12-2008

Developmental Toxicity Testing: Protecting Future Generations?

Jarrod Bailey
Physicians Committee for Responsible Medicine, jarrod.bailey@crueltyfreeinternational.org

Developmental Toxicity Testing: Protecting Future Generations?

Jarrod Bailey

Physicians Committee for Responsible Medicine, Washington, DC, USA

Summary — A recent editorial is discussed, which implied that animal-based developmental and reproductive toxicology tests will continue to be crucial, that the thalidomide disaster could have been prevented by more animal testing, and that tests on juvenile animals would help to protect children (as developing adults) from the adverse effects of pharmaceuticals. It is argued that animal tests in these scientific areas do not provide reliable data that are predictive for human responses and, even if they did, the tests are too expensive and time-consuming for application to the very large number of substances that need to be tested. It is estimated there are already more than 100,000 man-made chemicals to which humans may be exposed on a regular basis, and it is therefore widely accepted that *in vivo* developmental toxicology could not possibly be used to assess all new and existing chemical substances, due to the scale of its demand upon time and resources. It is therefore imperative that alternatives such as those outlined above are embraced, further developed, accepted and used — as a matter of urgency.

Key words: *animal testing, birth defects, developmental toxicity, teratology, toxicity testing.*

Address for correspondence: *J. Bailey, Senior Research Consultant, Physicians Committee for Responsible Medicine, 5100 Wisconsin Avenue, N.W., Suite 400, Washington, DC 20016, USA.*
E-mail: *jarrod.bailey@newcastle.ac.uk*

Introduction

A recent Guest Editorial in *Lab Animal Europe*, "*Developmental and reproductive toxicity testing: a potted history*" (1) affirmed that "the ultimate goal of developmental and reproductive toxicology testing is to provide data for risk assessment… to protect our future generations." While this is not in doubt, the article implied that this would involve a continued dependence on animal-based tests.

However, this laudable goal cannot be achieved via such means. There is ample evidence to show that this 'data for risk assessment,' when obtained from animal tests, is not suitable for this purpose — it neither permits confident human-risk assessment, nor serves to protect our future generations.

Reliability and Concordance of Animal Data

The data presented here summarise some of the salient points of a comprehensive, systematic and non-selective study completed in 2005 (2), which examined the developmental toxicity results for almost 1400 substances in 12 different species, by using several databases and reference texts.

Historical data — across twelve species

An analysis of the responses of up to 12 animal species to 11 *groups* of known human teratogens (grouped by drug class/chemical nature) revealed significant discordance: the positive predictability ranged from 75% for the hamster, down to 40% for the rabbit, which also exhibited a false-negative rate of 40%. The mean positive predictability rate in the six species most frequently used historically (mouse, rat, rabbit, hamster, primate, dog) was less than 55%, and the number of equivocal results remained high at just under 25%.

Furthermore, there were 139 animal results across different species for 35 *individual* substances positively linked with human teratogenicity. Just over half (56%) of the animal results were positive. This poor predictability was underlined by a US Food and Drug Administration (FDA) report which detailed the responses of the mouse, rat, rabbit, hamster and monkey to 38 known human teratogens, in which the mean percentage of correct positives was only 60% (3). This report also analysed 165 compounds known to be non-teratogenic in humans, for which the mean negative predictive value for these five species was 54%.

Contemporary data — for two species, rat and rabbit

For *groups* of known human teratogens, results from tests in the rat correlated with human classifications in 64% of cases; for the rabbit, this correlation was 40%. When this analysis was focused on 35 *individual* substances known to be associated with human teratogenesis, the positive predictability for the rat

was 61%, with 29% of the results falsely negative. The rabbit was positively predictive in 41% of cases, but produced false-negative results for 56% of substances. A small number of equivocal results were obtained in both these species (Table 1, depicting our analysis of results and classifications contained in Schardein's book [4]).

The performances of the rat and rabbit in teratology tests were further elucidated by examining the results for the 20 chemicals used in the ECVAM validation studies on three non-animal alternative methods for developmental toxicology (Table 2). Nine of these substances had a human risk classification with which to compare the rat and rabbit results:

— four 'unlikely' human teratogens produced two negative, one equivocal and one positive result in the rat, and 3 negative and 1 positive result in the rabbit.

— three 'minimal-to-small risk' human teratogens produced three positive results in the rat, and one positive in the rabbit (the other two were not tested).

— two 'moderate-to-high risk' human teratogens produced positive results in both the rat and the rabbit.

Based on these data, it can be argued that the tests in the rat and the rabbit are not sufficiently predictive to justify their use, and that the high rate of false-negatives also raises concern over their human relevance and applicability. This lack of predictive power is also underlined by the statistic that, of 3301 substances tested prior to 1993, 37% were classified as definitely, probably or possibly teratogenic in animals, but fewer than 2.3% of these substances were linked to human birth defects. (4).

Summary and Conclusions

The predictive nature of animal developmental toxicity tests

Contrary to Moxon's assertions that animal-based developmental toxicology provides valuable data to

Table 1: The results from rat and rabbit teratology tests for known human teratogens

	Total no. of substances	+	+/–	–	True positive (%)	False negative (%)
Rat	31	19*	3	9	61%	29%
Rabbit	27	11	1	15	41%	56%

*The total numbers of results in the rat and rabbit for the 35 known human teratogens (categorised by Schardein [4]) are shown, followed by the number of positive (+), equivocal (+/–) and negative (–) conclusions. The final two columns reveal the percentages of rat and rabbit results that represented True Positives and False Negatives for these substances. * = one of these results was strain-dependent.*

Table 2: The results from rat and rabbit teratology tests for nine chemicals used in ECVAM validation studies on non-animal alternative methods that had human risk classifications

	Human teratogenic potential								
	Unlikely			Minimal-to-small risk			Moderate-to-high risk		
	+	+/–	–	+	+/–	–	+	+/–	–
Rat	1	1	2	3			2		
Rabbit	1		3	1			2		

The results for those nine substances for the rat and the rabbit, where they existed, are provided for chemicals classified as posing 'unlikely', 'minimal-to-small' and 'moderate-to-high' human teratogenic risks.

enable reliable human risk assessment, the examination of substantive data from decades of animal-based teratology revealed significant variability in positive and negative predictability, and high rates of false-positives, false-negatives and equivocal outcomes across twelve species. These tests are therefore not suitable for their intended purpose.

Further, while accepting the inherent scientific and technical challenges involved, Moxon implies a necessary role for juvenile animals in assessing the safety of pharmaceuticals for children, based on an acknowledgement of important differences between children and adult humans and their susceptibilities to the effects of pharmaceuticals. It may be true that potential adverse events cannot be assessed ethically and safely in paediatric clinical trials, and that therefore some form of effective preclinical investigations must be conducted. However, one must be sceptical that animal teratology studies of any kind are fit for this purpose, given the statistics presented here and their illustration of the lack of predictability of this approach.

Alternative approaches to developmental toxicity tests

Fortunately, a number of alternatives to animal testing exist or are in the course development, with the potential to improve the field of developmental toxicology in terms of time, cost and, most importantly, improved human predictability. Thus, it is anticipated that alternatives to the current animal-based methods will greatly enhance the number of substances that can be evaluated for potential developmental toxicity, at lower cost and in a shorter time frame. It is estimated there are already more than 100,000 man-made chemicals to which humans may be exposed on a regular basis (5), and it is therefore generally accepted that *in vivo* developmental toxicology could not possibly be used to assess all the new and existing chemical substances due to the scale of its demands upon time and resources (6).

Computer-based systems, such as expert systems and structure–activity relationship (SAR) analyses, and physiologically based pharmacokinetic modelling (PBPK), have already been responsible for the elimination of many animal tests in the pre-screening of candidate drug compounds. The recently established US Environmental Protection Agency (EPA) ToxCast™ programme employs the computational modelling of high-throughput screening data to assist in prediction of the potential toxicities of chemicals to humans, incorporating reproductive endpoints, as well as other endpoints, including a variety of biochemical and developmental assays (7).

The use of lower organisms, embryo stages and cell, tissue and organ cultures, was endorsed as scientifically validated in 2001. The Embryonic Stem Cell Test, for instance, uses two permanent murine cell lines to screen for teratogenic potential (8–12), and has scored highly on predictability, precision and accuracy in independent validation studies. It is already considered to be more reproducible, provide easier end-points, present no problems with respect to 'route of exposure,' placental transfer and metabolic differences, and is devoid of the confounding factors associated with animal tests, such as intra-species variability, environmental factors, differences in metabolism, placental and other anatomies, absorption, sensitivity, metabolic activation, routes of administration, dose levels and strategies. It provides a means to establish vital mechanistic models of teratogenic action, via gene expression analysis, for example, which will decrease the cost and increase the number of chemicals evaluated for developmental toxicity, could reduce the human impact of the false-positive and false-negative results generated by animal models, and could also greatly reduce the numbers of animals used (2, 13). As human cell culture and other technologies improve, new protocols will evolve, that will enable an even closer *in vitro* approximation of *in vivo* human teratogenesis (14).

In addition, risk assessment can be aided by better information and data comparison and data sharing, and also by valuable human studies and birth-defect registries, which have identified many important human teratogens (2, 15).

In conclusion, the article by Moxon (1), which implied that animal-based developmental and reproductive toxicology tests will continue to be crucial, that the thalidomide disaster could have been prevented by more animal testing, and that tests on juvenile animals would help to protect children (as developing adults) from the adverse effects of pharmaceuticals, has little or no scientific basis. Animal tests in these scientific areas do not provide reliable data that are predictive for human responses and, even if they did, the tests are too expensive and time-consuming for application to the very large number of substances that need to be tested. It is therefore imperative that alternatives such as those outlined above are embraced, further developed, accepted and used — as a matter of urgency.

Acknowledgement

This essay, and the original study on which it is based, was funded by the Physicians Committee for Responsible Medicine (PCRM), Washington DC, USA.

References

1. Moxon, M. (2008). Developmental and reproductive toxicity testing: a potted history. *Lab Animal Europe* **8**(3), 3.
2. Bailey, J., Knight, A. & Balcombe, J. (2005). The future of teratology research is *in vitro*. *Biogenic Amines — Stress & Neuroprotection* **19**, 97–145. Available at: http://www.curedisease.net/reports/teratology.pdf (Accessed 14.07.08).
3. Anon. (1980). United States Food and Drug Administration. Caffeine: deletion of GRAS status, proposed declaration that no prior sanction exists, and use on an intern basis pending additional study. *Federal Register* **45**, 69,817.
4. Schardein, J.L. (1993). *Chemically Induced Birth Defects*, 902pp. New York, NY, USA: Marcel Dekker.
5. Anon. (2006). *European Parliament Memo. Q & A on the new chemicals policy, REACH. MEMO/06/488*, 16pp. Brussels, Belgium: European Parliament. Available at: http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/06/488 (Accessed 12.11.08).
6. Kotwani, A., Mehta, V.L. & Iyengar, B. (1995). Aspirin by virtue of its acidic property may act as teratogen in early chick embryo. *Indian Journal of Physiology & Pharmacology* **39**, 131–134.
7. Anon. (2008). *ToxCast™ Program: Predicting Hazard, Characterizing Toxicity Pathways, and Prioritizing the Toxicity Testing of Environmental Chemicals*. Research Triangle Park, NC, USA: National Centre for Computational Toxicology, EPAS. Available at: http://www.epa.gov/comptox/toxcast/ index. html (Accessed 08.10.08).
8. Genschow, E., Scholz, G., Brown, N., Piersma, A., Brady, M., Clemann, N., Huuskonen, H., Paillard, F., Bremer, S., Becker, K. & Spielmann, H. (2000). Development of prediction models for three *in vitro* embryotoxicity tests in an ECVAM validation study. *In Vitro Molecular Toxicology* **13**, 51–66.
9. Genschow, E., Scholz, G., Brown, N.A., Piersma, A.H., Brady, M., Clemann, N., Huuskonen, H., Paillard, F., Bremer, S. & Spielmann, H. (1999). Development of prediction models for three *in vitro* embryotoxicity tests which were evaluated in an ECVAM validation study. *ALTEX* **16**, 73–83.
10. Scholz, G., Pohl, I., Genschow, E., Klemm, M. & Spielmann, H. (1999). Embryotoxicity screening using embryonic stem cells *in vitro*: correlation to *in vivo* teratogenicity. *Cells, Tissues, Organs* **165**, 203–211.
11. Scholz, G., Genschow, E., Pohl, I., Bremer, S., Paparella, M., Raabe, H., Southee, J. & Spielmann, H. (1999). Prevalidation of the Embryonic Stem Cell Test (EST) — A new *in vitro* embryotoxicity test. *Toxicology in Vitro* **13**, 675–681.
12. Spielmann, H., Pohl, I., Döring, B., Liebsch, M. & Moldenhauer, F. (1997). The embryonic stem cell test (EST), an *in vitro* embryotoxicity test using two permanent mouse cell lines: 3T3 fibroblasts and embryonic stem cells. *In Vitro Toxicology* **10**, 119–127.
13. Adler, S., Pellizzer, C., Hareng, L., Hartung, T. & Bremer, S. (2008). First steps in establishing a developmental toxicity test method based on human embryonic stem cells. *Toxicology in Vitro* **22**, 200–211.
14. Adler, S., Lindqvist, J., Uddenberg, K., Hyllner, J. & Strehl, R. (2008). Testing potential developmental toxicants with a cytotoxicity assay based on human embryonic stem cells. *ATLA* **36**, 129–140.
15. Polifka, J.E. & Friedman, J.M. (1999). Clinical teratology: identifying teratogenic risks in humans. *Clinical Genetics* **56**, 409–420.