rather than spontaneous diseases. Veterinarian Michael Fox, in recounting Schwabe’s view, calls for greater collaboration between veterinary and medical researchers.  

Computer-Aided Drug Design: discovering new drugs is largely a trial-and-error process, costly in terms of time, money, and animals. It takes eight years, on average, to screen a new substance from the seven thousand to eight thousand novel compounds created each year and to bring it into medical practice. Fortunately, methods are being developed to replace this shotgun approach with the more directed approach of computer-aided drug design. Three-dimensional computer graphics and the theoretical field of quantum pharmacology are being used in efforts to design drugs with particular specifications. These efforts are based on the lock-and-key mechanism of drug action; that is, drugs must be the right shape and composition in order to “dock” with their targets and trigger their effects. Color graphics help visualize this process. Although computer-aided drug design is in its infancy and is highly theoretical, there are indications that progress is being made. Several “drug designers” have been included on new drug patents for aid in discovering drugs. A new drug being tested clinically for effectiveness against high blood pressure was designed with computer methods. Perhaps it is not surprising, then, that several pharmaceutical companies now employ such “drug designers.” Much of the work in computer-aided drug design is apparently being conducted in Britain, where it has received some financial support from the Lord Dowding Fund. However, researchers at the University of Pittsburgh are collaborating with British researchers in attempts to use computer-aided methods to design a drug to treat sickle-cell anemia. New efforts such as these hold great promise for reducing animal use by revolutionizing the process of drug discovery.

Case Studies: The LD50 Test And the Draize Test

MUCH OF THE public outcry against the use of animals in toxicity tests has centered on the LD50 test and the Draize test. It is not surprising, therefore, that much of the research into developing alternatives in toxicity testing has been directed at these two tests. Substantial progress in this research has been made during the last five years.

The LD50 Test

The LD50 test was developed in 1927 to standardize the potency of potentially poisonous substances destined for human use, such as diphtheria toxin, digitalis extract, and insulin. Although not originally designed to do so, the test gradually became incorporated into routine toxicity programs for testing new chemicals. Government regulations in the U.S. and abroad specified the LD50 test for evaluating new drugs, food additives, cosmetics, household products, industrial chemicals, and pesticides. Each year in the U.S., four to five million rats, mice, guinea pigs, and, less frequently, rabbits, dogs, and primates, are subjected to this test.

In the LD50 test, test substances are force-fed, inhaled, injected, or applied to the skin of animals. Of these variants—the oral, inhalation, injectable, and dermal LD50 tests, respectively—the oral LD50 is the most common. It produces signs of poisoning including bleeding from the eyes, nose, or mouth; labored breathing; convulsions; tremors; paralysis; and coma.

The classical LD50 test uses large numbers of animals to derive a numerical index of toxicity (the LD50 value). This approach has two major scientific problems. First, the test is of limited value in protecting human health. This limitation stems primarily from an overemphasis on the LD50 value. Sometimes, little or no additional information (such as poison symptoms, body organs affected, and specific cause of death) is gathered. This important information could be derived from relatively few animals. According to D.V.W. Parke, the “counting of cadavers” should be replaced by full clinical and postmortem studies using fewer animals.

Even when the LD50 value is supplemented with clinical and pathological information, public health officials can still be at a loss to infer the maximum safe dose of the test substance in humans. The LD50 provides the median lethal dose, not the safe dose. Moreover, the lethal dose, as ill-suited a measure as it is, still has...
Calculated 1D50 values for the same chemical can vary substantially among laboratories (inter-laboratory variation) and among laboratory animal species (inter-species variation). For example, a study of inter-laboratory variation was recently recommended by the British Toxicological Society.

A second problem with the classical LD50 test is its unnecessary precision. Large numbers of animals are used to derive a precise estimate of LD50, yet that estimate can be applied to humans in a rough manner only. According to Rowan, "If the LD50 figure of a compound is 100 milligrams per kilogram of body weight for a mouse, it could easily be anywhere between 10 and 10,000 milligrams per kilogram body weight for a human being."

The 1D50 figure of a compound is 100 milligrams per kilogram of body weight for a human being. According to toxicologist G. Zbinden of the University of Zurich:

"The marked species differences in acute toxicity are well recognized, making it impossible to predict the human lethal dose from the results of animal experiments. With such enormous variations between species, it is clear that the knowledge of the LD50 in a mouse or a rat does not provide much support for the prognosis in a human case of acute poisoning."1

A second problem with the classical LD50 test is its unnecessary precision. Large numbers of animals are used to derive a precise estimate of LD50, yet that estimate can be applied to humans in a rough manner only. According to Rowan, "If the LD50 figure of a compound is 100 milligrams per kilogram of body weight for a mouse, it could easily be anywhere between 10 and 10,000 milligrams per kilogram body weight for a human being."

The precision of the classical LD50 test is also called into question by regulatory practice. Most LD50 testing is conducted according to regulatory guidelines. Ironically, the same guidelines that call for precision usually specify that LD50 values are to be lumped in limited numbers of broad categories for labeling purposes. Thus, all of that precision, gained at such cost in animal suffering, is lost in categorization!

Some of these alternatives are modifications that would require fewer animals:

- One test uses six to ten animals to determine the Approximate Lethal Dose (ALD). This test was discussed earlier as an illustration of reduction alternatives in toxicity testing.

- The Limit test involves giving a small group of animals (ten to twenty) a single dose of a test substance. If no ill effects are seen, no further testing at higher doses is required. The accepted maximum dose for the test depends on the nature of the substance being tested. The Limit test is especially useful for relatively harmless substances, which would necessitate unrealistically large doses in the classical test.

- In the Up-and-Down test, each animal receives a single dose, but that dose changes as the six or so animals are sequentially tested. The dose is lowered after signs of severe toxicity develop or is raised after an animal survives one week without such signs. The resulting information is evaluated in a commonly available computer program. The test yields a reasonable estimate of the LD50.

- Other techniques also use fewer animals than the classical LD50 and yet yield LD50 figures of satisfactory precision. These include the "moving averages" technique and a graphical method suggested by Molinengo.

These modifications of the LD50 test use substantially fewer animals than the classical test and, therefore, qualify as reduction alternatives. The Limit test, by its very nature, is also a refinement alternative in that it reduces exposure of animals to pain-inducing doses. All of these modifications could qualify as refinements if animals that were acutely suffering and dying were instead painlessly killed and counted among those that died or exhibited severe toxic reactions. This refinement was recently recommended by the British Toxicological Society.

Many toxicologists who conduct acute toxicity tests such as the LD50 make at least some use of these alternatives, especially the Limit test. The U.S. cosmetics industry substituted the Limit test for the classical LD50 test and thereby reduced animal use by seventy-five to ninety percent, according to a trade association survey. And Allied Corporation has abandoned the classical LD50 test in favor of...
the "Up-and-Down" test. Animal use was cut in half. This innovation and others are yielding more information, cutting costs, and reducing the stress of those animals that are used.14 These examples may represent just the tip of the iceberg: the consensus among participants at the second symposium of the Center for Alternatives to Animal Testing was that reduction alternatives could completely replace the classical LD50 test.15

Alternatives to the classical LD50 are not limited to modifications of the test itself. They also include mathematical models, in vitro techniques, and alternative species.

Mathematical models are being developed to predict LD50 values without using animals. The most promising of these models is that of Kurt Enslein and his colleagues at Health Designs, Inc. The model predicts the oral LD50 values for rats, based solely on a chemical's structure and properties. The model, created through an analysis of nearly two thousand chemicals that had already been tested in rats, was evaluated by generating predictions on the LD50 values of another 900 compounds that had already been tested. The predicted values were similar to the actual values obtained in animals.

The researchers concluded that their model could be used competitively with the rat LD50 test. It has many advantages: elimination of unnecessary animal testing, lower cost, faster response, and greater repeatability. K. Enslein suggested a number of applications: (1) estimating the doses to be used in animal-based LD50 tests (this application could spare animals from being tested at doses that are too small or large to be meaningful), (2) selecting least toxic compounds by obtaining estimated LD50s on similar compounds before they are synthesized, then ranking these estimates to decide which compounds to investigate further (this application could spare animals from being tested with highly toxic substances); and (3) supplying data for any acute toxicity studies as needed.16

The major limitation of the model is that it cannot, as yet, generate estimated LD50 values for all compounds, owing to technical problems. Enslein and his collaborators have discussed this and other limitations of their model, adequately addressed their critics, and discussed future plans to improve the model and render it more understandable to toxicologists.17 This latter development will hasten the development and possible application.

The model is likely to be used initially as a preliminary screen, backed up by animal testing. During this period, the model could be improved. If it then inspires confidence, it may totally replace LD50 testing in animals.

Cell-culture alternatives to the LD50 test are being developed by a research program coordinated by FRAME.18 The program involves four laboratories in the United Kingdom and is financially supported by numerous commercial and nonprofit organizations. The aim of the program is to develop a tier approach to acute toxicity testing:

Level 1: In vitro testing for gross toxic effects on fundamental properties of cultural cells,

Level 2: In vitro testing for specific toxic effects on particular target organ cells, and

Level 3: In vivo testing, if necessary.

Work on this program is in progress. Preliminary results on Level 1 are encouraging. The fundamental property being examined is protein synthesis by rat cells. In this procedure, toxic test chemicals are added to these cells inhibit protein synthesis. The test yields an LD50 value, the dose causing fifty percent inhibition of protein accumulation. LD50 values are well correlated with in vitro LD50 values. Although cell-culture tests such as this one may never completely replace in vivo testing, their judicious use clearly has great potential.19

Another potential alternative to the LD50 test involves the use of less sentient organisms. Using a series of alcohol compounds as test chemicals, researchers recently obtained an excellent correlation between LD50 values in mice and inhibition of movement in tubifex worms.20 These findings need to be extended through the testing of other compounds that have already been tested on animals but not yet tested on worms.

Given the inadequacies of, and the alternatives to, the classical LD50 test, it is not surprising that support for the test is eroding in all quarters. Even toxicologists have criticized it. Dr. S.B. deC. Baker stated that acute studies such as the classical LD50 "are of little use and are expensive in animals. The main information they give is an indication of the… dose required to commit suicide."21 Zbinden called the LD50 "a ritual mass execution of animals."22 Dr. D.P. Rall, director of the United States-based National Toxicology Program, called the LD50 "an anachronism. I do not think the LD50 test provides much useful information about the health hazards to humans from chemicals."23 The Pharmaceutical Manufacturers Association, which represents 149 research-based pharmaceutical companies in the United States, stated that "Advances in toxicity testing now make it possible to conduct most drug-safety evaluation without the Classical LD50 test."24 Even the National Society for Medical Research, which promotes and defends the use of animals in biomedical research, has backed away from the LD50. Its new position is that "The routine use of the quantitative LD50 test is not now scientifically justified …"25

Despite these statements, the classical LD50 test has not been abandoned. A 1983 survey of toxicologists who conducted acute toxicity tests revealed that eighty percent used the classical LD50 test.26 Perhaps the only scientifically legitimate use for the classical LD50 test is in rare cases in which drugs have a narrow margin of safety, so that toxic levels have to be precisely determined.27 So why does the widespread use of the classical LD50 test persist? The main reason cited by the manufacturers and testing companies that participated in the 1983 survey was to satisfy regulatory requirements.28 Of course, by satisfying regulations, these companies may feel better armed against damage claims brought by consumers. Perhaps these companies have difficulties in breaking an old habit. For their part, regulatory agencies also seem worried about consumer backlash and seem to be plagued by bureaucratic inertia.

A major regulatory obstacle for products marketed internationally is the Organization for Economic Cooperation and Development (OECD). OECD guidelines require that the LD50 test be conducted prior to international marketing of products. Companies whose products have any chance of being marketed overseas must satisfactorily demonstrate safety. For their part, regulatory agencies also seem worried about consumer backlash and seem to be plagued by bureaucratic inertia.

The Draize Eye-Irritancy Test

The Draize test is a method of assessing the eye irritancy potential of various substances including cosmetics, toiletries, household products, ophthalmic drugs, pesticides, and industrial chemicals. The test was developed following passage of the Federal Food, Drug, and Cosmetic Act of 1938, which mandated (among other things) that cosmetics be free of substances poisonous or deleterious to the user. Today, the test is a routine component of toxicology programs and regulatory evaluations worldwide. However, prospects for developing and implementing alternatives appear promising.

The Draize test is performed almost exclusively on albino rabbits. It consists of modified several times since its adoption. A fixed dose (0.1 milliliters or 0.1 ml) of the test solution is placed inside the lower lid of one eye of six to eighteen rabbits.29 The lower and upper lids are then briefly held together to distribute the test substance on the eye surface. The other eye is left unused for comparison. The rabbits are restrained during the procedure and later immobilized in stocks to prevent them from moving in tubifex worms. Using a series of alcohol compounds as test chemicals, researchers recently obtained an excellent correlation between LD50 values in mice and inhibition of movement in tubifex worms. These findings need to be extended through the testing of other compounds that have already been tested on animals but not yet tested on worms.
VIII. Rabbits immobilized in stocks as part of a Draize test

from rubbing or scratching their eyes.

The rabbits' eyes are examined at specific times after exposure to the test substance (e.g., at 1, 24, 48, 72, and 168 hours). Damage to different parts of the eye is rated on separate scales. The maximum scores for damage to the cornea, conjunctiva, and iris are eighty, twenty, and ten points, respectively. These scores are added to yield an overall score for eye injury.

Eye irritation in the Draize test usually consists of reddening and swelling of the conjunctiva and iris and clouding of the cornea. Eye damage can be readily anticipated when substances such as hydrochloric acid, formaldehyde, alcohol, industrial solvents, drain cleaner, laundry soap, dish washing compounds, and shampoos are tested. Animals that survive the test with minor injuries are sometimes used for other laboratory studies, such as skin-irritancy testing, before they are killed.

The Draize test undoubtedly has been of some help in deciding whether or not substances are safe for human use. However, as the test for preventing ocular injury to humans, it leaves much to be desired. A major problem is that the test is unreliable. In cases in which particular substances were tested several times (either in the same laboratory or in different ones), it has not been uncommon for the same substance to be classified as an irritant in some instances and as a nonirritant in others. Such differing results may have been caused by variation in the scoring of similar degrees of eye damage or the haphazard distribution of the test substance on the eyeball.

The Draize test is also crude. It yields a score that is used to determine whether or not a test substance is an irritant—virtually a pass-fail test with an arbitrary cut-off point. For many substances, the important question for protecting human health is not whether a substance is an irritant, but how much of one it is.
doses are not only a humane refinement, but are also likely to yield results of greater relevance to protecting human health.

Using weaker dilutions of a test substance has the same effect as using smaller doses. N.J. Van Abbe recommends using dilutions when a substance is likely to cause severe reactions at the routine dose. Such dilutions have the scientific advantage of enabling finer discriminations to be made from the results. According to Van Abbe, the discrimination can also be enhanced by simultaneously comparing the results with those from a reference standard.

39 Use of noninvasive techniques.

One technique used to document eye damage in the Draize test involves killing the test animals and surgically excising eye tissue. However, several noninvasive and nonlethal refinements of this procedure are available. These include measuring corneal thickness using an optical device, measuring intraocular pressure using a hand-held instrument, and measuring the corneal reflex using a taut string and a simple device. If these procedures were adopted, all animals would survive the test.

40 Use of fewer animals.

The Draize test currently calls for six to eighteen rabbits. In a study investigating the effect of the number of test animals on the test’s precision, increasing numbers from one up to six yielded marked improvements in precision. However, increasing numbers to nine or twelve yielded “little further benefit when set against the increase in animal numbers.” Hence, six animals should suffice, in most cases.

A more far-reaching reduction alternative was suggested by Koeter and van Vliet, who recommended that a preliminary Draize test be conducted with only one animal. If severe irritation resulted, testing should stop. If irritation were less than severe, a few more animals could be tested, as necessary.

One can readily imagine other reduction alternatives to the Draize test. Dr. G. Flamm of the Food and Drug Administration recently recommended that any substance found to be an irritant at a low dosage should not be tested on more animals at higher doses.

It is unlikely that the Draize test could be refined to the point where all pain and stress were excluded. Even if anesthesia and weak dilutions of test chemicals were used, the rabbits would still be living in stockades, which is undoubtedly stressful. Indeed, “rabbits not infrequently break their backs as a result of struggling to escape” from the Draize stocks. This brings us to consider alternatives that would replace, or at least reduce the demand for, the Draize test.

The most widely known alternative replaces rabbits with chicken eggs. A portion of the eggshell and adhering membranes is removed from a fertilized egg when the embryo has developed for two weeks. This procedure exposes the “chorioallantoic membrane” (CAM), which surrounds the embryo. A small amount of a potential irritant is applied to a section of the CAM. A positive response can include cloudiness, inflammation, and proliferation of blood vessels, but since the CAM has no demonstrable nerve supply, the embryo feels no pain.

Initial results from the CAM test show a good correlation with results from the Draize test. This promising alternative is now in the “validation stage” and is being funded by various animal-welfare organizations in the United States.

Many potential alternatives to the Draize test involve in vitro systems. One is an organ culture of isolated eyes. Eyes can be obtained from human or animal cadavers, especially from sources such as slaughterhouses. An example of this organ-culture method is the Enucleated Rabbit Eye test, which has yielded promising results. The strengths and limitations of this type of test have been discussed by D.W. Swanson and M. York.

Another organ-culture system involves isolated corneas, as distinct from entire eyes. Rabbit or bovine corneas are incubated with suspected irritants. Irritancy is inferred from changes in corneal thickness, ratio of wet weight to dry weight, microscopic anatomy, and corneal enzymes.

Other potential in vitro alternatives to the Draize test involve cell-culture systems. The cells for these tests are derived from a variety of sources, including the human cornea, mouth lining, and blood (leucocytes); rabbit cornea; rat abdominal cavity; and mouse embryo and connective tissue. In these tests, chemical irritancy is inferred from a variety of end points, including cell death, cellular release of substances associated with irritation, cell membrane damage, changes in cell
movement or metabolism, and the rate of wound healing. Examples of these cell-culture alternatives include the Rat Mast Cell assay, the Fluorescent Diacetate test, the Haemolytic Activity test, and a variety of as-yet-unnamed tests.10

The Rat Mast Cell assay is already in limited use by the Johnson and Johnson Company. Mast cells are derived from connective tissue and are involved in inflammatory responses. The assay monitors the cells’ release of the chemical serotonin. J. McCormack reported that “the procedure has a high degree of correlation with in vivo test results. In addition, it is easy to perform, accurate, and repeatable, and it limits the scope of in vivo testing.”11 The assay is used as a screen to eliminate severe irritants from in vivo testing. However, only a single class of compounds was evaluated, so it remains to be seen whether the Rat Mast Cell assay has wider applicability as a substitute for the Draize test.

Cell-culture tests funded by Revlon are producing encouraging results. Two such tests monitor either anatomical changes in cells or inhibition of cellular uptake of an important chemical constituent.12 Rockefeller University researchers obtained excellent correlations between the results of the two tests and the Draize test. Perhaps the most promising cell-culture alternative is the Draize test, which is being developed at the Eye Research Institute (Boston) and Harvard Medical School. The test is based on the observation that when the surface of a rabbit or human eye receives a minor injury, healthy cells migrate over the wound and proliferate to heal it. Irritants chemicals slow this healing process. To investigate this inhibitory effect, researchers injure two types of rabbit corneal cells in vitro, normal cells and those treated with an irritant. The rate of wound healing is measured by staining the wound and using time-lapse photography. The degree to which a substance slows the response is an indicator of the substance’s toxicity.

Dr. A. Neufeld, one of the developers of this test, recently commented, “Not only is the Draize test a poor way to treat animals, but the in vitro method appears to be far more sensitive and far more relevant.”13 Preliminary results using human cells suggest that the methods developed for rabbit cells can be successfully applied to human cells.

A promising tissue-culture alternative to the Draize test utilizes excised strips of rabbit intestine. Some sixteen pieces can be isolated from a single animal. When suitably cultured, these strips will contract spontaneously for hours unless chemically poisoned. The test determines the concentration of test chemicals necessary to block fifty percent of the contractions. This test is based on the premise that some damage in the Draize test occurs when chemicals penetrate cells on the eye’s surface and damage cells at lower levels. The surface cells can be viewed as a penetration barrier to chemicals, intestinal cells mimic this barrier effect. The results of this test have compared very favorably with in vivo data. One rabbit could provide enough material for thirty experiments, and the technique could be used for two-thirds of the Draize tests currently performed.4

The outlook for major changes in routine eye-irritancy testing is bright. Research into alternatives to the current Draize test is active and varied, thanks largely to public outcry over the treatment of animals in this test. Refinements and reduction alternatives to the Draize can be implemented immediately. Implementation of replacement alternatives probably will be gradual, as alternatives are incorporated into a suitable battery of tests.

While the Draize test is still in use, alternatives can be used in a supplementary manner to screen out highly irritating substances and to determine doses that will yield mild reactions in the Draize test. Before considering any form of eye-irritancy testing, investigators should ask whether a particular substance needs to be tested at all. Certain substances need not be tested because they are almost certain to cause eye irritation. These include substances that are highly acidic or alkaline and those that are already known to be severe skin irritants.13

As with the LD50 test, efforts to make eye-irritancy testing more humane should be directed to government regulators as well as to product manufacturers and their testing laboratories. Such efforts should focus on eliminating variables in regulations that result in unnecessary replication and on circumventing bureaucratic inertia to accepting proven alternatives.

Recall that one reason for the persistence of the classical LD50 test is the claim that it is necessary to protect manufacturers against untoward legal action. A similar claim has been made with regard to the Draize test. However, it has been made in one legal action taken against a manufacturer of a shampoo that damaged someone’s eye, rabbit–eye testing was a minor part of the case. This case (United States v. An Article of Cosmetic...Beacon Castle Shampoo...) merits discussion here because both supporters and opponents of the Draize test claim that it supports their arguments.

The case was a civil suit brought by the Food and Drug Administration (FDA) against the manufacturer in the wake of an eye injury sustained by a young girl. The girl dropped a container of shampoo, and the contents splashed up in her eye. To support its case that the shampoo was dangerous and therefore should not have been marketed, FDA commissioned a Draize test in rabbits and a study of human volunteers.

Despite the fact that the shampoo injured the rabbits’ eyes, the court ruled against FDA and in favor of the manufacturer. In a discussion of this decision, the General Accounting Office emphasized that FDA failed to show that “the results of test on rabbit eyes can be extrapolated to humans.” This statement is significant because it apparently undermined the manufacturer’s defense—against—liability argument for conducting the Draize test.

Unfortunately, the issue of extrapolating from rabbits to humans was not the keystone of the judge’s decision.15 The primary reason for the ruling was that the FDA failed to show that the full concentrate of shampoo might get into the user’s eye under the usual conditions of use and that the user would not automatically flush out the eye.16

Nevertheless, the judge did state that the “rabbit studies, standing alone, do not warrant condemnation of this product.” The judge refused to accept extrapolations from rabbit–test results to human response without confirming data from research on human volunteers. In this case, FDA submitted conflicting and incomplete results of human studies. Although a complicated and multi-faceted case, the Beacon Castle decision does provide evidence that a court did not find rabbit–eye testing particularly helpful in determining the extent of human hazard.

Given the judge’s comments, it is rather surprising that a spokesperson for The Cosmetic, Toiletry, and Fragrance Association (CTFA), a manufacturer’s trade group, asserted that the case “provides support for use of the Draize test as a reliable method of substantiating that a product is safe for eye-area use.”16 According to the CTFA, the judge reasoned that rabbit eyes are more sensitive than human eyes in that the former have less capacity to tear and flush away an irritant; therefore, any chemical that does not injure rabbits’ eyes is not likely to injure human eyes.

The CTFA supersensitivity—as—an—asert argument is unsupported not only by the Beacon Castle case, but also by toxicological principles. Although supersensitive species are well-suited for confidently identifying harmless substances, the strength of a toxicity test should be its ability to detect harmful substances. If a test is supersensitive, it will overclassify substances as harmful. This would be the toxicological equivalent of “crying wolf.” The test’s results could easily be explained away, much as studies identifying cancer-causing substances in laboratory animals sometimes are dismissed because the huge doses utilized may cause cancer by overwhelming the body’s metabolism. Use of a less sensitive species or system could be more valuable in protecting human health, as well as more humane.