

Evading Causal Disanalogy

It Just Works

The arguments so far have shown that animal models are often not good CAMs of human biomedical phenomena. Since the proffered rationale for doing biomedical experiments on animals is that they uncover significant causal information about humans, this is a surprising and unfortunate conclusion.

However, as we noted earlier, there is a gap between the stated and the real rationale for animal experimentation. Any number of researchers think animal models need not be either strong or weak CAMs to be scientifically valuable. In fact, some would claim that questions about causal isomorphism, causal disanalogy, etc., are just red herrings, ways of diverting attention from the demonstrated success of animal experimentation. On this view we “just know” animal experimentation works.

"IT JUST WORKS"

Some researchers contend that we know by experience that animals are good models of human biomedical phenomena -- even if we don't understand why. All we need know is that the model serves some particular scientific purpose. That, researchers say, we can and often do know.

Katz (1981) suggests a taxonomy based on the purpose or use of the model rather than the degree of isomorphism between the animal model and the human situation. They are used because they are useful -- they work! He offers two model types. The *empirical/utilitarian* model (e.g., drug screening) is useful when a theoretical rationale may be offered, but is not needed. *Theoretical* models are used to test specific hypotheses

IT JUST WORKS

about etiology, mechanism, and so forth. Most models are not pure representatives of either type but have elements of both (Nooneman and Woodruff 1994: 9).

Some defenders of animal research straightforwardly acknowledge that there are clear differences between humans and animal test subjects, but claim the differences in no way undermine the value of animal experiments. For instance, when discussing tests to determine the carcinogenicity of saccharin, Giere says:

The relevance of animal studies was questioned for two reasons. First, humans are different from rats . . . Second, the amount of saccharin used (5 percent of the rats' diet) was quite large. Some critics calculated the 5 percent of the human diet corresponds to the amount of saccharin in 800 bottles of diet soda. Who drinks 800 bottles of diet soda a day. There is something to these criticisms, but not nearly as much as many critics thought . . .

As for the statement that humans are not rats, that is obviously true. But of the approximately thirty agents known definitely to cause cancer in humans, all of them cause cancer in laboratory rats -- in high doses. From this fact it does not necessarily follow that anything causing cancer in rats will also do so in humans. Again, it is difficult to justify basing practical decisions on the assumption that saccharin is an exception. And taking account of differences in dose and body weight, those fourteen cancers in ninety-four rats translate into about 1200 cases of bladder cancer in a population of 200 million people drinking less than one can of diet soda a day (1991: 232-3).

In other words, *we just know* that experiments on rodents work. That is, given our experience with rodent assays, and given that we have no particular reason to think that saccharin is a carcinogen only for rodents, then we can legitimately conclude that the current test results are relevant to humans. As it turns out, Giere's claim is misleading.

According to the International Agency for Research on Cancer (IARC) there are 26 (of 60,000) chemicals shown to be carcinogenic in humans. (The list of probable human carcinogens is somewhat longer). Giere's claim suggests rodent bioassays are a good way of determining cancer risk in humans. However, his claim

THE SCIENTIFIC ISSUES

is misleading because a test's *usefulness* is a function not just of its *sensitivity* (the proportion of human carcinogens that is carcinogenic in rats), but also its *specificity* (the proportion of human non-carcinogens that is non-carcinogenic in rats).

In the standard bioassay for carcinogenesis, researchers expose groups of rats and mice (usually about 50) to maximum tolerated doses of suspect substances for their entire lives. They then examine pathological manifestations in necropsied animals and compare them with control populations. The National Cancer Institute/National Toxicology Program has used this assay to test hundreds of substances. Private corporations have also used these assays to test an unknown number of substances to satisfy FDA or EPA requirements (Salsburg, 1983: 63). However, this standard carcinogenesis assay was never validated before use. As Salsburg notes:

Common scientific prudence would suggest that this assay be tried on a group of known human carcinogens and on a group of supposedly innocuous substances . . . before we either (1) believe that it provides some protection for society (sensitivity) or (2) believe it identifies mainly harmful substances (specificity). There is no substitute for such proper validation on any new bioassay. However lacking proper validation prior to its use, we might be able to examine the validity of the assay using the results of 200 or more compounds subjected so far to the bioassay . . . (1983: 63).

The results are not particularly encouraging -- even when considering the list of 26 known human carcinogens.

Of the 26 known carcinogens, humans are exposed to seven of them by inhalation. Salsburg comments on the sensitivity of these rodent bioassays as follows:

Most of these compounds have been shown to cause cancer in some animal model. However, many of the successful animal models involve the production of injection site sar-

comas or the use of species other than mice or rats. If we restrict attention to long-term feeding studies with mice or rats, only seven of the 19 human non-inhalation carcinogens (36.8%) have been shown to cause cancer. If we consider long term feeding or inhalation studies and examine all 26, only 12 (46.2%) have been shown to cause cancer in rats of mice after chronic exposure by feeding or inhalation. Thus the lifetime feeding study in mice and rats appears to have less than a 50% probability of finding known human carcinogens. On the basis of probability theory, we would have been better off to toss a coin (1983: 64).

Tossing a coin might be a good idea, since the direct cost of a rodent bioassay is \$1,000,000 per chemical tested (Lave, *et al.*, 1988: 631).

If rodent studies are not particularly sensitive, are they at least specific? As we noted in the last chapter, specificity may be as low as 0.05 (Lave *et al.*, 1988: 631) -- rodents have shown carcinogenic responses to 19 out of 20 probable human non-carcinogens. Evidence proving the limitations of animal tests for carcinogenicity has become so overwhelming that even governmental agencies are beginning to rely more on non-animal-based research methodologies (Brinkley 1993; Vainio, *et al.*, 1992: 27-39). More than a decade ago Salsburg concluded:

Presently the lifetime feeding study preempts the field. As long as it is considered to be useful in detecting human carcinogens this very expensive and time-consuming procedure will continue to drain the toxicological resources of society. This report questions its usefulness and suggests that it is time to consider alternatives (1983: 66).

However, this procedure is currently part of the governing scientific paradigm. As such, researchers tend to use it without question and often without attempts at further validation. In such circumstances, and when there are relatively few disasters from introducing new chemicals, people get the impression that the procedure "just works." Then, when there are failures, like FIAU discussed earlier, such failures are treated as aberrations.

Are there other ways of cashing out the claim that animal experimentation "just works." Are there ways of discerning the adequacy of animal experimentation without worrying whether the

model is causally analogous to the subject being modeled? Explaining what this could mean is no easy matter. However, some researchers have claimed that purely correlational models would “work” even if the models were not causally analogous with the system they supposedly model.

Parametric relations and correlational models

According to Woodruff and Baisden, animal models will be biomedically useful if they are appropriately correlated with the object they model. Thus, even when these models are causally dissimilar from the human condition they supposedly model, they may be useful, for example, when assessing drugs as *candidates for human clinical trials*. They note:

[A] particular behavior, such as activity in an open field, may be changed in a dose-related fashion by drugs that have a recognized clinical effect in psychiatric patients, but the rat behavior may not have any clear relationship to the human psychopathology. This model is not necessarily useful for the study of the cause and progression of the disease or of its pathophysiology. Rather, its validity relates to a consistent parametric relationship between the effect of a drug on this behavioral measure and the clinical efficacy of the same drug (1994: 319).

The following example helps illuminate exactly what this means:

Because of the known correlation between the clinical effectiveness of commonly used anxiolytics and their ability to inhibit seizures in rats, the dose-response curve for inhibition of pentylenetetrazol seizures in rats is a fairly good predictor of the ability of a newly proposed anxiolytic compound, and this test is widely used as a preclinical screen by pharmaceutical companies (*Ibid.*: 320).

All that matters scientifically is observed behavior. The investigator is not " . . . forced to make assumptions about the cognitive structure of the rat and to construct intervening variables so as to explain the observed behavior in human clinical terms" (*Ibid.*: 320). Moreover, the substances do not have to be chemically related: Valium and pentobarbital both inhibit seizures in rats and are both anxiolytics, but they are not chemically related.

IT JUST WORKS

Woodruff and Baisden introduce four possible criteria for validating animal models in psychopathology (1994: 320):

- [1] Similarity of inducing conditions.
- [2] Similarity of behavioral states.
- [3] Common underlying mechanisms.
- [4] The dependent variable measured in the model should react to therapeutic intervention in a way that is predictable from the effects of the same intervention when applied to humans.

Models satisfying all four conditions would be similar to CAMs as we have described them. Correlational models, however, are valid if they satisfy condition [4]. That is why we think it is best not to see correlational models even as weak models, but rather as models exemplifying the “It Just Works Argument.” Researchers using correlational models do not claim that they reveal underlying causal mechanisms. Such models are simply instrumental tools that presumably reveal biologically significant information, even if we know not how. As Woodruff and Baisden explain:

The symptoms of interest presented in the model do not have to be analogous to those of the human disease. Therefore this type of model may be created without much knowledge concerning the pathology of the endogenous disease being modeled. As suggested above . . . the most frequent use of correlative models is as screening devices for new therapeutic treatments (*Ibid.*: 321).

Drug companies use these models as screening devices to identify drugs with the desired pharmacological properties:

Newly designed drugs that produce greater effects on the animal model without significant detrimental side effects would then be likely candidates for clinical trials (*Ibid.*: 319).

In summary, Woodruff and Baisden claim animal models may be valuable even if the model and the object modeled are causal disanalogous. Are they correct? It seems unlikely. Admittedly, researchers may sometimes find a dose-related correlation between the reactions of animal subjects and humans to the same drug. However, from that we should not conclude that drugs that are relatively safe in the animal model will be likewise safe in humans, especially when the drug is not chemically related to previously

discovered anxiolytics that also inhibit seizures in rodents. The causal details are all-important.

Correlations and carcinogens.

Correlational models are also used to validate rodent carcinogenicity bioassays. Salsburg describes this procedure:

. . .it is standard practice, when setting up a bioassay, to determine the operating characteristics of the assay. To do this, the bioassay is applied to some compounds that are known to be positive and to other that are known to be negative with respect to the property sought. Error rates are then determined that describe the sensitivity and the specificity of the assay (1983:63).

Is this, however, a plausible expectation? Suppose scientists knew that rats and humans responded in the same ways to previously tested chemicals 80% of the time. Could we then safely infer that if some new chemical were carcinogenic in rats, then there would be an 80% likelihood that the same chemical would be carcinogenic in humans? Such an inference would be plausible only if we had reason to think that the sample class tested represented all carcinogens. Under what conditions would that be a reasonable assumption?

Suppose, for the sake of argument, that twenty-four of the thirty chemicals, (mentioned by Giere), that are both human and rodent carcinogens, are members of a given chemical class -- perhaps they are aromatic amines, for example. Under these circumstances, the most these previous tests could establish is that rats and humans react similarly to this class of chemicals. That is, those findings might suggest that if a previously untested chemical of the same class was discovered to be carcinogenic in rats, then it would likely also be carcinogenic in humans. Perhaps, though, these findings wouldn't even show that. Suppose the new chemical was a member of a class of chemicals, several of which had been previously tested in rats and humans, and had been found carcinogenic in both species. The inference from rats to humans would be of predictive value (though hardly a 100% guarantee) only if the metabolic (causal) features of both rats and humans that led them to react similarly to other members of this class of chemicals, are the same features involved in the metabolism of the new

chemical. Otherwise the *questions of relevance* (both ontological and epistemological) raised in chapter two again rear their ugly heads.

In fact, we needn't rely on this theoretical argument. We have overwhelming empirical evidence that even if the drugs tested were of the same class, we could not assume that it would react in the same way as other members of that class. First, we know that rodents do not respond similarly to all members of a given chemical class. In tests of 65 distinct aromatic amines, rats developed tumors in response to only 35 of those chemicals. In tests of 34 nitro aromatics and heterocycles, rats developed tumors in response to 17. For 18 Azo compounds, they developed tumors in response to 10 of them (Gold, *et al.*, 1989: 214). In short, even if we know that rats develop tumors when exposed to some members of a chemical class, we do not know that they will respond similarly to all members of that class.

We cannot even assume that animals of a given species will respond similarly to substances with virtually identical chemical structures. For example, benzopyrene-(a) causes cancer at several distinct target sites of several species; whereas benzopyrene-(e) is not carcinogenic at all, although the only difference between the chemicals is the arrangement of their respective benzene rings. Phenobarbital and sodium barbital are liver-carcinogens in rats and mice, while closely related amobarbital and barbituric acid are not. The phorbol ester, TPA, turns out to be a mouse skin-carcinogen, whereas its analog, phorbol, is not carcinogenic anywhere. Finally, one by-product of the manufacture of TNT, 2,4 dinitrotoluene, is a liver carcinogen in rats and mice, whereas 2,6 dinitrotoluene is not. (We are indebted to Dr. Lynn Willis of the Department of Pharmacology and Toxicology at the Indiana University Medical School for these examples).

We find a similar phenomenon in developmental toxins such as thalidomide. Thalidomide is teratogenic in several species of primate, and certain strains of rabbit. Yet as Schardein notes:

Several other thalidomide analogues, including WU-334, WU-338, and WU-420 had no teratogenic activity in primates, while a number of substituted isoindolines, quinazolines, and benzisothiazolines were not teratogenic in the rabbit (1985:233).

In short, some chemicals that are structurally related to thalidomide do not induce the same effects.

THE SCIENTIFIC ISSUES

Of course, were the chemical not a member of the class in question, we would have no reason to assume that the same causal mechanisms were involved in metabolism (e.g., the same cytochromes P-450); so we would have no reason to think we could legitimately extrapolate the test findings from rats to humans. Findings in test animals will be relevant to humans only if they share relevant metabolic mechanisms, and there are no significant causal disanalogies. If we don't know their mechanisms are similar, then we cannot rationally extrapolate findings in past cases to future cases.

Of course, we have not identified the carcinogenic potential of many chemicals, nor do we know many mechanisms that produce cancer. What we need to know, if general predictive inferences from rats to humans are to be strong, is that the previously tested chemicals are representative of all chemicals to be tested. But think for a moment about what it means to say that the sample class of carcinogens represents all carcinogens. It is to say that the chemicals *cause* similar *effects* in rats and humans. In short, there is no way to avoid it: researchers cannot do without causal knowledge of these biological systems.

Yet, according to Lave, *et al.*, the standard rodent carcinogenesis bioassay, does not uncover causal mechanisms:

For almost all of the chemicals tested to date, rodent bioassays have not been cost effective. They give limited and uncertain information on carcinogenicity, generally give no indication of mechanism of action, and require years to complete (988: 633).

Hence, mere correlational knowledge will not suffice. We have no reason to think that the mechanisms of action in the animal model are causally similar to the mechanisms in humans simply because we find some mathematical correlation between the behavior of these two systems.

Finally, some commentators have claimed that in rodent bioassays for carcinogenicity, the relevant correlations are the results of regulators' fiat and not empirical evidence. As Lave, *et al.*, argue:

. . . attributes of carcinogenicity such as potency, route of administration, site of tumour, histopathology, strength of evidence, pharmacokinetics and extent of malignancy are ignored in our analysis. This positive-negative simplification and the assump-

tion that any chemical carcinogenic in mammals is carcinogenic in humans, do not reflect the growing sophistication of current mechanistic research in health assessments. But neither the regulatory agencies in the United States nor the IARC use such data in their decisions . . . (1988: 631).

To relate this discussion to the epistemological problem of relevance, we are justified in thinking the systems are causally similar only to the extent that we have detailed knowledge of the conditions and mechanisms of metabolism in *both* humans and animal models. Yet toxicologists rarely have this knowledge. In fact, the very purpose of toxicological screening programs is to determine safety without this detailed knowledge.

This problem of relevance creates another dilemma for animal researchers. If we did know (or were reasonably confident) that non-human animals were causally similar to humans, and thus, that inferences from one to the other had a high likelihood of truth, then we would already have to know a great deal about the mechanisms of human disease -- the very mechanisms the non-human CAM are designed to reveal. That is, the very evidence that would justify the belief that animal models are strong CAMs of human systems would be the very evidence that would diminish its usefulness.

History shows "it works": evidence of historical benefits.

A somewhat different, albeit related, instrumentalist argument goes like this: surely we just know, from (a) surveys of primary research literature; and from (b) histories of medicine, that the general practice of animal research is a powerful source of biomedically significant information about humans. Often this justification of animal experimentation lurks in the background of all justifications of the practice. Consider, for example, the *prima facie* case for research summarized in chapter 1. That case, as exemplified in the *AMA White Paper* and the *Sigma Xi Statement of the Use of Animals in Research* rests on claims about the specific historical benefits of animal research. Similar claims can be

found elsewhere in the research literature (Smith and Boyd 1991: 25-29; Leader and Stark 1987: 470-4).

The first thing to notice is that this response, even if defensible, does not show that animal models are good CAMs of human conditions. Even if merely reciting historical episodes did show that animal research had been valuable, it would not show that animal models were good CAMs. That is, even if the primary literature did reveal that animal experiments were a vital source of information about humans, the advocates recounting of them in public policy documents would not enable us to extract the historical role that animal CAMs played from the roles that other uses of animals might have played.

Moreover, surveys of primary research literature are not an effective way of determining the success of animal experimentation. For although it is likely that such sources of information will report some failures of research, it will likely seriously underreport manifest dissimilarities between animals and humans. Sometimes underreporting of failures is intentional. There is mounting evidence that even some "giants" of biomedical science have misreported data that conflicted with their anticipated finding. Louis Pasteur, often identified as *the* paradigm of what a scientist should be, manipulated, suppressed, and even fabricated data to insure that the "evidence" supported his preconceived notions (Geison 1995).

However, we do not want to suggest that scientists are dishonest. Indeed, we needn't make such an assumption to explain why failures of science are underreported. If a researcher is trying to discover the nature of human hypertension, and conducts a series of experiments on a hamster only to discover that the animal cannot develop hypertension, then the investigator will likely not report the findings -- not because he wants to suppress relevant information, but because many other scientists just won't be interested in that information. Even when scientists do report negative findings, other scientists are less likely to read and discuss them -- especially if the results do not help explain the failure. These facts are recognized by researchers:

One of the reasons that many contributors have missed the point is that they have drawn conclusions from published data, which represent only a small sample of the many screening tests performed. Moreover, these represent a biased

IT JUST WORKS

sample because of the generally greater interest in positive results and the tendency of editors, whether of a sensational newspaper or an erudite journal, to cater to the tastes of their readers. Consequently, lessons gained from the high proportion of negative results and borderline cases that occur in practice are lost, as are also the occasional positive responses which regrettably never see the light of day, for commercial or political reasons (Palmer 1978: 216).

It is true that there is no such thing as a failed experiment in the sense that we can always learn from our failures. Nonetheless, scientists are more likely to publish their successes. Their successes are more likely to be read and discussed, if the findings are consonant with the current paradigm. Therefore, it is misleading to assess the fecundity of the practice of animal experimentation simply by tallying successes -- or even ratios of successes to failures -- in the extant research literature.

Documenting the success of animal experimentation by citing standard "histories" of biomedical research is likewise difficult. When historians of medicine discuss the history of some biomedical advance, they typically underreport failed experiments, even when those experiments appear in the primary research literature. Historians tend to report -- or at least highlight -- only those events crucial to understanding the current state of the science. Failed experiments, usually vital to the actual development of science, are often underreported, perhaps, because of their ubiquity.

This is not to question either the accuracy of these historical reports by, or the integrity of, medical historians. Careful studies of the history of medicine can be extremely instructive. The question here is not the accuracy of the facts, but how those facts are interpreted in discussions of the effectiveness and societal benefits of biomedical research. Given the human tendency to rewrite even our personal histories in light of our present beliefs (Ross 1989: 342-4), it would be surprising if medical historians did not write the history in a way that articulates their current understanding of that science. Since the use of non-human animals as CAMs is integral to the current paradigm in the biomedical sciences, we should not be surprised to find that these histories often emphasize

the apparent "successes" of the paradigm. This does not show that animal experiments have been useless, but it gives us a further reason to think that it is no simple matter to substantiate their successes. More is required than counting "successes" reported in the literature.

After all, scientists themselves often caution lay people for depending unduly on anecdotes. Most of us will, in less cautious moments, leap to conclusions based on simple anecdotal evidence – either our own or that offered by others. We may hear about someone who reports being cured of liver cancer after taking a regimen of particular vitamins and herbs. Such a report might, in some circumstances, warrant further study. However, this sort of anecdote can never prove that this medicinal mixture cured anything, let alone cancer. The belief that this anecdote constitutes evidence of the curative powers of these herbs would be, in the scientists' eyes, a modern form of alchemy that they would properly deride as decidedly unscientific.

This is not just arm-chair speculation. Defenders of research used just this argument when trying to justify legislation requiring proper regulation of food and drugs. Eventually what that meant was that all new drugs must be tested on animals before being tried in humans. In a treatise which inspired passage of the bill establishing the Food and Drug Administration in the US, Samuel Hopkins Adams identified the problems of anecdotal evidence::

The ignorant drug-taker, returning to health from some disease which he has overcome by the natural resistant powers of his body, dips his pen in gratitude and writes his testimonial. The man who dies in spite of the patent medicine -- or perhaps because of it -- doesn't bear witness to what it did for him. We see recorded only the favorable results: the unfavorable lie silent. . . Do while many of the printed testimonials are genuine enough, they represent not the average evidence, but the most glowing opinions which the nostrum vender can obtain. . . (1906: 4).

Yet when it serves their purposes, the defenders of animal experimentation often resort to just this type of anecdotal evidence to defend the biomedical *status quo*. They act as if anecdotal evidence were a scientifically respectable measure of the success of animal experimentation. However, the recitation of examples and anecdotes can never be a *measure of success*. As the quotation from

Adams suggests, those who try to justify the practice through a simple reading of the historical literature often succumb to two pitfalls. They may be duped by the *shotgun effect* and may also unintentionally commit the *fallacy of selective perception*.

Researchers may succumb to the shotgun effect when they cite their past successes as a rationale for the continuance of the practice. The practice of animal experimentation is a multi-billion dollar enterprise. Researchers conduct thousands of experiments annually. Thus, we should not be surprised to find some substantial successes, when we survey the practice over several decades. If you fire a shotgun (with thousands of pellets) in the general direction of a target, there is a good chance that several pellets will hit the target.

The researcher then commits the fallacy of selective perception if he counts the hits and ignores the misses. This fallacy is one we all are prone to commit. If there is some view to which we are antecedently committed, we often focus on evidence that supports our view and downplay evidence that conflicts with it. These tendencies are further complicated by two factors. In the *'Hard-to-Measure-Benefits' defense* of animal research, researchers claim it is difficult to judge when a pellet has missed the target, while in the *Numbers Game*, they offer artificially low estimates of the numbers of animals used in experiments, and thus skew the estimates of the ratio of hits to misses.

The 'Hard-to-Measure-Benefits' Defense.

Researchers do not try to justify their practice in just one way. They deftly move back and forth between a series of defenses of their practice. If objectors challenge, for instance, the "It just works" argument and the argument from "evidence of historical benefits," then defenders of animal research may resort to the "Hard-to-measure-benefits" argument. According to this argument, we should not ask for evidence that the practice is efficacious since in all scientific research such evidence is generally not forthcoming. However, this overstates the case. The success of a scientific practice may be difficult to measure, but that does not mean that we should not strive to ascertain its success scientifically. To the extent that researchers cannot measure the benefits of a practice, to that extent, at least, they should not claim to know that the practice is beneficial.

THE SCIENTIFIC ISSUES

For example, the authors of the *Sigma Xi* statement caution against hastily estimating the scientific importance and significance of any particular experiment involving animals:

[T]he body of scientific data generally increases by painstaking research that advances knowledge in small, incremental steps. Many such advances are usually needed to produce significant breakthroughs, and the value and importance of individual experiments are difficult to assess until the entire process has been completed. Therefore, it often is impossible to estimate the value of such experiments soon after they are finished, and thus to consider their worth in relation to any animals that may be used in the work (76).

At least in the short-term, they claim, demanding incontrovertible evidence of the significance and utility of research is inappropriate. Moreover:

[N]ot only is it difficult to predict the value of results before an experiment is performed, or even immediately afterward, but the ultimate value may be unrecognized for some time. In advance of contributions to a line of research or other applications, we cannot determine with certainty which results will have applications, what these applications may be, or when that application will arise (76).

The researchers are correct: assessing the value of research in the short-term is often difficult. Demanding that every experiment be a success would be silly. After all, by its very nature most scientific experiments fail. However, how can defenders of experimentation square their caution about judging the value of animal experimentation with their strong claims about the substantial contributions of such research to human health? Thus, we accept that we must be cautious in evaluating the success -- or the failure -- of the practice. That helps explain why we think that these public policy advocates should not make exaggerated claims about the benefits of the practice.

In summary, it is one thing to contend that scientific research slowly and incrementally contributes, in unpredictable ways, to human health. It is quite another to mislead the public with empirically unsubstantiated claims of immediate and direct benefits to humans.

The Numbers Game

We can determine the success of animal research only by measuring its benefits relative to its costs. In the previous section we explained why we often have difficulty precisely determining the benefits of animal experimentation. We likewise have difficulty judging at least one of its costs, namely, the number of animals used in those experiments.

Animal experimentation is not just science performed for its own sake. It consumes scarce health care resources. Resources spent on animal research are resources that cannot be spent on other forms of research. Thus, although some consumption of experimental subjects may be worth the benefits, a much larger use of animals may not be “worth it.”.

For instance, we determine the value of a gold mine not only by how much gold we retrieve, but by the ratio of gold to tons of ore mined. If we had boundless resources and wanted gold at any cost, then we may not be especially concerned if this ratio is quite large. If, however, we have scarce resources, we would be immensely concerned if the ratio of gold to waste was large, if we had to mine immense amounts of ore for a minor payoff. So it is crucial to determine, with some precision, the number of animals used in research.

Both sides of the debate recognize the importance of the “numbers” issue. Researchers, for example, go to some effort to explain that they don’t “waste” any of these resources. As the AMA explains it:

Research today involves intense competition for funding; for example, less than 25% of studies proposed to, and approved by, federal agencies each year are actually funded. Therefore, scientists on research evaluation committees are not likely to approve redundant or unnecessary experiments. Also, given the competition for funds, scientists are unlikely to waste valuable time and resources conducting unnecessary or duplicative experiments (1992:15).

This claim only shows, however, that researchers do not conduct (many) needless experiments. However, it still does not show that the benefits of experiments are worth the costs. We can determine that only after we know the number of animals used in experimentation. About the numbers of animals used, there is considerable disagreement. Those opposed to the practice offer

high estimates, while those in favor offer lower estimates. For reasons that will emerge below, there may be no straightforward way to settle the matter.

The AMA estimates that fewer animals are used:

The number of animals being used in biomedical research is not known. Animal activists place the figure as high as 150 million, but such estimates have no basis in any known data. An authoritative estimate was made by the Office of Technology Assessment (OTA) . . . The OTA examined data from both public and private sources and estimated that, for 1982 and 1983, the number of animals used in laboratory experiments in the United States was between 17 and 22 million (1992: 15).

Notice that even the OTA is itself uncertain about the number of animals consumed.

The AMA continues:

Surveys conducted by the Institute of Laboratory Animal Resources of the National Research Council indicate that the number of animals being used may be decreasing. In its 1978 survey, the Institute estimated the total was 20 million, a 40% reduction from the number noted in its 1968 survey (1992: 15).

However, it is difficult to accurately determine the total number of experimental animals used. Rowan *et al.*, state that "The statistics on laboratory animal numbers in the United States are crude and relatively unreliable" (1994: 1). Why can't we know exactly how many animals are used in research each year? Surely we could go to the library or some appropriate data base and find an answer. However, matters are not so simple. For instance, in 1984, Andrew Rowan estimated that the total number of animals used in research in the US was about 70 million. Yet the Institute for Laboratory Animal Resources (ILAR), quoted by the AMA, claimed only 20 million animals were used? Why the discrepancy? Rowan offers the following suggestions:

It is not clear why the ILAR survey should produce figures so much lower than other estimates, but one cue comes from the identification of the proportionate value of the research covered by the survey returns, namely \$2.2 billion. This is only 25% of the total annual expenditures for biomedical research

programs in the United States. Multiplying the ILAR 1978 survey figures by four gives a total of 80 million animals (1984: 67).

In short, estimating the actual numbers of animals consumed is tough. As Rowan notes, "Probably the best source of information on laboratory animal demand is the major commercial breeder. However, for various reasons, representatives of such companies are not particularly forthcoming on precise numbers" (1984: 69). This is not to say that evidence cannot be gathered. However, it is not a simple observational matter; we must make inferences from the evidence we do have. For instance, Rowan explains how, using information from the Charles River Laboratories -- a large breeding facility -- we might estimate the number of animals used each year.

Probably the best information on Charles River's production is found in a stock market analysis (Brown & Sons 1981). The report notes that Charles Rivers produces 22 million animals annually, more than 5 million of which are produced overseas. This would indicate that domestic (i.e., US) output is approximately 16 million animals. The report indicates that that Charles Rivers holds 20% of the total domestic market. Thus, extrapolation would indicate that about 70 million rodents are produced each year for the American market. This would not include rabbits, dogs, cats, frogs, and birds. The first three of these species probably account for about 1 million animals while the last two account for 5 to 10 million animals annually (1984: 70).

Rowan now contends that his original figures are probably off by as many as 20 million animals a year. "I now believe that my 70 million estimate may have been high (the actual total may have been around 50 million produced and 35-40 million actually used). I am also reasonably certain that animal use has declined . . . My estimate is that the decline is around 40% although it may be more (or less)" [Private correspondence, quoted with permission].

Minimally the best evidence suggests that the actual number of animals used is empirically underdetermined. Moreover, the best estimates are, in fact, somewhere between the estimates of the AMA and estimates by animal activists. As Rowan explains:

I think the AMA estimate may be low (it may be double that) but it is almost impossible to come up with accurate estimates of total use across the USA. The USDA reports and the ILAR stats only have time-series data on six species (primates, dogs, cats, rabbits, hamsters and guinea pigs) leaving out mice, rats and birds which account for 85% or more of total use in other countries. So we are still reduced to inferences and wild guesses. [Private correspondence, quoted with permission]

Since we cannot determine the numbers of animals used with any certainty, that makes it more difficult to judge the scientific success of the practice. As we will explain in chapter 15, the number's debate will be especially relevant to an assessment of utilitarian defenses of animal experimentation.

Alternate hypotheses

Some leading biologists are skeptical of the claims that modern medicine has singlehandedly caused massive increases in longevity. It is true that average life expectancies have risen dramatically over the last hundred years. Nevertheless, is this -- as researchers suggest -- the result of modern medicine prolonging the lives of the elderly and sick, or is this -- as objectors suggest -- largely the result of a massive reduction in infant mortality?

R.C. Lewontin notes that in the last century, respiratory diseases were major causes of death.

They died of tuberculosis, of diphtheria, of bronchitis, of pneumonia, and particularly among children they died of measles and the perennial killer, small pox. As the nineteenth century progressed, the death rate from all these diseases decreased continuously. Smallpox was dealt with by a medical advance, but one that could hardly be claimed by modern scientific medicine, since the smallpox vaccine was discovered in the eighteenth century . . . By the time chemical therapy was introduced for tuberculosis in the earlier part of this century, more than 90 percent of the decrease in the death rate from that disease had already occurred (1991: 43-44).

But what is the cause of falling rates of mortality? As we mentioned in the *prima facie* case against animal experimentation, some theorists have hypothesized that improvements in sanitation have

played a major role. However, there is some dispute about this. Lewontin continues:

The progressive reductions in the death rate were not a consequence, for example of modern sanitation, because the diseases that were major killers in the nineteenth century were respiratory and not waterborne. It is unclear whether simple crowding had much to do with the process, since some parts of our cities are quite as crowded as they were in the 1850s. As far as we can tell, the decrease in death rates from the infectious killers of the nineteenth century is a consequence of the general improvement in nutrition and is related to an increase in the real wage. In countries like Brazil today, infant mortality rises and falls with decreases and increases in the minimum wage (1991: 45).

There is certainly some reason to think Lewontin might be correct. Tuberculosis (often drug-resistant strains) has reappeared primarily in our inner cities, among the urban poor -- a population notoriously malnourished, and perhaps further weakened by the consequences of drug abuse.

Our best guess is that both sanitation measures *and* improved nutrition and economic well-being have been jointly responsible for the increase in life-span. For instance, the upswing in government support for biomedical research -- especially research using animals -- did not begin until after World War II. The largest increases occurred after 1970. Yet the overwhelming majority of the increase in life span occurred before even the modest increase in expenditures in the early 1950s. For instance, in 1900 an individual had less than a 66% chance of reaching the age of 40. By the 1950, when the increase in biomedical research was just beginning, an individual had a 91% chance of reaching the age of 40 -- a 60% increase in the probability that a person would reach that age. By 1991, the probability had increased to 95% -- less than a 6% increase in the probability that a person would reach the age of 40 (NCHS 1995: 13).

In short, only a relatively small increase in lifespan occurred after the big increase in biomedical research. So, whatever the cause of the decline in mortality, the cause for the decline was not solely or even primarily interventionistic medicine. In short, medicine has doubtless made some contribution to this decline. Moreover, no doubt some of this contribution from scientific medicine is derived

from research on animals. But it would be a gross mistake to claim that scientific medicine (especially that derived from animal research) was a major causal factor in this decline. The decline in death rates is a complex phenomenon with a complex cause, a cause that includes contributions from scientific medicine, to be sure, but also includes social factors.

Overall Evaluation of Applied Research

Researchers claim that non-human animals can be used as CAMs to uncover underlying causal mechanisms of human disease. We disagree. We have argued that animal tests are unreliable as tests to determine the causes and properties of human disease. Available evidence and the theory of evolution lead us to expect that evolved creatures will have different causal mechanisms undergirding similar functional roles. Or, more precisely, we can never know in advance if there are no causal disanalogies. Therefore, we can never be confident that condition (3) is satisfied and direct inferences from animal test subjects to humans will be questionable.

Some researchers will concede our point. However, they will say, the real benefits of animal experimentation come from the use of animals in basic research. However, before we can discuss basic research in any detail, we must turn to discuss a new development in animal research which has both applied and basic elements: transgenic animals.