Is This Study Believable?
Examples from Animal Studies with GM Foods

Bruce Chassy
Department of Food Science and Human Nutrition
40 National Soybean Research Lab
University of Illinois
Urbana, Ill. 61801

Wayne Parrott
CAGT/Department of Crop and Soil Sciences
111 Riverbend Road
University of Georgia
Athens, Ga. 30602-6810

It is not uncommon to read news articles or even scientific reports about feeding studies conducted on genetically engineered (GM) crops that claim the GM crop caused some sort of harm. Everyone has heard that one shouldn't believe everything that's in print, and these reports are a good example of why that's as true today as ever. It's not just newspapers that must be read with a grain of salt, since even reports found in the best scientific journals can be in error.

This article discusses how animal studies are done, what can be learned from them, and describes a few common pitfalls of animal studies. The article provides a framework for deciding whether to believe a sensational claim by considering the following points:

1. How are, and why are, animal studies with foods and feeds used by scientists?
2. Make sure the samples tested are comparable samples
3. Test composition to make sure the tests and controls are comparable
4. The need for an acceptable balanced and nutritious diet
5. Why the dose is important
6. What statistics do and don't tell us
7. The importance of peer review and scientific publication
8. Guidelines for dealing with conflicting information
9. Ethical considerations.

Historically, animal studies — rather than human testing — have been very effective in predicting if a given chemical is toxic, if the chemical has useful pharmaceutical effects, or if it is nutritious. The fields of toxicology and nutrition depend on doing animal studies of this kind. With the advent of GM crops, animal feeding studies now have been used by some to test the safety of whole foods — i.e., whole tomatoes, potatoes, corn grains, or soybean. It should be noted at the outset that whole foods are very different from the pure chemicals traditionally used in feeding studies, since whole foods contain hundreds, if not thousands, of distinct chemical compounds.

The premise is that animals will still be exposed to any potentially harmful chemical that might be present in a whole food. However, the complex composition of whole foods means that animals can only be exposed to very low concentrations of the potentially toxic component, and that such component may interact in ways that would not be observed when testing a pure chemical. As a consequence, tests with whole foods have a limited ability to detect adverse effects (Chassy, et al. 2004; Chassy, et al. 2008; van Haver, et al. 2008).

The tests also are extremely difficult to conduct properly, thus producing misleading results or erroneous conclusions.

Some consumers have asked why GM foods are not tested in humans. As the reader will appreciate from this article, while whole-food feeding studies are difficult to do in animals, they are essentially impossible to conduct successfully in humans. Humans are terrible experimental animals when it comes to diet studies. Humans are all genetically different from one another while laboratory animals are often inbred to ensure genetic uniformity. This means that each animal will react very similarly to the others; in contrast, no two humans would be expected to react exactly alike. Furthermore, humans travel about rather than stay in the controlled environment of a cage. They have many competing interests and do not follow lab protocols, and they probably wouldn't accept a diet that contained, for example, 30 percent soybeans for a year or two.

In spite of the obvious limitations and challenges, animal feeding studies with novel foods are conducted in many countries, and the results are published. A list of factors that confront investigators and which should be considered when interpreting animal feeding studies is considered in the following sections. If any one of these is not properly controlled in a study, the results should be set aside until a valid study is published in a high-quality peer-reviewed scientific journal, and the study is independently corroborated in another laboratory.

In order to facilitate reproducibility and allow laboratories to compare their results, investigators are expected to follow published international guidelines that describe in great detail exactly how to perform these complicated studies. As such, one factor that says a study is well done is if it follows international guidelines, such as those published by the OECD, EFSA, EPA, and FDA (see Marshall 2007 for references to guidelines). If a study does not follow published international guidelines, its results should be viewed with great skepticism.

A Good Vintage

Everyone has heard it: In some years, wines are better than other years, and are declared to be a particularly good vintage. Wine from different years tastes different because it is chemically different, yet the winegrapes came off the same plant. The reason is that each year has a unique temperature and rainfall pattern that affects how the plant grows and the chemical composition of its fruit.

The same is equally true of food and feed crops, to the point that certain regions are famous for producing agricultural products of especially high quality, as a result of their unique soils and growing conditions, which alter the chemical composition, and hence the taste of a plant. The French even have a word for it — *terroir*, meaning a taste of place.

There can still be alterations in the chemical composition of crops due to growing conditions, but these would not be noticed if the alteration did not alter their flavor. For example, one study in

Illinois found that isoflavone levels in one soybean variety differed depending on the part of the state where the soybean was grown, ranging from 47 mg/10 g in Girard to 171 mg/10 g in DeKalb (Eldridge and Kwolek 1983).

Why is this important?
The inherent variability over different growing seasons or locations will cause compositional differences in the food that is being fed to the animals. To determine the effects of one chemical in a whole food, whole foods with and without that one chemical are fed to animals. However, if other chemicals in the whole food are changed, the researcher has no way to tell if any effects seen are due to the chemical being tested, or to changes in the types and amounts of other chemicals that differ in the foods being compared.

Hence, properly conducted animal feeding studies must use feed grown at the same time and in the same place in order to be valid. Even when the food comes from the same place and the same growing season, scientists must measure the composition to ensure that animals receive the same diets and that there are no unexpected changes. For example, in the case of the soybean study cited above, the different levels of isoflavones, which are hormone-like compounds, can lead to differences in the animal's growth, physiology, organ composition, behavior, and reproductive outcomes.

Many crops such as corn also will have different levels of natural toxins like aflatoxin and fumonisin, both of which affect animal health. If a study doesn't use comparable test materials, it simply can't produce valid outcomes and should be disregarded.

(*The authors do not know of even one feeding study with GM crops that concluded an adverse health effect or a potentially adverse change had occurred when the properly produced control and test foods were used*.)

Comparing apples to apples

A Granny Smith apple looks different and tastes different from a Red Delicious apple because it is different in terms of its natural ingredients, such as sugars and organic acids.

Why is this important?
Because of their compositional differences, the same results cannot be expected from feeding Granny Smith apples to animals as would be if the animals were fed Red Delicious apples. The same applies to different varieties of feed crops, which can vary between themselves as much as apples can. In the previously mentioned study on soybean isoflavones, different varieties grown at the same site also showed large variation in isoflavones.

To use apple terms, a study using Granny Smith apples can only be compared to another study using Granny Smith apples grown in the same orchard and harvested at the same time. The scientific term used to denote this scenario is that feeding studies must be conducted using "near isolines," or simply "isolines" for short, grown in the same place in the same season.

In the case of GM plants, isolines differ only by the engineered gene, thus singling out the gene product for study. Any study that does not use isolines is not valid. The failure to use isolines may be the reason tissue-specific changes have been observed in mice fed GM soybean (Malatesta et al. 2002).

(*The authors do not know of even one feeding study with GM crops that concluded an adverse health effect or a potentially adverse change had occurred when isolines were used for control and test samples*.)

There are some things even a rat won't eat

One of the original reports that claimed adverse effects from GM crops was a study in which rats were fed GM potatoes (Ewen and Pusztai 1999). The rats — whether they were eating conventional or GM potatoes — eventually showed alterations in the intestinal lining.

Why is this important?
This infamous rat experiment shows the several limitations in feeding studies. First, raw potatoes are toxic to rats, so a proper experiment with rats and raw potatoes can never be conducted. Second, the inserted gene product is usually present at such low concentrations in a food that the animal must be fed large amounts of the food in question in order to get enough of the test chemical into the animal.

It is an easy matter to test the pure protein by feeding it to animals, but it is not so easy when the animals must be forced to consume large amounts of potatoes. Feeding large amounts of a single food at the expense of a balanced diet leads to nutritional imbalances. Imagine if a person were fed only baked potatoes. After a while, the person would become malnourished and thus appear ill. These symptoms are not an indication that the baked potatoes are harmful — they are simply an indication that the person is not eating a balanced diet.

Hence, feeding studies must be carefully designed to ensure that nutrient imbalances do not appear as harmful effects of food. The only way to establish diet identity is to do a complete chemical and nutritional analysis of control and test diets. Many studies that make claims about adverse effects of GM foods do not provide such analysis. One journalist concluded that about all Pusztai's study proved is that rats hate potatoes (MacKenzie 1999).

The dose makes the poison

Several studies mentioned above were conducted with a single concentration, or dose, of control and test material in the diet (Pusztai 1999, Malatesta 2002, Velimirov 2008). An even more flawed example is a widely publicized study that did not appear in a peer-reviewed scientific journal in which rats were allowed to eat out of common food dishes and to choose between animal chow or the soybeans being tested (Marshall 2007). Obviously, the investigators in this case not only didn't study multiple doses, they didn't even know which, if any, of the rats consumed the food being tested.

Studies that investigate the effect of a diet containing only a single dose of the material being tested are not restricted to rodents. Two studies were published claiming that pollen from engineered corn was harmful to monarch butterflies (Losey, et al. 1999) and caddisfly larvae (Rosi-Marshall, et al. 2007). Both studies suffer from several serious flaws, but a key one is the lack of a dose response.

Why is this important?

For a study to prove that a chemical is toxic, the toxicity symptoms must become more serious as the dose of the poison is increased. This is the classic way in which the toxic nature of a poison is demonstrated by toxicologists. A study that uses only one dose is not enough to prove something is a toxin.

Any study that uses random doses is hopelessly lost. If several doses are used, and there is no dose response, then whatever effect has been observed cannot be attributed to a toxin. If the symptoms become more severe as the dose is increased then the material being tested may be toxic, but other direct tests might be necessary to confirm this. And of course, the effect would have to be reproduced by an independent research team.

In keeping with the concept that the dose makes the poison, it is also important to ask if the study used a realistic amount of the material being tested. If the study used a dose that is thousands or even millions of times higher than an animal or person could realistically consume, the study is not really relevant. Everything we consume is potentially toxic to us if consumed in large amounts. For example, table salt is harmless when used at normal levels, and is even necessary for proper body function. However, about 7 ounces of table salt is enough to kill half of the 150-lb people who would eat it all at once, and as little as 2 ½ ounces of salt eaten all at once has been known to be enough to kill a 150-lb person.

The father of toxicology, Paracelsus (1493–1541), said it very well almost 500 years ago when he said: "All things are poison and nothing is without poison, only the dose permits something not to be poisonous," or as more commonly stated, "the dose makes the poison."

Lies, damn lies, and statistics

Scientists study small samples, and then use statistics to project the results to larger groups or populations. Statistics are based on the results obtained with a few animals in a few experiments, and because of natural biological variability not all animals give the same results for each measurement that is made on them. This effect of variation due to chance alone is taken into account when experiments are designed.

By convention, experiments are designed to arrive at the right answer 95 percent of the time — that is, on average, 5 percent of the time, the experiments will say there is what is called a statistical difference between the control and test groups. There is a temptation to automatically say that the effect is due to the food, when in reality it isn't 5 percent of the time.

Why is this important?
A recent report (Séralini, et al. 2007) reviewed a study that had been conducted on food safety of engineered maize. The study found that about 30 out of 500 comparisons of measurements (for example heart weight, kidney weight, body weight, etc.) between animals fed conventional feed and animals fed GM feed showed a difference, as opposed to the expected number of 25. Séralini claimed this meant that GE feed produced different outcomes than conventional feed, but expert statisticians called this a very good agreement between anticipated and observed differences.

When the study used a stricter 1 percent error rate, so that 99 percent of the measurements would be within the normal range, only four (as opposed to the expected five, on average) of the studies showed any effect. This is again good agreement, and what it means is that the study shows that the anticipated number of differences was found. Regulators in Europe rejected the claims made by Séralini et al. (see the European Food Safety Authority's statement at http://www.efsa.europa.eu/EFSA/efsa_locale-1178620753812_1178621165358.htm).

Statistics must deviate greatly from the 5 percent average value before it is possible to claim what statisticians call *statistical significance*. Another way of looking at this is that there will always be differences found between two groups of animals. Most of these differences are the result of random biological variation and have no biological significance. Differences are only important if they cause an adverse effect on health or nutrition.

Many of the animal studies that have been announced in the media that claim adverse effects of GM crops are really seeing simple biological variation that occurs when many measurements are made on test animals. These studies have also failed to conduct the necessary follow-up studies to verify the results.

Caveat lector

Science has set in place a system that is supposed to be self-correcting and self-policing. When scientific results are real, and not due to chance or poorly conducted experiments, they are repeatable by other scientists. Only time can tell if reported results are reproducible. This suggests that the public should not jump to a conclusion when they hear about a new research result on the television or read about it in the paper.

One way to ensure that experiments get repeated is by publishing them and making them public. However, they should not be published in just any newspaper or web site, but rather in peer-reviewed journals. Before new research results can be published in a peer-reviewed journal, the report is reviewed by a group of experts, who then attest, to the best of their abilities, that the experiment was conducted properly and the results are probably being interpreted correctly. In essence, the system helps separate the wheat from the chaff.

Why is this important?
Remarkably, many of the studies that show GM foods have adverse effects have bypassed the peer-review process. Instead, the scientists announced the results in press releases, as did Arpad Pusztai in 1999 when he announced his findings on BBC (Marshall 2007, Velimirov 2008).

Science-by-press-release is not science at all, but rather propaganda. Always consider the source, and be very cautious about any results that have not been peer-reviewed. Remember too that even peer review isn't perfect, that's why a study needs to be verified independently by other scientists. Several examples of flawed studies that were published in peer-reviewed journals were noted above (Losey 1999, Rosi-Marshall 2007). Many of the studies that claim harmful effects of GM crops on animals have not been verified or could not be verified. That's the reason for *caveat lector* — literally, *let the reader beware*.

Conflicting information

There can be no doubt about it, GE crops are controversial. Divergent information has appeared in the media about whether these crops are safe to eat, safe to plant, and are good for the environment. Some are based on fact, others on speculation or misinformation.

Conflicting stories understandably cause confusion. This paper focuses on animal studies, but there are also claims about super-weeds, genes escaping into other organisms, the possible spread of antibiotic resistance, the loss of Monarch butterflies, and a host of others.

How can we tell if a study is valid?
The book *Tomorrow's Table* provides readers a check-list for determining if a study is believable (Ronald and Adamchak 2008). These points boil down to the following:

1. Determine the primary source of information
2. Check if the work was published in a peer-reviewed journal
3. Check if the journal has a good reputation for scientific research
4. Determine if there is an independent confirmation by another published study
5. Assess whether a potential conflict of interest exists
6. Assess the quality of institution or panel
7. Examine the reputation of the author

The primary source of the information is important. Much of the misinformation circulating today comes from testimonials, such as reports of high-school experiments or casual observations by farmers. These have been picked up by the popular media, various web sites, and the blogosphere. Even when a story starts from a reputable source, the message gets distorted in the popular press.

Conflicts of interest often affect the message that many groups espouse. For example, the Organic Consumers Association wants to avoid competition from GM crops. The Austrian government funded the previously mentioned study (Velimirov 2008) to justify its stance against GM crops. Greenpeace depends on donations from people who are afraid of GM. Companies want to sell their products. Scientists need grants to support their research. It's always a good idea to try to understand what influences sources of information.

Finally, it is not just the reputation of an author that is important, but his or her background and training. Not surprisingly, some of the most controversial reports come from people with training in totally unrelated fields. Sometimes they have no experience in genetics or agriculture. Animal feeding studies have been conducted by investigators with no prior experience in designing or interpreting animal-based research.

It is pretty clear from the foregoing discussion that if research isn't published in a peer-reviewed journal *and* accepted by a consensus of the scientific community, it probably shouldn't be trusted. Trust is a very important concept, too. Deciding what to believe really comes down to whom one trusts. As scientists, we believe that we can trust the scientific literature — but only if work is independently verified and becomes generally accepted by the scientific community.

Parting thoughts: Animal studies are important, but . . .

As mentioned previously, animal feeding studies with whole foods are especially difficult to perform and are particularly prone to errors. Animal studies can be used to provide vital information about new drugs, the nutritional value of foods and feeds, and the potential toxicity of certain substances. Valuable information can be obtained if they are done correctly and follow the recommended guidelines.

Animal studies with whole foods should be done only when there is a clear scientific hypothesis that can be tested by the study design. Most of the studies that claim adverse effects of feeding GM foods and which have been sensationalized in the media do not meet the criteria of proper study design and conduct, do not test a hypothesis, and do not perform appropriate statistical analyses. Some have also not been published in the peer-reviewed literature. A study that fails to meet any of these criteria should be disregarded no matter how scary the claim.

When should animal studies be done?

Most people agree that it is wrong to test potentially toxic substances in humans, and some think it's also wrong to use animals. Proponents of animal studies say that they save human lives and relieve suffering, while opponents say that it's not right to use animal testing for any reason. The debate cannot be settled here, but there are clear ethical issues involved when animals are used for studies that cannot hope to provide answers. Therefore, many animal feeding studies with whole foods appear to be ethically unjustifiable and should be greeted with skepticism, if not criticism, by the general public and scientific community.

-- *Photos by Corbis*

Cited Literature

Chassy, B., J.J. Hlywka, G.A. Kleter, E.J. Kok, H.A. Kuiper, M. McGloughlin, I.C. Munro, R.H. Phipps, and J.E. Reid. 2004. Nutritional and safety assessments of foods and feeds nutritionally improved through biotechnology: Prepared by a task force of the ILSI International Food Biotechnology Committee, Comprehensive Reviews in Food Science and Food Safety, vol. 3. (http://members.ift.org/NR/rdonlyres/27BE106D-B616-4348-AE3A-091D0E536F40/0/crfsfsv3n2p00350104ms20040106.pdf).

Chassy, B., M. Egnin, Y. Gao, K. Glenn, G.A. Kleter, P. Nestel, M. Newell-McGloughlin, R.H. Phipps, and R. Shillito. 2008. Recent developments in the safety and nutritional assessment of nutritionally improved foods and feeds. Comprehensive Reviews in Food Science and Food Safety 7 (1), 50–113 (http://www.blackwell-synergy.com/doi/abs/10.1111/j.1541-4337.2007.00029_1.x).

Eldridge, A.C., and W.F. Kwolek. 1983. Soybean isoflavones — effect of environment and variety on composition. Journal of Agricultural and Food Chemistry 31:394–396.

Ewen, S.W.B., and A. Pusztai. 1999. Effect of diets containing genetically modified potatoes expressing Galanthus nivalis lectin on rat small intestine. Lancet 354:1353–1354.

Losey, J.E., L.S. Rayor, and M.E. Carter. 1999. Transgenic pollen harms monarch larvae. Nature 399:214–214.

MacKenzie, D. 1999. Unpalatable truths. New Scientist. p. 1018. April 17, 1999.

Malatesta, M., C. Caporaloni, S. Gavaudan, M.B.L. Rocchi, S. Serafini, C. Tiberi, and G. Gazzanelli. 2002. Ultrastructural morphometrical and immunocytochemical analyses of hepatocyte nuclei from mice fed on genetically modified soybean. Cell Structure and Function 27:173–180.

Marshall, A. 2007. GM soybeans and health safety — a controversy. Reexamined (an interview with B.M. Chassy, et al.) Nature Biotech. 25:981–987.

Ronald, P.C., and R.W. Adamchak. 2008. Tomorrow's Table: Organic Farming, Genetics, and the Future of Food. Oxford University Press.

Rosi-Marshall, E.J., J.L. Tank, T.V. Royer, M.R. Whiles, M. Evans-White, C. Chambers, N.A. Griffiths, J. Pokelsek, and M.L. Stephen. 2007. Toxins in transgenic crop byproducts may affect headwater stream ecosystems. Proceedings of the National Academy of Sciences of the United States of America 104:16204–16208.

Séralini, G.E., D. Cellier, and J.S. de Vendomois. 2007. New analysis of a rat feeding study with a genetically modified maize reveals signs of hepatorenal toxicity. Archives of Environmental Contamination and Toxicology 52:596–602.

van Haver, E., G. Alink, S. Barlow, A. Cockburn, G. Flachowsky, I. Knudsen, H. Kuiper, D.P. Massin, G. Pascal, A. Peijnenburg, R. Phipps, A. Poting, M. Poulsen, W. Seinen, H. Spielmann, H. van Loveren, J.M. Wal, and A. Williams. 2008. Safety and nutritional assessment of GM plants and derived food and feed: The role of animal feeding trials. Food and Chemical Toxicology 46:S2–S70.

Velimirov, A., C. Binter, and Univ. J. Zentek. 2008. Biological effects of transgenic maize NK603xMON810 fed in long term reproduction studies in mice. Forschungsberichte der Sektion IV Band 3/2008. http://bmgfj.cms.apa.at/cms/site/attachments/3/2/9/CH0810/CMS1226492832306/forschungsbericht_3-2008_letztfassung.pdf.