Control Performance of Medaka Extended One Generation Test Designs

N. Burden[1], S.G. Lynn[2], V. Mingo[3], C. Mitchell[4], E.R. Salinas[5] and J.R. Wheeler[6]

[1]NC3Rs, London, UK, natalie.burden@nc3rs.org.uk; [2]US EPA, Office of Pesticide Programs, Washington DC, USA*; [3]Corteva Agriscience, München, Germany; [4]HESI, Washington D.C., USA; [5]Bayer AG, Monheim, Germany; [6]Corteva Agriscience, Bergen op Zoom, The Netherlands.

**This poster does not necessarily represent EPA policy.*

NC 3Rs 20 YEARS

Pioneering Better Science

Project aims

1. **Collect and evaluate historical control data for Medaka Extended One Generation Test (MEOGRT) test guideline (TG) studies and studies conducted in the spirit of the MEOGRT (e.g. medaka multigeneration test - MMT), to investigate cross-laboratory and study design differences.**
2. **Form a knowledge base that could be used to improve: a) interpretation of the data; b) test performance and c) potentially the test design.**
3. **Inform on areas for improvement; ensure best use of the *in vivo* data; and aid in evaluating new approaches (e.g. new approach methodologies) that may be alternatives to *in vivo* TGs in future.**

Table 1. Biological validity criteria set out for each generation (F0, F1, F2) in the final OECD TG 240 and/or EPA OCSPP 890.2200 representing the minimum standards for acceptable study performance.
Such criteria are specified to ensure that chemical effects are detectable. EPA OCSPP 890.2200 states that "Failure to meet a single performance criterion, while a warning sign, would in general not be expected to compromise the performance of the entire test. However, failure in several criteria or failure to meet the fecundity criterion could result in the rejection of the test". Criteria related to reproduction are given in **bold**.

Biological validity criteria
Mean **Fecundity** > 20 eggs/pair-day (F0)
Mean **Fecundity** > 20 eggs/pair-day (F1)
Fecundity replicate performance > 65% with > 20 eggs/pair-day (F1)
Mean **Fertility** ≥ 80% (F0)
Mean **Fertility** ≥ 80% (F1)
Mean Hatch ≥ 80% (F1)
Mean Hatch ≥ 80% (F2)
Mean Larval Survival ≥ 80% until 3 weeks post fertilization (F1)
Mean Subadult Survival ≥ 90% from 3 – 12 weeks post fertilization (start of reproduction phase) (F1)
Mean Adult Survival. USEPA only: ≥80% (F0)
Mean Adult Survival. USEPA only: ≥80% (F1)
Mean Subadult Weight (mg). USEPA only: ≥100 mg for each sex (F1)
Mean Subadult Length (mm). USEPA only: ≥20 mm for each sex (F1)
Mean Intersex Rate. USEPA only: XX♂*<5%; XY♀<2% (F1)

Background

- *In vivo* TG assays used to assess the apical effects of chemicals interacting with endocrine pathways can use a substantial number of laboratory animals, and their outcome can lead to regulatory actions.
- They should be sufficiently reliable and robust.
- The MEOGRT (OECD TG 240/EPA OCSPP 890.2200) is a highest-tier *in vivo* assay designed to provide comprehensive data on adverse effects and endocrine-relevant endpoints for key aspects of the fish life cycle. It is conducted using the medaka species (Japanese rice fish, *Oryzias latipes*).
- Currently no set mechanism to review established TGs and assess their utility or performance, but such a review is recommended in the MEOGRT OECD TG.
- Studies following the final published MEOGRT study design were not conducted as part of the validation process. The 12 validation studies were MMTs which include half the number of control replicates vs. the MEOGRT and assessments continue past F2 hatch (see oecd.org/chemicalsafety/testing/MMT-Integrated-Summary-Report.pdf).
- High-quality historical control data can be used to better understand the performance of tests, endpoint relevance, facilitate understanding of variability within a method, and aid interpretation of study data.

Methods

- Control data collated for 25 control groups from 24 independent studies in 9 labs:
  - 3 MMTs including validation studies, 1 OECD TG 229 test extended out until assessment of hatchability in the F2 generation, 9 MEOGRTs, and 1 modified MEOGRT (additional endpoints).
- One MEOGRT study contained water control and solvent control groups, which have been treated independently.
- Data assessed across 17 specified biological validity criteria within OECD TG 240 and/or EPA OCSPP 890.2200 (Table 1) for:
  - Number of control groups for which the relevant data were reported.
  - Number of control groups meeting the relevant validity criterion.
  - Percentage of control groups meeting each validity criterion, where reported. We define a "reasonable" failure rate for a validated study design if a criterion is met in ≥90% of the control groups.
- Assessment conducted for:
  - All 25 control groups.
  - The 14 MMT/MMT-like study control groups (≤ 6 control replicates in each study) vs. 11 MEOGRT study control groups (≥ 12 control replicates in each study).

Results

All control groups

- For over half of the criteria assessed (9/17), the relevant validity criterion was not achieved in at least 90% of the control groups (range 71-89% of control groups meeting the criterion).
- Includes 4/5 criteria relating to reproduction.
- Almost all the criteria with "high" success rates (7/8) relate to growth and survival endpoints.
- Mean of 2 validity criteria failures per control group, and 2 criteria not reported per control group (ranges of 0-6 validity criteria failures and 0-6 criteria not reported, across all the control groups).
- Reproduction endpoints appear to be the most challenging in terms of laboratories' ability to meet the criteria.
- Concerning, considering the special importance of these endpoints for regulatory purposes - which is why the EPA TG states that these specifically need to be met.
- Evidence that even experienced labs are unable to meet all of the validity criteria in the final MEOGRT TG.
- For F1 Mean Intersex Rate: XY♀ criterion, the EPA criterion of <2% intersex is unachievable unless there is zero XY♀ intersex, as the presence of intersex in only 1 of the 24 fish will provide a rate of 4.2% which could explain why this criterion had one of the highest failure rates.

MMT vs. MEOGRT control groups

- The proportion of MEOGRT control groups meeting the relevant validity criterion was higher for almost all criteria vs. MMT.
- For the MMT control groups, 10/17 criteria assessed were not achieved in at least 90% of the control groups (range 40-86% of control groups meeting the criterion), including for all 5 criteria relating to reproduction.
- For the MEOGRT control groups, 4/17 criteria were not achieved in at least 90% of the control groups (range 73-82% of control groups meeting the criterion).
  - Includes the two criteria related to F1 fecundity.
- Fewer validity criteria failures per control group in the MEOGRT control groups vs. MMT control groups (mean of 2 failures for MMT (range 0-6) vs. 1 (range 0-4) for MEOGRT.
- MEOGRT control groups had a higher rate of reported validity criteria (mean of 2 not reported, range 0-4) vs. the MMT studies (mean of 3 not reported, range 0-6).
- Increased replication in the MEOGRT may enhance compliance with biological control performance criteria.
- It is also possible however that the unbalanced replication in the MEOGRT design could lead to a higher incidence of false positive outcomes in the lower replication treatment groups for some endpoints.

Discussion and conclusions

- High likelihood of one or more validity criteria failures, particularly those related to reproduction endpoints increases the potential for studies to be repeated.
- High likelihood of validity criteria failures for MEOGRT studies not ideal considering the animal and financial burden of the studies, and current lead times to have them placed within the few experienced laboratories.
- Further work needed to:
  - Assess which of the many validity criteria are critical to ensure study success and data interpretation.
  - Identify those which are too ambitious and require a revision to ensure they are met >90% of the time.
- Further analysis will now be conducted on the larger dataset, including examination of the effect of lab experience on the success of studies, descriptive statistics for each endpoint and assessment of inter- and intra-laboratory variation, and variance component analysis (within study, between study, between lab, between solvent type where applicable).
- Conducting analyses such as this and demonstrating the issues with data when these TGs are employed in practice will assist all stakeholders (CROs, sponsors and evaluation bodies) to better determine what is realistically achievable and improve interpretation of the data, and so reduce the number of unnecessary repeat studies.

1. US EPA (2015). Endocrine Disruptor Screening Program Test Guidelines. OCSPP 890.2200: Medaka Extended One Generation Reproduction Test (MEOGRT). EPA No. 740-C-15-00, July 2015.
2. OECD (2015). Test No. 240: Medaka Extended One Generation Reproduction Test (MEOGRT), OECD Guidelines for the Testing of Chemicals, Section 2, OECD Publishing, Paris. doi.org/10.1787/9789264242258-en.

Call for data

This dataset largely comprises published/accepted studies. It would be useful to have data from studies that have not been reported (e.g. as part of internal validation exercises, or where there are significant validity criteria failures), as this will give further insight into the issues laboratories may be experiencing.

We would greatly appreciate ANY further MEOGRT/MEOGRT-like data to expand our analysis. We also hope to conduct a similar analysis of OECD TG 234 – Fish sexual development tests and OECD TG 241 – Larval amphibian growth and development assays.

Contact **Natalie Burden** (natalie.burden@nc3rs.org.uk) or **Connie Mitchell** (cmitchell@hesiglobal.org).

Figure 1. Exposure and measurement endpoint timelines within the MEOGRT.
Adapted from Table 1 in OCSPP 890.2200.

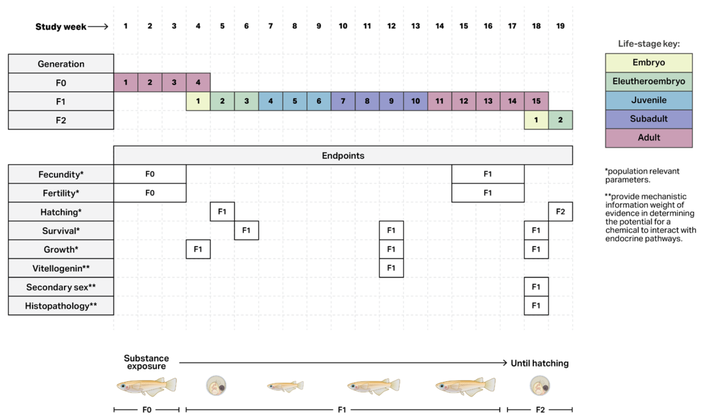

Figure 2. Pooling and repopulating replicates throughout the MEOGRT.
The figure represents one treatment or ½ of a control. Replicated from Annex 7, OECD TG 240.

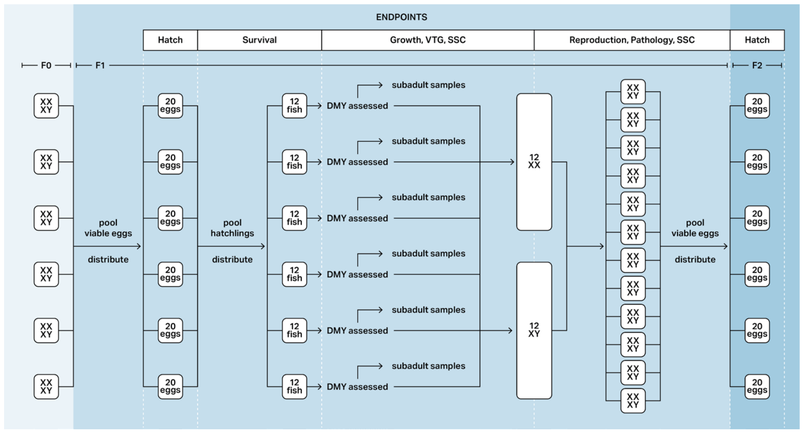

Figure 3. Percentage of control groups meeting each of the 17 biological validity criteria associated with MEOGRT endpoints, where this was reported.
(n=16-25 for "All" control groups depending on criterion) across all groups (green), and split between the 14 MMT control groups (blue) and the 11 MEOGRT control groups (purple). Line at 90% indicates the minimum "reasonable" failure rate for a validated study design for each criterion.