that perform well in validation programs are said to be validated. They are then ready to be used in lieu of the animal tests, or at least be reviewed by government authorities for regulatory acceptance.

Validation typically is expensive, time-consuming, and logistically cumbersome. It entails ensuring that an alternative method has been developed to the point that it is ready for validation (a process known as test optimization or prevalidation). One then has to standardize the test protocol, enlist participating laboratories, identify a sufficient number

of chemicals that have already been tested in the conventional animal test, identify sources of the chemicals to be tested, code the chemicals so they can be tested "blindly," and ensure that participating labs are following the protocol exactly, as well as collect, analyze, interpret, and publish the data. Ideally, one should also develop a "prediction model" that specifies, prior to the outcome, how the alternative data will be used to predict the animal data. The details of the validation process have been worked out only since the late 1980s and put into practice in the 1990s, in connection with alternative methods exclusively (e.g., Goldberg et al., 1993). (In the past, and to a certain extent even today, animal tests have not been subjected to the rigors of validation.)
Validation trials are plagued by a variety of problems. Alternative tests have not always been prevalidated. Available animal data are often of dubious quality, quite apart from the issue of their relevance to humans. Participating laboratories do not always follow test protocols exactly. Variability in the animal data often clouds data analysis and interpretation (see above).

The validation process can be particularly frustrating for animal protectionists, given the problems mentioned above. Because animal tests themselves were never formally validated, validation looks like an extra hurdle that only alternative tests must face. Also, animal data are used as the de facto gold standard, giving them the appearance of 100 percent accuracy and all but ensuring that the alternative tests look inferior. Moreover, the high degree of variability of animal data slows down the replacement of these tests. Small wonder that some animal protectionists view validation as a political hurdle for maintaining the status quo.

Cosmetic and Household Product Testing

The use of animals to test drugs and other therapeutic agents has the support of a majority of the American public, but there is much less support for the animal testing of products that are deemed less essential, such as cosmetic or household cleaning products. For example, 60 percent of a sample of 1,000 American adults opposed the use of animals in cosmetics testing, compared to 43 percent and 20 percent opposing animal testing of over-the-counter medicines and prescription drugs respectively (Ward, 1990). About 90 percent of the sample said they would purchase cosmetics that had not been tested on animals.

Animal protectionists have supported public opposition to cosmetic testing on animals by encouraging consumers to buy only "cruelty-free" personal care products. Lists of companies that do not test on animals are available from many animal protection organizations. Over the years, those for and against "cruelty free" designations have come to realize that the issue of animal testing of cosmetics is not as simple as it may have appeared initially. Animal testing can be carried out on the finished product or on individual ingredients. It can refer to testing done not only by the company whose name appears on the product label, but also by upstream manufacturers, contract testing laboratories, or ingredient suppliers. Moreover, nearly all chemicals have been animal tested at one time or another, if not for cosmetic purposes then for some other industry. Consequently, the precise meaning of some "cruelty free" claims has been unclear. Some

European authorities have taken steps to limit or prohibit such animal testing claims in marketing.

Several American animal protection organizations sought to address these issues by forming, in 1996, the Coalition for identical logo available to companies for use on Consumer Information on Cosmetics (CCIC). The CCIC developed a tough new standard that addresses testing of ingredients as well as finished products and allows companies to come on board by selecting a date by which they will no longer conduct or commission animal testing. The CCIC and its partners in Canada and Europe all have similar standards and an packaging and in advertising.

Regulatory practices governing personal care products are somewhat murky. Even though the Food and Drug Administration (FDA) has no explicit animal testing requirements for cosmetics (except in the case of certain coloring agents that are tested for carcinogenicity), the agency has historically used animal toxicity data as its de facto gold standard to settle safety issues. Similarly, many companies have historically felt that the only way to obtain the necessary information to assure the safety of workers and consumers and satisfy regulatory expectations is through testing on animals. However, this situation has changed considerably during the past 10 to 15 years, as substantial progress in reducing animal testing has occurred throughout the industry. Several years ago Avon and Revlon (amongst others) announced that they would no longer conduct animal testing. Mary Kay announced a temporary moratorium on animal testing (later made permanent). More recently, L'Oreal stated that it would not test finished cosmetic products on animals. Gillette has not conducted animal testing on its consumer products during the past several years. Recently, Colgate-Palmolive announced a moratorium on animal testing of adult personal care products, and Procter & Gamble announced an end to animal testing of its current lines of personal and home care products, except where required by law. A number of large companies that still conduct animal testing market most of their cosmetic and other consumer products without recourse to animal testing; only in a few cases do they judge it necessary to conduct some animal tests. (See Table 2 for a summary of company announcements on animal testing.) These announcements and practices certainly do not mean that products from these companies will no longer be safety tested. Avon and L'Oreal, to mention two examples, are actively developing alternative test techniques in their laboratories and have developed a range of in vitro systems to help them assess the safety of both products and ingredients. In addition, both companies have extensive historical databases on their ingredients and on their product lines, which allow them to predict, with a high degree of confidence, how new formulations might react when applied to human hair or skin. Also, cosmetics companies routinely conduct tests of their products (which have very low toxicity as a result of 60 years of refinement) on human volunteers.

In addition, Avon further seeks to protect itself by buying new chemical ingredients with which they are unfamiliar only if the suppliers also provide a standard set of animal toxicity data. L'Oreal has taken a different approach and has not foresworn the testing of ingredients on animals when they consider such testing necessary to protect consumer safety.

None of the large companies that have renounced animal testing in the past several years has made a big play of that fact in their marketing campaigns. This reflects a continuing uneasiness about the issue. Their trade association in the United States, the Cosmetic, Toiletry, and Fragrance Association (CTFA), claims that appropriate animal testing is still vital to ensure the safety of the industry's products.

By contrast, many smaller companies have used the fact that their products are not animal tested as a key part of their marketing. The Body Shop, which sells personal care products in its own retail stores, advertises and labels it products with the phrase "Against Animal Testing." Tom's of Maine, a manufacturer of natural oral care products, includes the statement "Cruelty-free, tested for safety without the use of animals" on its product labels. John Paul Mitchell, which distributes its hair-care products through salons, has the statement "Tested by hairstylists, not on animals" on its products. All three of these companies have grown substantially over the past 15 years.

Thus, consumers are presented with contrasting messages about the importance of animal testing. On the one hand are those companies that argue that some animal testing is still necessary. On the other are those companies that advocate a "cruelty free" approach, as well as animal activists who note that there are plenty of adequate alternative testing methods and that we do not need more lipsticks or dishwashing detergents anyway. How are consumers to sort out these issues?

It is clearly possible for a company to produce many cosmetics without conducting animal tests. Some of the large companies have started to do so and many smaller companies have done so for some time. But there are differences of opinion in the industry about whether alternative tests can completely replace animal testing of cosmetics and, if not, whether companies have an ethical obligation to forego animal testing, even if it means not moving forward with a promising product. The implementation of new safety techniques is aided by the fact that most cosmetics are nontoxic, so the risks for a cosmetics company of not testing on animals are less than those for companies that manufacture a wide range of chemicals or products that vary from nontoxic to very toxic (e.g., oven cleaner).

Some companies continue to hold on to the option of testing chemicals or products on animals, conducting such testing when they consider it warranted but at the same time putting considerable resources into a continuing search for and implementation of alternatives. Some of the large household product companies like Procter & Gamble and Unilever fall into this category. While they still conduct some animal testing, they can be considered to be actively seeking solutions, as opposed to simply being unresponsive to the growing public concern about animal testing.

An examination of progress in two specific areas of testing---the Draize and LD50 tests---illustrates some of the challenges faced by those who seek alternatives to animal testing. These are 2 of the tests that animal protectionists have found most objectionable; both have historically been used in the testing of cosmetics and household products. Campaigns against these two tests have led to significant modifications in test protocols,

considerable research in in vitro toxicology to find alternatives, and major changes in regulatory attitudes about animal tests and potential alternatives.

The Draize Eye Irritancy Test

During the Second World War, animal-based protocols were developed to assess the effects of chemical warfare agents on eye irritancy. In 1944 John Draize and his colleagues developed a scoring system to grade eye damage; adverse effects on the cornea accounted for almost 80 percent of the maximal score. Since the war, the Draize test (as it became known) became the standard procedure for estimating the eye-irritancy potential of a wide variety of products, including shampoo, hair spray, deodorant, detergents, drugs, and pesticides (see Frazier et al., 1987, for a comprehensive review). In the standard version of the Draize test, a chemical or product is placed in one eye of a rabbit, usually without local anesthetic, while the other eye is used as a control. Irritation levels are observed over several days and damage to the cornea (e.g., opacity), conjunctiva (conjunctivitis), and iris (iritis), as well as discharge, are recorded and combined into a single score. The maximum score possible is 110, which usually means destruction of the eye. Albino rabbits were chosen for the test because they have large, unpigmented eyes in which it is relatively easy to observe inflammation and irritation. Rabbits' eyes are generally more sensitive to irritating agents than are humans' eyes; consequently the test has a relatively low risk of giving false negative results (i.e., mischaracterizing an irritant as a nonirritant). This sensitivity is unfortunate for the test animals but is considered a valuable feature from a regulatory perspective because the chances are good that a substance with little or no effect on a rabbit will be safe for a human eye.

In the mid-1970s, animal activists began protesting against the use of animals for the safety testing of cosmetics and, in particular, the use of the Draize test (see Rowan, 1984, for more details). These protests initially had relatively little impact; an official of the Cosmetic, Toiletry, and Fragrance Association (CTFA) in Washington, D.C., declined to explore any initiatives to develop alternatives. However, in 1979 New York--based activist Henry Spira began a campaign against the eye-irritancy test, organizing almost 400 animal protection groups to join the Coalition Against Rabbit Blinding Tests (see Table 3 for a timeline of this and other developments).

Initially, Spira approached Revlon, identified as one of the leading cosmetics companies, and asked the company to devote 0.1 percent of their annual profits (then the equivalent of approximately $170,000) to research and development on alternatives to the Draize test. Revlon rejected the request and passed the matter on to the CTFA for its consideration. Spira then mounted a year-long campaign against Revlon that ended with the company announcing at the end of 1980 that it was setting up a research program at Rockefeller University, in New York City, to develop an alternative to the Draize test. Revlon, which had not welcomed the attention it had received, also suggested that the rest of the industry might join in their initiative. The other cosmetics companies then banded together and established a $1 million fund that was awarded to the Johns Hopkins School

of Hygiene and Public Health in 1981 to establish the Center for Alternatives to Animal Testing (CAAT).

When Revlon and the CTFA set up these alternatives research programs, there was very little research being done on potential alternatives to the Draize test. A pilot cell culture study had been conducted at Hazleton Laboratories in the United Kingdom, and Procter & Gamble was attempting to gain regulatory approval for its low-volume eye test (which used one tenth of the standard dose in the Draize test, based on evidence that it gave results that correlated better with human data and caused less damage to rabbits' eyes). Cynics among both the scientific and animal protection communities felt that the Rockefeller and Hopkins initiatives were mainly exercises in public relations. Some of the scientists believed the technical rationale for starting the projects was very weak and some animal protectionists argued that neither university was really interested in pursuing alternative methods.

The cynics have been proved wrong. Nearly 20 years after the Revlon and CTFA initiatives began, the situation has changed dramatically, thanks largely to the changes set in motion by the Spira-led campaign against the Draize test. Many cosmetic and household product companies now have active in vitro toxicology programs and all have made major modifications in the way they conduct their safety testing. Corporate scientists who saw the initial research programs as little more than public relations exercises have become convinced of the scientific merit of pursuing alternative approaches. While the FDA and other regulatory authorities still recognize the Draize test or its variants as the final standard for eye irritancy, regulatory attitudes have changed considerably over the past several years and numerous modifications and potential replacement batteries are now under serious consideration (see below).

Alternatives to the Draize Eye Irritancy Test

The area of eye-irritancy testing can be used as an example of how a combination of common sense, small modifications, and innovative new technology is revolutionizing our approach to toxicity testing (see The Alternatives Report, Volume 3[5/6], 1991, for a summary and Food and Chemical Toxicology, Volume 31, issue 2, 1993, for an extensive discussion of eye-irritancy modifications arising out of a regulatory agency workshop held in September, 1991).

The search for alternatives to the rabbit eye-irritancy test began in earnest in 1981 with creation of the CAAT. Since then, the following changes in practice and modifications to the rabbit test have become sufficiently accepted to be endorsed by a wide variety of regulatory authorities, even if formal approval of the modifications has not always been forthcoming.

It has been well known that strongly acidic and alkaline substances are eye irritants. Now, companies routinely identify strong acids and bases as eye irritants without confirming this fact in an animal test. It has also been shown that it is possible to reduce the number of rabbits used in the Draize test without compromising safety standards. Instead of using

the standard 6 or more animals, one can judge whether a test agent should be labeled as an eye irritant in 3 animals or even fewer. The procedure involves dosing the eye of a single rabbit. If a positive response is observed, then the agent could be labeled an irritant without further animal testing. However, if the response is negative (or confirmation of the positive response is required), then an additional 2 rabbits can be used. If one or both rabbits are negative, then the substance is labeled a nonirritant. If a positive response is observed in both additional animals, the substance should be labeled an eye irritant. This reduction alternative was shown to provide almost exactly the same classification as the use of 6 rabbits. Scientists from the Environmental Protection Agency (EPA) and the Food and Drug Administration (FDA) reviewed individual rabbit data previously submitted to these agencies and conducted a statistical analysis of all possible groups of three animals (out of the 6 to 9 tested for each agent) and found almost no difference in overall classification. As a result, regulators at the EPA and FDA have become much more willing to adopt the 3-rabbit protocol.

Finally, Procter & Gamble has been attempting to gain official approval of its low-volume eye test (LVET), which uses one-tenth the standard dose in the eye (0.01 vs. 0.1 ml/kg). The lower dose produces less trauma in the rabbit eye and hence qualifies as a refinement alternative. To date, the LVET has not been accepted by U.S. regulatory authorities, although Procter & Gamble has found the test to be both effective and more humane.

One of the first modifications to the rabbit test that was explored was the use of local anesthetics. Unfortunately, after years of investigation, there are still many questions about the utility of local anesthetics to prevent short-term pain, especially while other modifications are steadily reducing the need for local anesthetics.
While the above modifications and changes in practice for Draize testing were being explored and argued, a wide range of new test systems for eye irritancy have been developed and promoted. These include the following (see Frazier et al, 1987 for more details):

- Several cell culture approaches are now available, including the widely used Neutral Red assay, which measures cell viability, and the Fluoroscein-Leakage assay, which measures the integrity of the junctions between cells. The Neutral Red assay was developed by Ellen Borenfreund as part of the Revlon-sponsored research at Rockefeller University (see above).
- Several approaches using the chorio-allantoic membrane (CAM) of the developing chick embryo have been developed, including the original CAM assay and its European variant, the Hen's Egg Test. The CAM itself is not supplied with nerves, so irritants applied to this membrane are presumed not to trigger a pain response. (However, the chick embryo is killed in the assay.)
- The use of bovine eyes from slaughterhouse material has been shown to be promising as a prescreen for eye irritancy. If a test agent produces a positive reaction in the Bovine Corneal Opacity and Permeability (BCOP) assay, then it can be labeled as an eye irritant without further animal testing. (Animal advocates have some ambivalence about a test based on slaughterhouse byproducts, but

welcome the BCOP assay as an interim step in moving away from the Draize test.)

- The EYTEX series of assays, developed by In Vitro International, relies on purely chemical approaches to assess irritancy. These assays involve a solution of protein and other biological macromolecules that changes in clarity when a test agent is added. The test is simple, fast, and relatively cheap, and a number of companies and government laboratories claim to have produced good results. However, other laboratories disagree and question how and why this proprietary mixture of chemicals should work.
- Several companies have developed artificial skin systems that can be used to assess the irritancy potential of test agents. These systems are bioengineered by seeding an artificial basement membrane with skin fibroblasts and sometimes keratinocytes (the major skin cell type) to produce a product that simulates some aspects of human skin. Procter & Gamble has worked with one of these companies, Advanced Tissue Sciences (now defunct), to produce a test protocol that can assess the irritancy potential of test materials that are solid and water-immiscible, an assessment not possible in the aqueous medium of a cell culture.
- A test system called the Silicon Microphysiometer has been developed to record a very sensitive measure of cellular metabolic rate. Although it is an expensive piece of equipment, it has considerable potential in the laboratory and it can also be used to assess a cell culture's recovery from perturbations caused by a test agent (hence simulating to some extent the recovery of the eye from an irritant reaction).
- Health Designs Inc. has developed a computer model that can provide a rapid, preliminary assessment of eye irritancy. One application is in quickly screening large numbers of related compounds, with promising compounds moving on to more definitive testing.

It is clear that considerable progress has been made in developing and implementing alternatives to the Draize test. While no single alternative or battery of tests has yet been validated as a total replacement of the Draize test, notwithstanding a large-scale international validation study (Balls et al., 1995a), the number of animals required in such testing has been reduced and could be further reduced. In addition, the animal distress caused by this testing has also been reduced by appropriate pre-screening programs. It should also be noted that no one had ever undertaken a careful and systematic analysis of human eye injury caused by chemicals. The key cellular and organic changes that occur during irritancy have recently been evaluated in a project funded by Procter & Gamble and conducted by International Life Sciences Institute. As a result, we now are beginning to develop the type of mechanistic understanding of eye injury that will permit us to identify and validate more-predictive alternatives.

While the effort invested in developing alternatives to the Draize test has been considerable compared with most other animal tests, substantial progress has been made in other areas of testing as well. This progress has facilitated the decision by several companies to stop testing on animals altogether, as mentioned above. These companies have decided that they can rely on a safety-evaluation process that no longer involves

testing on animals, either by the companies themselves or by contract testing laboratories. A typical approach to safety assessment of a new product by such a company might involve reviewing the following: the corporation's historical toxicity database on the product line, the physical characteristics of the product (e.g., pH), the relevant structure-activity relationships for key ingredients, available animal data on the ingredients in the product, and data from a variety of in vitro tests (perhaps including a CAM-based test and the BCOP assay). The final step might be testing on paid human volunteers. Regulatory authorities are also becoming more comfortable with some of the new non-whole animal tests.

As a result of advances in alternative methods and changes in corporate responsiveness to public pressure, there is evidence that the use of animals in cosmetic testing has declined substantially (Figure 1). In Great Britain, where good statistics are available, the number of animals used annually in the testing of cosmetics and toiletries has dropped from 28,600 to 590 between the end of the 1970s and 1998 (Anon., 1993, Anon., 1999). Indeed, Great Britain has instituted a de facto ban on animal testing of cosmetics by revoking all existing licenses to conduct such testing and announcing that no further licenses would be issued for this purpose.

The LD50 Test

The LD50 test was originally developed to standardize batches of powerful biological medicines such as digitalis (Trevan, 1927). Each batch of the drug varied in potency. Consequently, it was important to have a method to help ensure that new preparations were of uniform potency before being sold to pharmacists. Trevan used the lethality in the LD50 test as a means of gauging potency. The technique required the use of from 60 to 200 animals, usually mice, for each LD50 determination.

The LD50 later became one of the first toxicity tests to be conducted on any chemical or product, and the LD50 value itself came into use as a baseline toxicological measure. The LD50 value is the dose that kills 50 percent of a group of animals to which it is administered, hence the term lethal dose 50 percent, or LD50. The dose is usually administered by mouth, but dermal, inhalation, and intravenous LD50s can also be determined. The test protocol for the classical LD50 entails first estimating the approximate range of lethal toxicity and then administering several doses around that range to 5 groups each of males and females. The animals are observed for up to 14 days. Those who survive are euthanized, and the tissues of all the animals, including those who die, should be examined pathologically.

The LD50 test began to come under attack from animal activists in the 1970s. The initial criticism was on humane grounds (poisoning animals to death by force-feeding toxic substances to them provided an easy and legitimate target). Later, technical criticisms of the test published in the toxicological literature were also used (e.g., Morrison et al., 1968), adding important weight to the campaigns. For example, it was determined that the LD50 value should not be regarded as a biological constant because so many factors---including the animals' species and strain, age, gender, diet, bedding, ambient

temperature, caging conditions, and time of day--- can all affect the LD50 value obtained, sometimes up to one thousand fold. In 1981 Zbinden, a well-respected toxicologist, published a review that concluded there was little public-health justification for conducting the classical LD50 test (Zbinden, Flury-Roversi, 1981). (See Table 4 for a timeline of this and other developments.)

Aside from the variability in test results, there were two other key technical criticisms of the LD50 test. First, the influence of so many variables on the LD50 value, as mentioned above, means that there is little point in using large numbers of animals in order to achieve a statistically precise LD50 figure. In other words, a test agent that had a rat LD50 of 100 mg/kg (milligrams of chemical per kilogram of the animals' body weight), precisely determined, could cause human toxicity at a dose of from 1 to 10,000 mg/kg. It would not make much sense to determine that the rat LD50 had a standard deviation of 15 given the considerable uncertainties in extrapolating to other species. Thus, it would be quite sufficient to know that the lethal dose for a rat was approximately 100. Consequently, the LD50 test should be recognized as providing, at best, only a ballpark estimate of human lethality. The humane advantage of eliminating the demand for statistical precision is that one could cut the number of animals required to determine lethal doses by 80 percent or more without compromising human safety standards at all. Second, animal activists were concerned about the LD50 endpoint. Why dose animals and simply let them die from poisoning? Some effort has been devoted to developing modifications of the LD50 test that attempt to avoid death as an endpoint. In these modifications, the animals are monitored closely and when they appear to be severely compromised, they are euthanized.

Given the technical criticisms, the campaign against the LD50 has focused on reducing the number of animals used in the test and on looking for nonlethal endpoints. At one level, the campaign has been very successful in that few individuals or organizations still defend the classical LD50. Nonetheless, the inertia of traditional practices is hard to overcome, and many classical LD50 determinations are still performed simply to complete product registration tables and satisfy the demands of regulatory authorities that have yet to hear, apparently, that nobody supports the classical LD50 measure.

Alternatives to the LD50 Test

A number of alternatives to the classical LD50 have been developed and several have been incorporated into guidelines on acute toxicity promulgated by the Organization for Economic Cooperation and Development (OECD), an international trade organization that includes the United States, Japan, and several member states of the European Union. The baseline of comparison for the alternatives is OECD Guideline 401. In the original version of Guideline 401 (issued in 1981), 10 animals (5 females and 5 males) were used at each of at least 3 dose levels selected to produce a range of toxic effects and mortality rates. (A few animals are used initially in a so-called "dose ranging" or "sighting study" if there is little information on what the lethal dose might be.) Guideline 401 was modified in 1987, calling for only 5 animals of one sex (females) at each dose (subtotal = 15

animals), with confirmatory testing at the LD50 level in males (adding another 5 animals). The minimum number of animals used was thus reduced from 30 to 20.

The alternative methods to Guideline 401 include the following:

- Fixed Dose Procedure---In 1984 the British Toxicology Society concluded that precisely determined LD50 values are rarely justified (Anon., 1984) and proposed an alternative test, the Fixed Dose Procedure (FDP) (van den Heuvel et al., 1987). Compared with the LD50 test (either older or newer versions of OECD Guideline 401), the FDP is both a reduction and refinement alternative (using clear signs of toxicity, termed "evident toxicity," not lethality, as its endpoint), yet it still provides data for product labeling and classification. The average number of animals used in the FDP is 14 (van den Heuvel, 1990). The FDP was incorporated into OECD Guidelines as Guideline 420 in 1992.
- Acute Toxic Class Method---In Germany, regulators developed the Acute Toxic Class (ATC) method as an alternative to the precisely determined LD50. Death is still the endpoint, but the ATC is a reduction alternative, typically using only 6 to 12 animals per test. It was incorporated as OECD Guideline 423 in 1996.
- Up and Down Procedure---American scientists developed the Up and Down Procedure (UPD) as a third alternative to the LD50 test. As with the ATC method, death is still the endpoint, but the UDP uses only 8 animals per test on average. The UDP yields a "point estimate" of the LD value, unlike the FDP and ATC Methods, which classify the LD50 into one of several ranges. The UPD was incorporated as OECD Guideline 425 in 1998.

With three alternatives to the LD50 included in its guidelines (420, 423, and 425), the OECD announced in June 1999 that it would shortly delete Guideline 401. This was welcome news. Even with alternative methods already on the books, Guideline 401-both old and new versions-is still being used in many countries. In 1997, for example, over 139,000 animals were used in Great Britain to determine classical LD50 values (Home Office, 1997). In November 1999 the UK announced it would no longer issue project licenses for Guideline 401 toxicity testing.

Meanwhile, hope is building that in vitro tests will supplant the use of animals in acute toxicity testing. The Multi-center Evaluation of In Vitro Cytotoxicity, known as MEIC, has produced interesting results in its assessments of in vitro acute toxicity tests. In a departure from previous evaluations of in vitro tests, animal data were not used as a de facto gold standard. Instead, both the newly generated in vitro data and preexisting LD50 data were compared with human data on lethal blood concentrations, gleaned from hospital records and similar sources. This work has been spearheaded by Bjorn Ekwall and the Scandinavian Society for Cell Toxicology.

The changes that have taken place in attitudes to the LD50 test and in acute toxicity testing generally are based on a mixture of public pressure, moral concern, and sound scientific argument. While some activists would like to see faster progress and complete replacement of animals, the speed of change of both testing practices and toxicological

attitudes over the past 15 years actually has been quite surprising. And we are getting closer to completely replacing the use of animals in acute toxicity testing with cell culture systems, perhaps in conjunction with computer modeling.

The International Scene

Much of the scientific work on alternatives has been coordinated and conducted in Europe, driven by public pressure in European countries, particularly in the United Kingdom, Germany, and the Netherlands. The ongoing attention in the European Union (EU) to the issue of alternatives to animal testing has had a spillover effect in the United States, where the issue had otherwise lost some momentum in the 1990s.

The European Union Cosmetics Directive has been a focal point for ongoing alternatives work in Europe. Enacted in 1993, it called for a ban on the marketing in Europe of any cosmetics tested on animals after 1998, including products manufactured in the United States. The marketing ban was to be implemented in 1998, but an escape clause allowed the EU government to postpone it until 2000, declaring that insufficient progress had been made in developing alternatives to the animal tests used in assessing cosmetic safety. While animal protectionists were disappointed by the delay, the directive kept some pressure on the cosmetics industry to continue developing alternatives. Various conferences spearheaded by the COLIPA, the European cosmetics trade association, have been held to provide updates on progress.

A new amendment to the EU cosmetic directive (the 7th) has been proposed. It calls for an immediate ban on any animal testing of finished products. Any animal testing of cosmetic ingredients would be banned when alternatives become available or within 3 years of implementation of the directive, regardless of the availability of alternatives. The European Commission has stated that only one two-year postponement to the ingredients testing ban would be considered. Consequently, an ingredients testing ban would take effect no later than 5 years after the implementation of the new amendment. The 7th amendment would apply only to EU countries, due to potential problems with World Trade Organization rules.

Apart from encouraging industry to invest in the development of alternative methods, the Cosmetics Directive also led directly to the establishment of the European Centre for the Validation of Alternative Methods (ECVAM) in 1993. The EU created ECVAM, in part, to provide a mechanism for the validation and regulatory acceptance of alternatives for the safety assessment of cosmetics. These alternatives would then have enabled the cosmetics ban to take effect. ECVAM also has a broader mandate to speed the pace of progress on alternatives generally, by coordinating the development, validation, and regulatory acceptance of promising methods.

In addition to ECVAM, another important institution on the international scene is the OECD (see above), which promotes international trade through, among other activities, producing and harmonizing test guidelines acceptable to all 27 member countries. As we already mentioned, the OECD has been a major player in the LD50 issue, promulgating

guidelines for the alternative assays (the Fixed Dose Procedure, the Acute Toxic Class Method, and the Up and Down Procedure) and agreeing to phase out use of the LD50 test (Guideline 401). The OECD operates on the principle of "mutual acceptance of data," which means that member countries are obligated to accept data generated according to OECD guidelines. This principle is not always honored in practice, particularly in the United States and Japan. Animal protectionists will be watching to see whether regulatory agencies in their respective countries accept data from the various LD50 alternatives and reject data from the LD50 test (401).

It is extremely important that testing protocols be harmonized internationally to reflect the latest developments in alternative methods. Otherwise, companies would be forced to adopt a lowest common denominator approach to toxicity testing, perhaps using the latest alternatives but also conducting the traditional animal tests to satisfy draconian requirements in the most regressive countries.

Success Stories

Until recently there was no formal process for regulatory agencies in the United States to evaluate alternative methods for acceptance. Each agency assessed the strengths and weaknesses of new tests on a more or less ad hoc basis. Moreover, there was little interagency coordination for tests that were applicable to multiple agencies. Consequently, individuals and organizations backing new alternative tests had to take their case to several agencies independently. This was frustrating not only to developers of alternative tests, but also to the companies that wanted to use these tests and to the animal protectionists who supported alternative methods.

Recognizing this problem, a coalition of representatives from industry, animal protection, and academia joined forces to lobby the U.S. Congress for the inclusion of favorable language in the 1993 National Institutes of Health (NIH) Revitalization Act. They were successful. The language called upon the National Institute of Environmental Health Sciences (NIEHS), one of the NIH institutes, to identify the key features of the validation and regulatory acceptance of alternative methods and to establish a process for the regulatory acceptance of these methods. In response to the legislation, the NIEHS in 1994 spearheaded the establishment of the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM), which consists of representatives from more than a dozen federal agencies. In 1997 ICCVAM published its landmark report, Validation and Regulatory Acceptance of Toxicological Methods. In 1998 ICCVAM, which is staffed by people who have full-time obligations to their home agencies, expanded its capability by establishing a support center, the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). (For a timeline of this and other developments associated with the validation and regulatory acceptance of alternative methods, see Table 5.)

With this center in place, ICCVAM began reviewing alternative methods for regulatory acceptability. The process involves establishing a scientific peer review panel that evaluates the submission of the one or more organizations backing the method under consideration. The committee then evaluates the strengths and weaknesses of the new method and issues its recommendations, which may include modifications in the new

procedure. The recommendations are then taken up by the appropriate individual agencies. Given the involvement of agency personnel throughout the review process, a positive recommendation by the peer review committee is considered tantamount to federal government approval. As of August, 2000, ICCVAM-associated peer-review panels have endorsed two alternative methods: the Local Lymph Node Assay and Corrositex® (see below).

The establishment of ICCVAM in the United States and ECVAM in the EU created official government bodies charged with judging whether proposed alternative methods have been adequately validated and are acceptable for regulatory purposes. (ECVAM's mission also includes the important function of actually coordinating the validation of new methods.) To date, ICCVAM or ECVAM has approved the following alternative toxicity tests:

The Local Lymph Node Assay---The Local Lymph Node Assay (LLNA) is the first test method to have gone through the ICCVAM regulatory acceptance process, including peer-review by an expert panel. In 1998 ICCVAM's expert panel approved the LLNA as an alternative to the Guinea Pig Maximization Test (GPMT) for assessing allergic contact dermatitis. The LLNA, which uses mice, is both a reduction alternative and refinement alternative in that it requires fewer animals and is less painful than the GPMT. Peer reviewers determined that the LLNA is also less expensive and less time consuming, but no less accurate, than the GPMT. Allergic contact dermatitis is an allergic reaction that occurs after skin comes in contact with certain inflammation-causing substances. Cosmetics, metals, plants, and medications are just some of the substances that can cause such a reaction. According to federal statistics, 10,200 guinea pigs experienced unrelieved pain and distress in the GPMT in fiscal year 1999 (Adams, 2000). The ICCVAM Web site (iccvam.niehs.nih.gov) provides the peer-review committee's report on the LLNA, as well as details about ICCVAM, its approval process, and its plans for future activities.

The Transcutaneous Electrical Resistance Assay---The Transcutaneous Electrical Resistance (TER) Assay was accepted by ECVAM in March 1998 as a valid method for assessing the corrosive effects of chemicals on skin. The TER Assay is an in vitro test that uses rat skin cells and is based on the observation that corrosive substances reduce the electrical resistance of skin. A formal validation comparing the results from the TER assay and data from previous corrosivity tests on animals indicated that the alternative test was useful for testing a diverse group of chemicals. The TER Assay technically is a replacement alternative because it replaces the use of animals as test subjects. Unfortunately, the test does involve killing animals to obtain fresh skin cells. On the positive side, only 1 animal is killed to supply skin cells to run the assay, whereas standard corrosivity testing involves up to 3 animals per chemical. Also, the TER Assay can be considered a refinement in that the animals themselves are not exposed to any pain or distress associated with corrosivity testing. For more information about the TER Assay, see the Web sites for the Institute for In Vitro Sciences and the Invitroderm Company.

Episkin---The Episkin assay and similar artificial skin models were accepted by ECVAM in March 1998 as another valid replacement of corrosivity testing in animals. Episkin consists of layers of human skin cells on a collagen matrix. Chemicals are applied to Episkin and irritation/corrosivity is measured by the release of pro-inflammatory agents. A formal validation study comparing the results from Episkin and data from previous corrosivity tests on animals indicate that the alternative test is useful for testing a diverse group of chemicals and substances, such as cosmetics, sunscreens, and other topically applied products. For more information about Episkin, see the Web sites of the Institute for In Vitro Sciences and the Invitroderm Company.
Corrositex® ---In June 1999 an ICCVAM expert panel approved Corrositex® as an alternative method for assessing skin corrosivity. When Corrositex® is used, a chemical or chemical mixture being tested is placed on a type of artificial skin barrier made of collagen. Beneath that layer is a liquid containing a pH indicator dye that changes color when it comes into contact with the chemical being tested. The corrosivity of a chemical is determined by the time it takes for the chemical to penetrate the artificial skin and produce a color change. According to Williams Stokes, the associate director for animal and alternative resources for the NIEHS, "Current regulations usually require 3 animals for each chemical that is evaluated for skin corrosivity and dermal irritation. Since there are more than 2,000 chemicals introduced each year, [the use of Corrositex®] could result in a considerable reduction in the use of laboratory animals to identify corrosives." The text of the peer-review committee's report on Corrositex is available at iccvam.niehs.nih.gov.

3T3 Neutral Red Uptake Phototoxicity Test---In July, 1998 the EU officially accepted an in vitro test for the assessment of phototoxicity, an important consideration for sunscreens and other topical products. The acceptance of the 3T3 NRU PT Test marked the first time that an alternative test has been accepted by the EU's official sanctioning body responsible for pharmaceutical and cosmetic products, Directorate General (DG) III. Additionally, it was the first time that an in vitro toxicity test had been accepted following rigorous experimental validation under "blind" conditions, a validation procedure that no animal-based test had ever undergone. Previously, 3T3 NRU PT was formally accepted by two other important regulatory bodies of the EU: ECVAM's Scientific Advisory Committee and the DGXI, an EU government ministry responsible for environmental and industrial chemicals. Consequently, the 3T3 NRU PT Test is the only EU-sanctioned toxicity test for phototoxic potential both for industrial and environmental chemicals and for pharmaceutical and cosmetic products. 3T3 NRU PT is a cytotoxicity test in which mouse embryo-derived cells of the 3T3 cell line are exposed to test chemicals with and without exposure to type A ultraviolet (UVA) light. Cytotoxicity (cell death) is measured as inhibition of the capacity of the cell cultures to take up a vital dye (neutral red). For a chemical to be labeled as having phototoxic potential, the assay requires a significant increase in toxicity in the presence of UVA.

New Large-Scale Animal Testing Programs
Despite the positive developments in the animal testing arena--including the establishment of ECVAM and ICCVAM, the validation and regulatory acceptance of several alternative tests, and the decline in the numbers of animals used in safety testing--

-there are a few disturbing developments in the United States. The Environmental Protection Agency (EPA) recently launched three new testing programs that, as originally planned, would rely heavily on animal tests and all but ignore alternative approaches. Moreover, two of the three programs would subject hundreds of thousands of animals to potentially painful or lethal animal tests. Fortunately, the EPA has begun to modify at least some of these programs in light of recommendations and criticism from the animal protection community and pro-alternatives voices within the scientific community.

The High Production Volume Chemical Testing Program

The High Production Volume (HPV) Chemical Testing Program seeks to generate a standardized toxicity profile on chemicals used widely in commerce (i.e., the 2,800 chemicals produced or imported into the United States in quantities of 1 million pounds or more per year). Many of these chemicals have some toxicity data associated with them, but the HPV Program aims to generate the "missing" data and to make the resulting toxicity profiles publicly available.

The toxicity tests selected for the HPV Program are those of the Screening Information Data Set (SIDS), developed by the OECD. The SIDS battery includes tests for both health and environmental effects, but the health effects tests have drawn the bulk of the criticism from those concerned about the HPV program's emphasis on animal tests. The health effects tests include animal tests for a number of endpoints, including acute, subchronic, developmental, reproductive, and genetic toxicity. Critics estimated that the HPV Program could consume over 1 million animals.

The HPV Program was announced by Vice President Al Gore in 1998 as a partnership of the EPA, the Chemical Manufacturers Association, and the Environmental Defense Fund. The EPA called for testing to begin in 1999 and extend through 2004. Animal protection organizations criticized the EPA for developing the program with little public input. In response, the EPA held a series of public meetings. The CAAT also organized meetings to assess how alternative methods could be incorporated into the program. As a result of these meetings and substantial pressure from animal protectionists, the EPA made a series of announcements over the course of several months, pledging changes. The announcements culminated in an October 1999 letter from the EPA to the chemical companies involved in the HPV Program, outlining considerable modifications. The key modifications include the following:

A two-year delay in testing individual chemicals (those not grouped into categories), so that pending alternative tests can be validated and incorporated into the program

A $4.5 million commitment from the NIEHS, the National Toxicology Program, and the EPA for the development of nonanimal test methods

A partial amnesty for chemical companies to reveal and share testing data previously withheld from the EPA, thus reducing the number of new tests that will be conducted

A testing exemption for certain tests on specific types of chemicals (e.g., "closed system intermediates")

EPA examination of the totality of information on chemicals and, based on these analyses, possibly allowing companies to avoid conducting certain tests, rather than require a rote "check list" approach that includes many animal tests

EPA encouragement of participating companies to use in vitro genetic toxicity testing to generate any needed genetic toxicity screening data
Reducing the numbers of animals used
EPA encouragement of participating companies to maximize the use of scientifically appropriate categories of related chemicals and structure activity relationships, thereby testing representative chemicals instead of all chemicals in the selected groups
These changes were sufficient for various animal protection organizations to suspend their grassroots campaigns against the HPV Program, though the organizations are still advocating for the vigorous implementation of the changes outlined above. Further information about the HPV Program can be found at The HSUS's Web site and at the EPA Web site.

Endocrine Disruptor Screening Program

The EPA is also developing a program to screen chemicals that may cause harmful effects in people by disrupting their *endocrine system* (the hormone-secreting glands that regulate important bodily functions such as blood-sugar levels). Some scientists have suggested that such endocrine disruptors are prevalent among the chemicals with which we come into contact and may cause problems such as low sperm count in mature males and premature sexual development in young girls. Congress took up this issue and enacted legislation that called upon the EPA to establish an endocrine disruptor (ED) screening program and to adhere to an ambitious timetable.
The EPA established the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) and charged it with coming up with an appropriate battery of tests. EDSTAC's proposed battery relies heavily on animal tests. Thousands of chemicals would be run through this battery, unproven in its relevance ED screening, making it one of the largest animal testing exercises in history.
Animal protectionists and sympathetic scientists have criticized EDSTAC's proposed battery on several grounds: it is based on an unproven hypothesis about widespread harm to people from endocrine disruptors, it has never been validated for the purpose of ED screening, and it ignores alternative approaches. There is considerable concern that this initiative has so much momentum that it cannot be stopped or overhauled, despite its flaws. Further information can be found at the EPA website.

Voluntary Children's Chemical Evaluation Program

The Voluntary Children's Chemical Evaluation Program (VCCEP) is a third EPA program that has caused concern among those seeking alternatives to animal testing. TheVCCEP's goal is to generate toxicity profiles for chemicals to which children face a high exposure. As with the HPV and ED programs, the VCCEP's proposed testing battery relies heavily on animal tests. The details of the initiative are still being worked out. The EPA first thought it would impose the testing on industry by issuing a mandatory call for testing. The agency decided in 1999 to abandon that approach and seek an agreement with industry to have the testing conducted voluntarily. The EPA then held a series of "stakeholder" meetings to solicit advice from the chemical industry, children's health advocates, animal protectionists, and others, regarding which chemicals should be tested,

what tests should be used, and how the program should be structured. Prior work by the agency and its advisory committees had already led to the proposed test battery. The EPA has issued a draft list of approximately 50 chemicals for the first round of testing.
The VCCEP is related to the HPV Program in two ways. First, the chemicals to be tested will be drawn primarily from the 2,800 HPV chemicals. Second, both programs are part of Vice President Gore's Chemical Right to Know Initiative. The animal protection community is hopeful that it can convince the EPA to make the same sorts of positive changes in the VCCEP as it has done for the HPV program. Further information about the VCCEP can be found at the web sites of The HSUS and the EPA.

Concluding Remarks

The use of animals and alternatives in safety testing has changed considerably over the past twenty years. The animal protection community has pressed industry and government for humane reform. Academic centers and corporations devoted to developing alternatives to animal testing have become established, as has a federal government process for assessing validation and regulatory acceptance of alternative methods. New methods are moving through this process. More and more corporations are announcing internal bans or cut-backs on animal use. Industry has more or less adopted the goal of zero animal use. And most observers are increasingly recognizing the limitations of animal-based tests as predictors of human safety.

Technical innovation itself, apart from humane concerns, is also driving the move away from animal testing. One potentially exciting technology known as toxicogenomics involves the use of new genetic engineering techniques. It has been shown that a range of genes are switched on in response to damage or strain caused by certain toxic agents. One company, Xenometrix, has begun to exploit this responsiveness by combining the genetic material that is responsible for switching on these genes with a gene that will produce a colored product when it is switched on. The resulting cell culture will change color when it starts to react to a particular type of toxic agent. Using this technology, one can produce a variety of cell cultures (including those with human cells) that will respond to specific toxic insults. Eventually, it is hoped, the pattern of toxic insults to the different cell cultures might produce accurate predictions of what the agent might do in humans and also provide basic information on the possible mechanisms of toxicity. This would be a far cry from animal tests developed decades ago that essentially use animals as "black boxes."

Reaching the goal of zero animal use is a long-term process and there will be setbacks---witness the recent EPA animal testing proposals. Some decision-makers may have taken the mid-1990s lull in activist attention to the animal testing issue as a sign of lack of interest, rather than a shift in priorities in the face of positive momentum on the testing issue. The furor over the EPA's HPV program indicates that the animal protection movement has not abandoned the animal testing issue and, indeed, is willing to mobilize considerable energy and resources to combat negative developments on this front.

References

Adams, C. 2000. Cuteness counts: Guinea pigs will be guinea pigs no more. *Wall Street Journal*, Jan. 5, page B1.

Anonymous. 1984. Special report. A new approach to the classification of substances and preparations on the basis of their acute toxicity. *Human Toxicology* 3:85-92.

Anonymous. 1993. Scientific procedures on living animals in Great Britain in 1992. *Alternatives To Laboratory Animals* 21:406-408.

Anonymous. 1999. Scientific procedures on living animals in Great Britain 1998. *FRAME News 48:6-8.*

Balls, M., B.J. Blaauboer, L. Bruner, R.D. Combes, B. Ekwall, J.H. Fentem, R.J. Fielder, A. Guillouzo, R.W. Lewis, D.P. Lovell, C.A. Reinhardt, G. Repetto, D. Sladowski, H. Spielmann, and F. Zucco. 1995b. Practical aspects of the validation of toxicity test procedures. The report and recommendations of ECVAM Workshop 5 ("Amden II report"). *Alternatives To Laboratory Animals* 23:129-14.

Balls, M., B. Blaauboer, D. Brusick, J. Frazier, D. Lamb, M. Pemberton, C. Reinhardt, M. Roberfroid, H. Rosenkranz, B. Schmid, H. Spielmann, A-L. Stammati, and E. Walum. 1990. Report and recommendations of the CAAT/ERGATT workshop on the validation of toxicity test procedures ("Amden I report"). *Alternatives To Laboratory Animals* 18:313-337.

Balls, M., P.A. Botham, L.H. Bruner, and H. Spielmann. 1995a. The EC/HO international validation study on alternatives to the Draize eye irritation test. *Toxicology In Vitro* 9: 871-929.

Bruner, L.H., G.J. Carr, M. Chamberlain, and R.D. Curren. 1996. Validation of alternative methods for toxicity testing. *Toxicology In Vitro* 10:479-501.

Frazier, J.M. 1990 Scientific criteria for validation of in vitro toxicity tests. Environment Monographs no. 36. Paris: Organization for Economic Co-Operation and Development (62 pp.).

Frazier, J., S. Gad, A. Goldberg, and J. McCulley. 1987. A Critical Evaluation of Alternatives to Acute Irritation Testing (*Alternative Methods in Toxicology*, Volume 4). New York: Mary Ann Liebert, Inc.

Gold, L. 1998 What do animal cancer tests tell us about human cancer risk? Overview of analyses of the carcinogenic potency database. *Drug Metabolism Reviews* 30(2):359-404.

Goldberg, A.M., J.M. Frazier, M.S. Dickins, O. Flint, S.D. Gettings, R.N. Hill, R.L. Lipnick, K.J. Reskers, J.A. Bradlaw, R.A. Scala, B. Veronesi, S. Green, N.L. Wilcox, and R.D. Curren. 1993. Framework for validation and implementation of in vitro toxicity tests. *In Vitro Cellular and Developmental Biology* 29A:688-692.

Home Office. 1997. *Statistics of Scientific Procedures on Living Animals in Great Britain*. London: The Stationary Office (58 pp.).

Morrison, J., R. Quinton and H. Reinert. 1968. The purpose and value of LD50 determinations. Pp. 1-17 in *Modern Trends in Toxicology* (eds. E. Boyland and R. Goulding). London: Butterworths.

OECD. 1996. Final Report of the OECD Workshop on Harmonization of Validation and Acceptance Criteria for Alternative Toxicological Test Methods ("Solna Report"). Paris: Organisation for Economic Co-operation and Development (60 pp.).

Parascandola, J.P. 1991. Historical perspectives on in vitro toxicology. Pp.87-96 in *In Vitro Toxicology: Mechanisms and New Technology* (eds. A. Goldberg and M. Principe) (*Alternative Methods in Toxicology, Volume 8*). New York: Mary Ann Liebert, Inc.

Rowan, A. 1984. *Of Mice, Models and Men.* Albany: State University of New York Press.

Trevan, J. 1927. The error of determination of toxicity. *Proceedings of the Royal Society,* Section B, 101:483-514.

Van den Heuvel, M., D.G. Clark, R.J. Fielder, et al. 1990. The international validation of a fixed-dose procedure as an alternative to the Classical LD50 Test. *Food Chemical Toxicology* 28 (7): 469-482.

Van den Heuvel, M.J., A. Dayan, and R. Shillaker. 1987. Evaluation of the BTS approach to the testing of substances and preparations for their acute toxicity. *Human Toxicology* 6:270-291.

Ward, A. 1990. Consumers at odds with animal testing. *Advertising Age* February 26.

Zbinden, G. and M. Flury-Roversi. 1981. Significance of the LD50-test for the toxicological evaluation of chemical substances. *Archives of Toxicology* 47(2):77

Glossary

Acute	Short-term, typically referring to the toxic effects exhibited up to a month after a substance is given to an animal in a single (usually large) dose.
Biologicals	Biologically active substances that are similar or identical to compounds produced in the body, such as hormones.
Carcinogenicity	The ability of an agent to cause cancer.
Chronic	Long-term, typically referring to toxic effects exhibited by animals after multiple doses of or continual exposure to an agent. These tests are generally performed for greater than three months.
Chorio-allantoic Membrane	The membrane just inside the shell of a avian egg.
Conjunctiva	The clear membrane that lines the inner surface of the eyelid and, in some species, can be extended across the front surface of the eye.
Endocrine System	A system of glands that release hormones into the blood stream.
Endpoint	An outcome in a test or experiment that constitutes a predetermined stopping point.
False Negative	Characterizing a substance as nontoxic (e.g., non-carcinogenic) when it is truly toxic (e.g., carcinogenic).
False Positive	Characterizing a substance as toxic (e.g., carcinogenic) when it is truly nontoxic (e.g., non-carcinogenic).
Hazard	A danger; hazard assessment gauges whether or not a chemical causes an adverse effect, thereby making it a potential danger.
In Vitro	Literally "in glass," referring to procedures conducted in a test tube or some artificial environment.
In Vivo	Literally "in life," referring to procedures done within the body
Iritis	Inflammation of the iris (the colored part of the eye that controls the amount of light that enters the eye).
Keratinocyte	An epidermal cell that make keratin (a protein that forms the outer layer of epidermal structures).
Mutagenicity	The ability of a substance or chemical to cause mutations (a change in genetic material).
pH	A chemical symbol referring to the acidity (low pH values) or alkalinity (high pH values) of a solution.
Phototoxicity	The ability of sunlight to trigger or increase the toxicity of a substance.

Potency	The ability of a substance to cause strong chemical effects.
Prediction Model	As part of the validation process for an alternative test, such a model would specify, in advance, how the alternative data will be used to predict animal data.
Pre-validation	The process during which an alternative test protocol is improved and standardized, prior to validation; also known a "test optimization."
Risk Assessment	An assessment of a substance's potential to injure people, taking into account the substance's ability to cause toxic effects ("hazard"), the dose at which those effects occur, and the likely exposure of people to the substance.
Subchronic	Medium-term, referring to tests in which an animal is exposed to a substance for generally one to three months, with any toxic effects determined during that time period.
Teratogenicity	The ability of a substance to cause fetal malformation.
Test Optimization	See Pre-validation
Toxicity	The state of being toxic or poisonous.
Validation	The process by which the results of a test or technique are shown to be reliable and relevant for a particular purpose.

Tables

Table 1. Some Animal-Based Toxicity Tests	
Test	**Description**
Carcinogenicity	Measures the tumor/cancer cell-producing potency of an agent by administering the agent and observing either quantity of tumors or cell transformation
Corrositivity	Assesses the corrosive effects of a substance to the skin
Developmental Toxicity	Assesses the toxic effects of a substance to an organism throughout development and growth
Draize Eye Irritancy	Assesses the toxic effects of a test substance to the eye
Draize Skin Irritancy	Assesses the irritancy of a test substance to the skin
Functional Observational Battery	A group of noninvasive tests done to evaluate dysfunction caused to animals exposed to agents
Hershberger Assay	Assesses the ability of a chemical to stimulate or inhibit androgenic responses in testes and secondary sex organs
LD50	Assesses the ability of a substance to cause death and other adverse effects with a single dose
Reproductive Toxicity	Assesses the toxic effects of a substance on the reproductive capabilities of an organism
Teratogenicity	Assesses the effects of a chemical on the developing fetus
Uterotrophic Assay	Assesses the ability of a chemical to stimulate or inhibit estrogenic responses of the uterus

Table 2. Company Announcements Regarding Animal Testing	
Company	**Announcements**
Avon	1989: announced a permanent end to animal testing. (They were also the first major company to discontinue use of the Draize Eye Test. However, they did continue to sell cosmetics that contain ingredients that had been tested on animals.)
Colgate-Palmolive	1984: announced a reduction in lab animal use by more than 50%. 1988: sponsored a postdoctoral fellowship in in vitro toxicology. 1989: announced collaboration with Marrow-Tech, Inc. to co-develop alternatives to product testing on animals. 1999: declared a moratorium on the use of animals for testing the safety of its adult personal-care products.
Gillette	1995: launched a program to contribute $100,000 annually toward research on alternative testing methods, in collaboration with The

	Humane Society of the United States. 1996: announced a moratorium on animal testing.
L'Oreal	1993: announced a halt to cosmetic testing on animals (included final products only and did not include tests on pharmaceuticals or ingredients.
Mary Kay	1989: announced a moratorium on animal testing. 1999: adopted the Corporate Standard of Compassion for Animals which indicates that their products and ingredients undergo no animal testing during the process of manufacturing.
Proctor & Gamble	1989: established the University Animal Alternatives Research Program and provided three-year $50,000-per-year grants. 1989: contributed $4.5 million to the development of alternatives. 1991: announced a collaboration with Marrow-Tech, Inc. to develop a system for screening new materials used in the treatment of oral diseases. 1992: P & G and Advanced Tissues Sciences, Inc. announced P & G's development of a new cultured human tissue test which they would use for ocular testing. 1993: lobbied in support of the 1993 National Institutes of Health Revitalization Act, which led to the formation of the Interagency Coordinating Committee for the Validation of Alternative Methods (ICCVAM). 1999: announced that it would discontinue the use of animal tests for its current lines of beauty, fabric and home care, and paper products, except where required by law.
Revlon	1989: announced an end to all phases of animal testing, becoming the first major company to do so. (In 1980, they gave a $750,000 grant to Rockefeller University for a project aimed at finding an alternative to the Draize Test).

Table 3. Timeline: Draize Test & Cosmetic Testing

Date	Event
1944	John Draize develops scoring system to standardize eye irritation testing
1971	Weil and Scala publish paper showing tremendous variability in Draize results, concluding Draize test not valid without major changes
1980	Henry Spira launches the Coalition Against Rabbit Blinding Tests
1980	The Center for Alternatives to Animal Testing is established at Johns Hopkins University
1991	The Interagency Regulatory Alternatives Group convenes a Draize workshop in Washington, D.C.
1993	European Union Cosmetics Directive calls for an end to the marketing of cosmetics tested on animals unless alternative methods are not developed
1999	United Kingdom cancels all licenses for cosmetic testing on animals,

	initiating a de facto ban

Table 4. Timeline: LD50 Test

Date	Event
1927	Trevan publishes paper describing the LD50 Test
1976	United Kingdom Home Office rules LD50 necessary
1981	Organization for Economic Cooperation and Development (OECD) issues Guideline 401 (traditional LD50 Test)
1981	Sweden holds a symposium entitled "LD50 and possible alternatives"
1981	Zbinden & Flury-Roversi publish their LD50 critique
1983	FDA publishes statement that it does not need LD50 data
1987	OECD revises Guideline 401 to reduce animal numbers in the LD50 test
1992	OECD adopts the Fixed Dose Procedure (Guideline 420) as an alternative to the LD50 test
1996	OECD adopts the Acute Toxic Class Method (Guideline 423) as an alternative to the LD50 test
1998	OECD adopts the Up and Down Procedure (Guideline 425) as an alternative to the LD50 test
1999	OECD agrees in principle to drop Guideline 401 (the LD50 Test) from its guidelines. The UK government halts its practice of issuing licenses to laboratories for the use of the traditional LD50 test
2000	OECD scheduled to drop Guideline 401

Table 5. Timeline: Validation and Regulatory Acceptance

Date	Event
1981	Johns Hopkins University establishes the Center for Alternatives to Animal Testing (CAAT)
1990	OECD issues monograph on validation (Frazier, 1990)
1990	"Amden I" report on validation (Balls et al., 1990) issued
1993	European Union establishes the European Centre for the Validation of Alternative Methods
1993	National Institutes of Health Revitalization Act calls for federal government to identify key features of the validation and regulatory acceptance of alternative methods and to establish a process for the regulatory acceptance of these methods
1994	US government establishes the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM)

1995	"Amden II" report on validation (Balls et al., 1995b) issued
1996	OECD issues "Solna" report on validation (OECD, 1996)
1997	ICCVAM publishes its landmark report, Validation and Regulatory Acceptance of Toxicological Methods
1998	ICCVAM establishes a support center, the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods
1998	ICCVAM endorses Local Lymph Node Assay for allergic contact dermatitis testing
1998	ECVAM approves the 3T3 Neutral Red Uptake Test for phototoxicity testing and the Transcutaneous Electrical Resistance Assay and Episkin and similar assays for skin corrosivity testing
1999	ICCVAM endorses Corrositex® for skin corrosivity testing
2000	ICCVAM hosts the International Workshop on In Vitro Methods for Assessing Acute Systemic Toxicity