Excerpts from ***FAQs on the Use of Animals in Science: A handbook for the scientifically perplexed*** (p49-53):

Why do you say that animals have failed as a valid modality for predicting human drug and disease response?

As the previous chapter has shown, the Theory of Evolution provides us a theoretical framework while advancements in genetic and molecular biology provide the empirical data as to why animals cannot adequately predict human drug and disease response. Mice are unique and interesting creatures in their own right, they are not simply men writ small! In this and the next chapter, we will see how the animal model has failed at the clinical level, particularly in the search for treatments and cures for some of the major diseases.

Animals and humans share many similarities in terms of the "stuff" they are made from (all have cells, genes, lipids, proteins, and so on), but they also exhibit many differences. At the subcellular and genetic level, where the vast majority of research is now taking place, *organizational differences* between animals and humans outweigh the similarities in ways that are relevant to a discussion of prediction. (Recognition of organizational differences draws our attention to the way "stuff" is put together, used and regulated).

However, it remains true that some of those who insist that animals can be *predictive* models do not understand the use of the word *prediction*; in this chapter, we will explore what prediction means in the world of science, and how animals fail to meet a proper scientific standard of prediction.

In science, how is prediction different from guessing correctly and finding correlations?

In science, guessing correctly or finding correlations are not the same as predicting the answer. A fundamental part of any theory or practice claiming predictability is its ability to predict the result of an experiment that has not yet been done. For example, the Second Law of Thermodynamics predicts that a perpetual motion machine can never exist, and that there are upper limits to the efficiency of engines, such as the one in your car. The field of astronomy uses the laws of physics to predict the location of planets.

Again: In science, if you claim predictability, you must be able to predict the result of an experiment that has not yet been done. This concept separates the scientific use of the word *predict* from the lay use of the word, which more closely resembles words such as guess and conjecture—as we saw with the fortune teller and the university history department.

Where does the burden of proof lie in terms of prediction?

Those who claim that animal models are predictive must demonstrate that this claim is correct. The evidential burden of proof resides with those who make the claims. Do you believe that Big Foot exists and is roaming the forests of the Pacific Northwest? Then the burden of proof is on you to prove to us skeptics that

Big Foot does exist. It is no use saying to the skeptic, "You haven't shown me that Big Foot isn't there."

Neither is it our responsibility to provide examples of scientifically viable and predictive tests that do predict carcinogenicity. A modality is either predictive or it is not, regardless of what else is available. Astrology is not a predictive modality for learning about future marriage prospects. Neither is anything else to the best of our knowledge, but that does not mean that astrology wins by default.

To paraphrase a popular adage, "show me the data." If someone suggests that an animal—say, a mouse—can predict human response to chemicals vis-à-vis carcinogenesis, she would need to provide data that support her claim. Perhaps no one animal alone is capable of predicting human responses. But when the same result occurs in two species—say a mouse and a monkey—then perhaps the results are predictive. But the need to provide the appropriate statistical data to prove that claim is the same.

What are the criteria for determining the predictive value of an animal model—in other words, what kind of statistical data would prove that an animal model is or isn't predictive?

In biology, many concepts are best evaluated by using simple statistics like the four described below. By using these four measures we can ascertain whether a test does what we want it to do—whether it is predictive. The criterion for predictability lies in how the test measures up in these four areas.

For example, in medicine we can use a blood test to determine whether someone has liver disease. In order to ascertain how well this test actually determines the health of the liver, we calculate the following:

1. The sensitivity of the test;
2. The specificity of the test;
3. The positive predictive value (PPV); and
4. The negative predictive value (NPV).

The *sensitivity* of a test is the probability of a positive test among people whose test should be positive. In this case, the sensitivity of the blood test in our example would be the probability of a positive test among people who do have liver disease.

The *specificity* of a test is the probability of a negative test among people whose test should be negative. In our example, the specificity of the test would measure the probability of a negative test among people who do not have liver disease.

The *positive predictive value* (PPV) of a test is the proportion of people with positive test results who are actually positive. In our example, the PPV of the test would measure the proportion of people with positive test results who actually have liver disease.

The *negative predictive value* (NPV) is the proportion of people with negative test results who are actually negative. In our example, the NPV of the test would measure the proportion of people with a negative test who actually do not have liver disease.

All these values are measured on a scale from 0.0 (being the lowest) to 1.0 (being the highest). Very few tests have a sensitivity, specificity, PPV, or NPV of 1.0. But in order for the test to be useful—in this case, to tell us if the patient actually has liver disease—it needs to be predictive far more often than not.

How do these measurements apply to animal tests?

Before we answer that question you need to understand how science in general works vis-à-vis what counts as evidence. Scientists conduct experiments of various types and then submit their results and thoughts to journals that have referees. Referees are scientists in the same field (*peers* of the scientist submitting the paper) who judge whether the experiments, results, and the scientists' thoughts about it all are worthy of being published. There are many flaws to this system but all in all it works. The question we are considering, whether animals are predictive for humans, would be answered in part by scientists comparing the results of animal tests with the results from humans taking the drug or suffering from the disease. These results would be subjected to the four tests described in the previous question. Then the resulting paper would be submitted and hopefully accepted to a scientific journal and the experiment and results made available to the scientific community. This entire process is sometimes referred to as *peer review* and the scientific journals are called *peer reviewed journals*.

We can show this best by a specific examples. The data from testing six drugs on animals were compared with the data from humans and published in a peer review journal [45]. The animal tests were shown to have a sensitivity of 0.52 and the PPV was 0.31. The sensitivity is about what one would expect from a coin toss, and the PPV even less. This is not considered predictive in the scientific sense of the word.

Two studies conducted in the 1990s and also published in peer reviewed journals were equally revealing. One showed that out of 24 human toxicities only 4 were found in animals [46]. In another study, in only 6 of 114 cases did clinical toxicities have animal correlates [47]. Fletcher reported on drug safety tests and subsequent clinical experience with 45 major new drugs. Some effects were seen only in animals, while others were observed only in man. The survey established that 25 percent of toxic effects observed in animals might be expected to occur as adverse reactions in man. [48] Lumley:

> In one small series in which the toxicity in clinical trials led to the termination of drug development, it was found that in 16/24 (67%) cases the toxicity was not predicted in animals. [49]

Many other studies have found the same lack of predictability.

The sensitivity, specificity, PPV, and NPV of animal models based on these studies are obviously suboptimal. Unless a test as a whole has a highly positive predictive value (PPV), the test is not useful. (See *Animal Models in Light of Evolution* for more such experiments.)