Regulatory Forum Opinion Piece*: New testing paradigms for reproductive and developmental toxicity - The NTP Modified One generation study and OECD 443

PAUL M.D. FOSTER

Division of the National Toxicology Program, National Institute of Environmental Health Sciences, National Institutes of

Health, Research Triangle Park, North Carolina, USA

*This is an opinion article submitted to the Regulatory Forum and does not constitute an official position of the Society of Toxicologic Pathology or the journal Toxicologic Pathology. The views expressed in this article are those of the author and do not necessarily represent the policies, positions, or opinions of his respective agencies and organizations. The Regulatory Forum is designed to stimulate broad discussion of topics relevant to regulatory issues in toxicologic pathology. Readers of Toxicologic Pathology are encouraged to send their thoughts on these articles or ideas for new topics to regulatoryforum@toxpath.org.

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

The author received no financial support for the research, authorship, and/or publication of this article.

Address correspondence to: Paul M.D. Foster, Division of the National Toxicology Program, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC, 27709, USA; e-mail: foster2@niehs.nih.gov.

Abstract

The NTP has developed a new flexible study design, termed the modified one generation reproduction study (MOG) that will encompass measurements of developmental and reproductive toxicity parameters as well as enable the setting of appropriate dose levels, including evaluation of target organ toxicity, for a cancer bioassay that begins exposure during gestation. This design is compared and contrasted with the new OECD 443 test guideline, the extended one generation reproduction study. The MOG study has a number of advantages, with a focus on F_1 animals, the generation of adequately powered, robust datasets, the provision of both pre-and post-natal developmental toxicity information, the measurement of effects on reproductive structure and function in the same animals and not least, that it does not employ the use of internal triggers in the design structure. The new design is also consistent with the principles of the 3R's.

One of the major roles of the National Toxicology Program (NTP) has been in the development of new toxicology test methods. Following two workshops (King-Herbert and Thayer, 2006, Thayer and Foster, 2007) that focused on the NTP selection of a new rat strain for all of its toxicological studies and that there would be a greater emphasis on early life exposures in the conduct of its cancer bioassays, it became apparent that there would be a need for some dose range finding study that involved early life exposure (gestation, lactation and continuing exposure through adulthood). At the same time NTP had shown that in its developmental and reproductive toxicity (DART) studies that we could markedly increase the power to detect post-natal developmental effects (including those consequent to in utero exposure) by simply retaining more of the offspring from each litter post-weaning (on most DART littering studies, only 1 male and female from each litter is retained), that would normally be culled or only given a cursory examination (Blystone *et al.*, 2010).

Taken together, the Program realized that in performing the necessary setting of dose levels and identification of target organ toxicity in order to undertake a perinatal cancer bioassay, it was possible at the same time to use animals already produced following exposure during gestation and lactation to develop additional, high quality DART information in a single design, which we have termed the modified one generation study (MOG) [http://ntp.niehs.nih.gov/ntp/About_NTP/BSC/2011/April/MOGDesign.pdf]. The design basically commences with time mated rats (although this could easily be adapted to mice) with exposure from gestation day (GD) 6 (implantation) continuously through pregnancy and lactation (with at least 20 pregnant dams per dose group). At weaning, the offspring (normally 4 males and 4 females per litter) are then assigned to different "testing cohorts". One group (10 males and females from different litters) will be continuously dosed for 90 days post weaning for assessment of target organ toxicity and clinical pathology - analogous to a standard 90 day study. In a second group, 1 male and female from each litter will be continuously dose until sexual maturity and bred (non-siblings) and then at GD 21 a routine evaluation of fetuses for external, visceral and skeletal effects would occur (analogous to a teratology study). A third group (1 male and female per litter) would be treated similarly to the second, but after breeding, the females would be allowed to litter and raise their offspring to weaning (and potentially beyond if required). This leaves one final group of approximately 30 male and female offspring that could be used for any other assessment of developmental toxicity (eg developmental neurotoxicity or immunotoxicity). Importantly, this approach allows a 10 week exposure period before mating (ie covers the whole period of rat spermatogenesis) such that any changes in structure of the reproductive organs can be directly correlated with functional outcome (eg fertility and fecundity) in at least 40 breeding pairs with necropsies on all the F_1 offspring for evaluation of developmental effects.

There have been a number of other international efforts to try to improve DART study designs. One effort from ILSI on agrochemical testing attempted to improve life stage testing , in what would be a very data rich toxicology environment with for example a food use pesticide (Cooper *et al.*, 2006) and again endeavored to make the maximum use of the animals already produced. This basic design was then taken up and amended by OECD, but with the idea that this could replace the multigeneration reproduction study and also be used for all chemicals (not just agrochemicals) where the toxicology data

portfolio would be much poorer. One of the major drivers for change was the efforts under REACH in Europe and the realization that approximately 60% of the animals used or produced in toxicity testing on a given agent would come from DART studies (van der Jagt *et al.*, 2004). This effort became the OECD 443, extended one generation test guideline test guideline (OECD, 2011). This design commences with adult male and female rodents (usually rats) which are exposed to test article for 2 weeks prior to mating with sufficient numbers to achieve 20 litters per dose group. This would equate to the period of sperm maturation in the epididymis of male rats. Following continuous exposure of both parents through the period of gestation and lactation the pups are allocated to groups and continuously dosed. 10 males and 10 female per group (from different litters) are assigned for developmental immunotoxicity evaluation, a similar number for evaluation of neurotoxicity at weaning and a third group of 10 males and 10 females from different litters for evaluation of neurotoxicity at sexual maturity. A remaining 40 male and female offspring would be exposed and taken to sexual maturity for necropsy, or if "triggered" half of these animals would be bred (non-siblings) to produce F_2 litters. The developmental neurotoxicity and immunotoxicity arms of the study, originally a requirement in the draft guideline, are now optional.

Offspring that have been exposed in utero and taken to adulthood represent a unique exposure population in toxicology and not to breed these animals routinely is missing a major opportunity to garner toxicity information throughout the reproductive cycle. Moreover, a laboratory will have to plan and cost the use of rooms, equipment and personnel irrespective of whether the trigger will be activated. It is the triggers to breed the F_1 animals that will also produce some significant issues for toxicology laboratories. The guideline states that one of the triggers will be an effect on parental fertility (implantations, pregnancy rate, gestational interval) **in the absence** of a corresponding treatment related reproductive organ histopathology. This implies that all the pathology on reproductive organs of the F_0 parents must be complete and appropriately analyzed before the animals of the F_1 generation reach 90 days of age. The F_0 parents will be necropsied at weaning of the F_1 pups (eg 21 days of age). So when will this data be available to make such a decision? One might anticipate some study stagger (not all the animals will get pregnant at the same time), the necropsy data will need to be collected, tissues fixed, trimmed, embedded, cut, stained, evaluated by a pathologist, statistics conducted, plus any potential discussions with a sponsor or regulatory authority before triggering the breeding - all in less than 10 weeks. This may be possible, but perhaps unlikely, under normal conditions and will certainly be at a cost premium. A second trigger for breeding the F_1 animals is if a finding is made on F_1 developmental landmarks (e.g. anogenital distance, nipple retention, puberty onset). These changes must be dose-related and **in the absence of bodyweight-mediated changes** in these animals. Vaginal opening in rats usually occurs around 32 days after birth, when the animals are in a rapidly growing phase. If one sees an advancement in this parameter in treated groups of say, 4 days, will their body weights be lower? Answer: Yes. Is this due to toxicity, or because they were younger? Similar arguments about study stagger and data collection and statistical analysis (now post-weaning for some of these parameters) would also apply.

I believe the major advantages of the MOG over OECD 443 include:

- In the MOG there is a focus on the F1 generation – the animals exposed from "womb to tomb"; The critical periods of the reproductive life cycle is not assessed in the OED 443 due to the lack of F1 mating and delivery of the F2.
- The MOG HAS No internal triggers – a decision on how to use the animals is made ahead of study start; the use of triggers in the OECD 443 has never been successfuIly accomplished
- The MOG is adequately powered and robust datasets (the use of 10 males and females for immuno- and neurotoxicity assessment in OECD 443 is insufficient except for the most profound effects);
- The MOG Maintains the 10 week pre-breed exposure period – this is based on sound biology of spermatogenesis and would be more important when, for example, a 90 day toxicity study is not available (for many chemicals);
- The MOG includes measuring reproductive structure and function in the same animals. These designs are supposed to provide critical information on reproduction for risk assessment and classification and labeling, and therefore should assess reproductive function in a comprehensive manner;
- The MOG incorporates the provision of other developmental outcome data (pre-natal developmental toxicity, teratology);
- The design of the MOG has the ability to set dose levels for a perinatal cancer bioassay through the generation of target organ toxicity data in offspring following early life exposure.

The MOG is also not without some issues, especially if the test article happens to have very significant developmental toxicity that would preclude having a suitable number of offspring to assign to the various testing cohorts. NTP has tried several options to ameliorate such effects including lowering dose levels during gestation/ lactation versus post weaning; commencing the study at a later time in gestation (eg GD 15 rather than GD6 to overcome early embryonic loss), or deciding that separate studies would be more amenable to study.

In conclusion, the MOG offers a number of very positive advantages over conducting individual DART and range finding studies. Our current experience is with 3 such studies that have now completed their in life portions and are in reporting phase. One of the test agents studied in the MOG was also studied in a direct comparison with the segmented ICH DART designs (using the same stock of rat, diet and dose levels) and in evaluation of the preliminary information (before QA and peer review) has produced completely equivalent data. Thus, we have Refined our toxicity study designs. We can Replace certain other standard toxicity studies by folding them into this particular design and will have Reduced overall animal use compared to conducting individual DART and 90 day toxicity studies.

References

Blystone, C.R., Kissling, G.E., Bishop, J.B., Chapin, R.E., Wolfe, G.W. & Foster, P.M. (2010). Determination of the di-(2-ethylhexyl) phthalate NOAEL for reproductive development in the rat: importance of the

retention of extra animals to adulthood. *Toxicol Sci,* **116,** 640-646.

Cooper, R.L., Lamb, J.C., Barlow, S.M., Bentley, K., Brady, A.M., Doerrer, N.G., Eisenbrandt, D.L., Fenner-Crisp, P.A., Hines, R.N., Irvine, L.F., Kimmel, C.A., Koeter, H., Li, A.A., Makris, S.L., Sheets, L.P., Speijers, G. & Whitby, K.E. (2006). A tiered approach to life stages testing for agricultural chemical safety assessment. *Crit Rev Toxicol,* **36,** 69-98.

King-Herbert, A. & Thayer, K. (2006). NTP workshop: animal models for the NTP rodent cancer bioassay: stocks and strains--should we switch? *Toxicol Pathol,* **34,** 802-805.

OECD (2011). Test No. 443: extended one-generation reproductive toxicity study, OECD guidelines for the testing of chemicals,Section 4: health effects.

Thayer, K.A. & Foster, P.M. (2007). Workgroup report: National Toxicology Program workshop on Hormonally Induced Reproductive Tumors - Relevance of Rodent Bioassays. *Environ Health Perspect,* **115,** 1351-1356.

van der Jagt, K., Munn, S., Tørsløv, J. & de Bruijn, J. (2004). Alternative Approaches can reduce the Use of Test Animals under REACH. Joint Research Centre, European Commission. Available at http://publications.jrc.ec.europa.eu/repository/bitstream/111111111/8790/1/EUR%2021405%20EN.pdf . Accessed 2/21/2014.