Considerations in Toxicology Study Design and Interpretation: An Overview

Gradient Corporation: Lewis, AS; Beyer, LA; Langlois, CJ; Yu, CJ; Wait, AD

Introduction

The use of animal data to evaluate the safety of chemicals in humans can be complex. Biological differences among species and the use of high experimental doses often make animal test results difficult to interpret with regard to human relevance. Despite the difficulties, animal studies have formed the cornerstone of toxicology and safety assessments. Animal studies are used to assess a variety of toxicological outcomes: potential noncancer effects, and genotoxicity, carcinogenicity, as well as adverse reproductive and developmental outcomes. Rodents are the most widely used animal models, but other types of animals with special sensitivities may be used for specific endpoints. While is it impossible to eliminate all uncertainty, toxicology studies can be designed to maximize utility and reduce the generation of ambiguous results. Some of the most important considerations of a well-designed study are route of exposure, dose selection, and animal model.

In animal testing, a route of exposure that reflects probable human exposure is key. Without an appropriate route of exposure, study interpretation can be difficult. For example, while it is easier to administer substances to animals using intravenous injection, toxicity information from these experiments may have little relevance to oral exposures. This is because the toxicity (or detoxification) of a compound may be highly dependent on absorption into the gastrointestinal tract and subsequent metabolism. In addition, oral exposures usually do not reflect dermal toxicity and vice versa.

Dose selection is probably the most complex, but important, decision from a study-design standpoint. One of overall goals of a toxicology study should be to identify a dose associated with no observable adverse effects (called a no-observable adverse effect level [NOAEL]). It is important to establish a NOAEL, which can then be used to establish a threshold of toxicity (*i.e.*, a dose below which no adverse effects are expected). Testing a single dose associated with no adverse effects, however, is often not adequate to determine safety. Because there is some degree of uncertainty between animal and human responses, doses (sometimes well above expected human exposures) should also be tested to demonstrate that there is an ample margin of safety (also called margin of exposure). Many regulatory agencies that register chemicals and products require testing up to doses that are overtly toxic to animals, not only to understand the types of toxic responses that can be expected at higher doses, but also to understand how great the gap (or margin) is between expected human exposures and doses associated with animal toxicity.

What constitutes an adequate margin of safety can be quite complex but, in general, reflects how confident one is that the NOAEL from an animal study will also be nontoxic to humans, even individuals with increased sensitivity (*e.g.*, children, the elderly). As mentioned above, biological differences between humans and test animals are one source of uncertainty. In addition, information about expected human exposures and inter-individual variability in response may also be poorly understood. It should be noted that any safety information based on a history of use in humans will undoubtedly influence perceptions on what constitutes an adequate margin of safety, but at present, there is no formal framework for incorporating this type of information into the design of animal studies. A risk-benefit analysis may also be an important component for certain types of compounds, particularly pharmaceuticals.

Regulatory entities within the US and internationally, often require a prescribed battery of standard studies for product registration. These requirements not only vary across countries, but among different agencies within the same country. The type(s) of proof of safety varies, in part, because the agencies involved in establishing testing guidelines (*e.g.*, Food and Drug Administration [US FDA], Environmental Protection Agency [US EPA], Organization for Economic Co-operation and Development [OECD], *etc.*) formulate testing regimens for different types of products; examples include industrial chemicals, cosmetics, food additives, dietary supplements, pesticides, and pharmaceuticals. In some cases, there is little or no regulatory guidance. For example, no specific toxicology testing protocol exists for dietary supplements in the US. Even when strict requirements exist, scientific experience is key. Today, regulations recognize the importance of scientific judgment in choosing the type of test, the test species, the duration of the test, and the dose(s) (US FDA, 2007; NRC, 2006). Further, the scientific underpinnings of our understanding of toxicity are continually changing (Hayes, 2008).

Standard Toxicology Tests

Presented below is a broad overview of standard toxicity tests. It must always be remembered, however, that knowledge of the product/ingredient, its structure, and its proposed use should guide decisions regarding toxicity testing. Standard tests by exposure duration are discussed. We have focused on oral exposures. Different tests would be necessary for different exposure routes (*e.g.*, dermal, inhalation).

Acute toxicity: This test, used to evaluate high, short-term exposures, may also help to determine dosing regimens in longer-term tests. If a high dose (*e.g.*, 5,000 mg/kg) is found to be survivable, no further acute testing is conducted (NRC, 2006). In general, a single dose is given and monitored for several days to weeks after dosing, although compound administration may take place several times within or

continuously throughout a 24-hour period. Aside from survivability, acute tests are useful, because they can reveal whether frank toxicity is sudden, delayed, time-limited, or continuous. The time to onset and resolution of toxicity can provide insight into compound attributes such as pharmacokinetics and bioavailability. Importantly, in some cases, previous human experience with a compound may make acute single-dose testing unnecessary (Wu *et al*., 2008).

Subchronic toxicity: Subchronic studies evaluate the adverse effects of continuous or repeated exposure over a portion of the average life span of experimental animals. Specifically, they provide information on target organ toxicity and are designed to identify NOAELs (NRC, 2006). They can also help determine appropriate dose regimens for longer-term studies. Exposure durations are typically 28 or 90 days. Administration of the chemical is determined by the route of human exposure. Animals are often observed for 2 or 4 weeks after the end of treatment for reversibility, persistence, or delayed occurrence of adverse effects. Typically, doses in subchronic studies are selected to define a dose-response relationship. Toxicological endpoints evaluated include clinical signs (gait, changes in skin, fur, eyes, posture), motor activity, sensory reactivity, body weight, food consumption, and clinical pathology tests (*e.g.*, blood and urine tests). At study termination, gross necropsy is performed and a full histopathological analysis is conducted.

Chronic toxicity: The purpose of a chronic study is to determine cumulative adverse effects of repeated daily oral dermal or inhalation exposures over a 12- to 24-month period (depending on test species). Because of the importance of chronic toxicity, testing is done in two mammalian species (one rodent [rat] and one non-rodent [dog]), as a general rule, though other species can be used with adequate justification. Dose selection is based on the results of subchronic studies. At least three dose groups and a control group should be used. The highest dose should only cause mild signs of toxicity, while the lowest dose should show no adverse effect (NRC, 2006). During the study, body weight and food consumption should be measured along with some clinical pathology tests. At the end of the study, all organ systems are analyzed and a full histopathologic analysis is done (NRC, 2006).

Carcinogenicity: Carcinogenicity bioassays determine cumulative neoplastic effects of repeated exposures over most of the lifetime of the animal. Study design is very similar to a chronic test, and testing is often combined. The carcinogenicity bioassay is conducted with rodents for a minimum of 24 months (rats) and 18 months (mice). Dose selection guidelines are similar to chronic toxicity studies; however, sample sizes are generally larger—50 rodents of each gender per dose group. Further, at the end of the study, extensive gross necropsy and histopathology are conducted to detect neoplasms (NRC, 2006).

Genotoxicity: Genotoxicity testing is performed to ascertain adverse effects on DNA, genes, and chromosomes. It is recognized that many human diseases, and specifically cancer, begin as mutations. Both *in vivo* and *in vitro* genotoxicity tests are available, but *in vitro* tests, which use cell cultures or extracts rather than whole organisms, are relatively quick and much less expensive than animal testing. These tests are designed to detect gene mutations, chromosomal aberrations, and various types of DNA damage.

Although not described here, tests of developmental and reproductive toxicity may also be standard for certain types of chemicals. The results of standard studies may indicate the need for more information on specific endpoints, such as neurotoxicity or immunotoxicity, for example. Also, it should be noted that in many instances, refined follow-up experiments can be useful (but not required) for

understanding the biological basis of an adverse effect, which can be helpful when evaluating the relevance of test results to humans (see below).

Study Interpretation

If adverse effects are observed in a toxicology study, the default assumption is that the compound *could* lead to similar toxic effects in humans. However, in reality, the toxic effects are often not seen in human populations for a variety of reasons including biological differences among species and the fact that high doses (higher than those experienced by humans) are often used in animal experiments. Investigation into potential human relevance can be a complex undertaking requiring a historical perspective on the sensitivity of certain test species to adverse effects, a detailed knowledge of animal and human biology, and an understanding of toxicological interactions on cellular and molecular levels. Several key lines of investigation are mentioned below, but it should be emphasized that evaluating human relevance must be performed on a case-by-case basis—there is no one-size-fits-all approach.

One important consideration is whether the observed effect is a high-dose phenomenon, such that the effect would be unlikely to occur at lower, typical human exposures. This is often the case for noncancer effects, which are expected to operate with a threshold of toxicity. For suspected carcinogens, where it is often assumed that any exposure leads to increased risk, establishing a threshold can be more difficult. However, there has been recent recognition that certain animal tumors occur only when experimental concentrations are high enough to cause tissue damage that leads to a reparative process (Meek *et al*., 2003).

It may also be important to have a thorough understanding of how the compound is metabolized; the process may be different in animals and humans, resulting in marked differences in toxicity. As an example, theobromine, a compound in chocolate, is fatally toxic to dogs in relatively small doses, because they metabolize theobromine so slowly. Less than 100 grams (or about 3.5 ounces) of unsweetened chocolate can be fatal to a 20-pound dog (Finlay and Guiton, 2005). On the other hand, humans can clearly tolerate and consume chocolate in much higher quantities.

In addition to metabolic differences, humans and animals can also vary biologically in a number of different ways that may modulate responses to chemicals. For example, in rodents, a class of chemicals called peroxisome proliferators binds to a receptor called PPAR-α, which causes abnormal increases in cell proliferation and eventual liver tumors. Although humans have the PPAR-α receptor,

they have substantially lower quantities compared to rodents, making humans resistant to PP-induced liver tumors (Klaunig *et al*., 2003). As another example, rodents are exquisitely sensitive to compounds that disrupt thyroid hormones, while humans have a greater ability to adapt and maintain normal thyroid hormone levels (NRC, 2005; US EPA, 1998).

In conclusion, while testing guidelines are prescriptive for many products, testing for other types of products may be flexible. Regardless, knowledge of the compound, how it will be used, and sound scientific judgment should guide testing decisions. Choosing the appropriate tests and dosing regimens that will demonstrate an adequate margin of exposure is a critical step in establishing human safety. If adverse effects are observed in animal studies, more refined toxicology studies may provide insight on the potential human relevance of observed effects and, depending on the outcome, may offer assurance that effects observed in animals will not occur in humans under intended use scenarios.

References:

Finlay, F; Guiton, S. 2005. "Filler: Chocolate poisoning." *Br. Med. J.* 331:633. Accessed at http://www.bmj.com/cgi/content/full/331/7517/633.

Hayes, AW (Ed.). 2008. *Principles and Methods of Toxicology (Fifth Edition.)* Taylor & Francis, Philadelphia, PA, 2270p.

Klaunig, JE; Babich, MA; Baetcke, KP; Cook, JC; Corton, JC; David, RM; DeLuca, JG; Lai, DY; McKee, RH; Peters, JM; Roberts, RA; Fenner-Crisp, PA. 2003. "PPAR-alpha agonist-induced rodent tumors: Modes of action and human relevance." *Crit. Rev. Toxicol.* 33 (6) :655-780.

National Research Council (NRC). 2006. "Toxicity Testing for Assessing Environmental Agents. Interim Report." National Academies Press, Washington, DC.

National Research Council (NRC). 2005. "Health Implications of Perchlorate Ingestion." National Academies of Science, National Academies Press, Washington, DC.

Meek, ME; Bucher, JR; Cohen, SM; Dellarco, V; Hill, RN; Lehman-McKeeman, LD; Longfellow, DG; Pastoor, T; Seed, J; Patton, DE. 2003. "A framework for human relevance analysis of information on carcinogenic modes of action." *Crit. Rev. Toxicol.* 33 (6) :591-653.

US EPA. 1998. "Assessment of Thyroid Follicular Cell Tumors." Risk Assessment Forum, EPA/630/R-97/002.

U.S. Federal Drug Administration (US FDA). 2000. "Guidance for Industry and Other Stakeholders: Toxicological Principles for the Safety Assessment of Food Ingredients. Redbook 2000." Centerfor Food Safety and Applied Nutrition. Accessed at http://www.cfsan.fda.gov/~redbook/red-toca.html. Last updated in 2007.

Wu, KM; Ghantous, H; Birnkrant, D. 2008. "Current Regulatory Toxicology Perspectives on the Development of Herbal Medicines to Prescription Drug Products in the US." *Food Chem. Toxicol.* 46:2606-2610.