Attachment 1

Follow-on Statistical Analysis on AMA and FSTRA Endpoints Measured in Recent EDSP Tier I Screening Studies

Prepared by:

Battelle
505 King Avenue
Columbus, Ohio 43201

Submitted to:

Dr. Sharlene Matten
Office of Science Coordination and Policy
Exposure Assessment, Coordination, and Policy Division
U.S. Environmental Protection Agency
Washington, DC 20460

June 12, 2019

BATTELLE

Table of Contents

List of Tables

Introduction

Under Task Orders 14 and 17 of EPA Contract EP-W-11-063, EPA commissioned a series of studies in which the 21-day Amphibian Metamorphosis Assay (AMA) (five studies) and the Fish Short Term Reproduction Assay (FSTRA) (16 studies) were applied to identify and characterize the adverse consequences of exposure to EPA-identified test substances for potential endocrine effects. On these studies, Battelle performed visual assessments of monotonicity in the dose-specific replicate and treatment means (within the AMAs) or the replicate means and median of replicate means (within the FSTRAs) for the continuous quantitative endpoints, following guidance provided in Section 5.3.1.3 of OECD (2006).

In response to Technical Directive 2 (dated March 11, 2019), Battelle applied statistical analyses of monotonicity on the continuous quantitative endpoints for these five AMAs and 16 FSTRAs using appropriate statistical methods described in OECD (2006). For each study, Battelle used the data stored in completed Data Entry Spreadsheet Templates (DESTs) as input to the statistical analyses for monotonicity which were presented in a separate monotonicity report (dated June 12, 2019). Incidences of one or more "mismatched" outcomes (or disagreements) between the "visual" and "statistical" monotonicity assessments were noted in all but one of these 21 studies (AMA study #397 had no disagreements).

In response to Technical Directive 2.1 (dated May 23, 2019), Battelle performed follow-on statistical analysis for those endpoints in the AMA and FSTRA studies (except steroid and anal fin papillae endpoints) in which disagreements were noted between the visual and statistical monotonicity assessments. This report presents the results of this follow-on analysis.

Statistical Methods

The statistical analyses applied to the affected endpoints identified those treatment groups whose measurements differed significantly from that of the solvent control group. The analyses were consistent with OPPTS 890.1100 for AMA studies and OPPTS 890.1350 for FSTRA studies. They were generally followed procedures described in the document "Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application." The methods used in a particular application of the statistical analysis were determined based on whether the earlier statistical tests for monotonicity concluded monotonicity or non-monotonicity:

- When a monotonic concentration-response relationship was concluded, the Jonckheere-Terpstra test was applied to the replicate median values in a step-down manner to assess the presence of significant treatment effects.
- When a non-monotonic concentration-response relationship was concluded, the data were evaluated for normality (using the Shapiro-Wilk's test) and for equal variances among treatment levels (using Levene's test).
  - If the hypotheses for normality and/or equal variances were rejected at a 0.05 level, a normalizing, variance stabilizing transformation was applied to the data and the tests repeated.
  - If neither hypothesis (of normality and of equal variance) could be rejected at a 0.05 level, the analysis to compare each treatment group to the solvent control proceeded using Dunnett's test on replicate means.
  - Otherwise, the analysis to compare each treatment group to the solvent control proceeded by applying the Mann-Whitney-Wilcoxon test on the replicate medians (with Bonferroni-Holm adjustment made to the p-values).

Survival data deemed to be monotonic were analyzed using the Cochran-Armitage test. For this test, the survival data were pooled across replicates within a treatment or control group. If monotonicity was not observed, Fisher's Exact test with a Bonferroni-Holm adjustment was performed.

Initially, potential statistical outliers among all continuous quantitative endpoints were assessed by fitting the endpoint data to an analysis of variance (ANOVA) model. The outlier analysis differed between the AMA and FSTRA studies:

- For AMA studies, those observed values whose studentized residuals from the ANOVA exceeded 3 in absolute value were flagged as potential statistical outliers.
- For FSTRA studies, those observed values whose residuals from the ANOVA exceeded the median residual plus three times the residual interquartile range (i.e., the difference between the 75^{th} and 25^{th} percentiles) were flagged as potential statistical outliers.
  - For fecundity and fertilization success endpoints, the ANOVA was fitted to the replicate means.
  - For other endpoints, the ANOVA was fitted to the individual animal measurements.

The statistical analyses were performed both with and without potential outliers included.

In the AMA studies, treatment groups achieving developmental stage 60 and above were given special statistical consideration. After stage 60, tadpoles showed a reduction in size and weight due to tissue resorption and reduction of absolute water content. Thus, measurements of wet weight and snout-to-vent length (SVL) were not appropriate for statistical analysis of differences in growth rates without accounting for the animals' developmental stage. Because an increased number of tadpoles showed development beyond stage 60 (≥20%) in more than one concentration, a mixed ANOVA model with a nested variance structure was fitted to data from all tadpoles to assess growth effects due to chemical treatments while accounting for any effects of late stage development on growth. This ANOVA model contained the following terms:

- Fixed effects for Late Stage ('Yes' if the developmental stage was 61 or greater, or 'No' otherwise),Treatment Group, and their interaction.
- Random effects for "Replicate nested within Treatment Group" and "Animal nested within Replicate".

If the data violated the ANOVA assumptions of normality or equal variance among treatment groups (through analysis of the model residuals as noted earlier), then a normalized rank-order transform was applied to the data. In addition to the standard ANOVA F-tests for the fixed effects of Treatment Group, Late Stage, and their interaction, the interaction F-test was "sliced" into two additional ANOVA F-tests for significant differences among treatment groups within each level of the Late Stage factor ('No' and 'Yes'). Further comparisons of treatment group means to the solvent control were performed at each level of the Late Stage factor. For monotonic concentration response variables, the Jonckheere-Terpstra step-down trend test was used while, for non-monotonic concentration-response, the Mann-Whitney-Wilcoxon test (with Bonferroni-Holm adjustment if the corresponding F-test was not significant) was used.

Results

The monotonicity study report identified 97 endpoints across 20 AMA and FSTRA studies in which disagreements were noted between the "visual" and "statistical" monotonicity assessments. This follow-on statistical analysis was applied to 90 of these endpoints, with seven endpoints corresponding to steroids (testosterone, estradiol) and anal fin papillae process counts being excluded from this analysis. Table 1a through Table 20a present the results of these statistical analyses, with any potential outliers included in the analysis, and with each table specific to one of the 20 AMA or FSTRA studies. Eleven of these 20 tables have a "b" version (e.g., Tables 2b, 3b, 5b) in which results are presented on analyses involving only those data not flagged as a potential outlier. When a "b" table is not presented, no potential statistical outliers were identified for the re-analyzed endpoints.

Discussion

Results of these statistical analysis show that 19 of the 90 endpoints, representing ten of the 20 AMA and FSTRA studies, had different outcomes between the statistical analyses that were driven by a "visual" assessment made on monotonicity (these analyses were performed on Task Orders 14 and 17) and the analyses performed here which were driven by the "statistical" assessment of monotonicity.

Some key findings of differences in statistical analysis results between the two sets of analyses were as follows:

- AMA 387, "Body Weight (g) (Day 21) for NF Stage > 60", both with and without outliers (Tables 2a and 2b)
  - While Treatment Groups 5, 4, 3, and 2 differed significantly from control following the "visual" assessment, only Treatment Groups 5 and 4 differed significantly from control following the "statistical" assessment.
- AMA 387, "HLL (mm) (Day 7)", without outliers (Table 2b)
  - While Treatment Group 3 differed significantly from control following the "visual" assessment, no treatment groups differed significantly from control following the "statistical" assessment.
- AMA 388, "HLL:SVL (Day 21)", both with and without outliers (Tables 3a and 3b)
  - While no treatment groups differed significantly from control following the "visual" assessment, Groups 5 and 4 differed significantly from control following the "statistical" assessment.
- FSTRA 411, "Female Blood Plasma VTG", both with and without outliers (Tables 5a and 5b)
  - While Treatment Groups 5, 4, 3, and 2 differed significantly from control following the "visual" assessment, only Groups 5 and 4 differed significantly from control following the "statistical" assessment.
- FSTRA 6126, "Fertilization Success", all data (with outliers) (Table 8a)
  - While no treatment groups differed significantly from control following the "visual" assessment, Groups 5, 4, and 3 differed significantly from control following the "statistical" assessment.

- FSTRA 6128, “Male Blood Plasma VTG”, all data (with outliers) (Table 9a)
  - While Treatment Groups 5 and 3 differed significantly from control following the “visual” assessment, Groups 5, 4, and 3 differed significantly from control following the “statistical” assessment.
- FSTRA 405, “Male Blood Plasma VTG (ng/mL)”, all data (with outliers) (Table 12a)
  - While Treatment Groups 5 and 4 differed significantly from control following the “visual” assessment, only Group 5 differed significantly from control following the “statistical” assessment.
- FSTRA 405, “Female Body Weight (mg)”, all data (with outliers) (Table 12a)
  - While Treatment Group 5 differed significantly from control following the “visual” assessment, Groups 5 and 4 differed significantly from control following the “statistical” assessment.
- FSTRA 405, “Male Body Weight (mg)”, all data (with outliers) (Table 12a)
  - While no treatment groups differed significantly from control following the “visual” assessment, Group 5 differed significantly from control following the “statistical” assessment.
- FSTRA 406, “Male Body Weight (mg)”, all data (with outliers) (Table 13a)
  - While Treatment Groups 5, 3, and 2 differed significantly from control following the “visual” assessment, no treatment group differed significantly from control following the “statistical” assessment.
- FSTRA 412, “Male GSI” and “Male Body Length (mm)”, all data (with outliers) (Table 15a)
  - While no treatment groups differed significantly from control following the “visual” assessment, Group 5 differed significantly from control following the “statistical” assessment.
- FSTRA 412, “Female Gonad Weight (mg)”, all data (with outliers) (Table 15a)
  - While no treatment groups differed significantly from control following the “visual” assessment, Groups 2, 3, and 4 differed significantly from control following the “statistical” assessment.
- FSTRA 105, “Fertilization Success”, all data (with outliers) (Table 17a)
  - While Treatment Group 4 differed significantly from control following the “visual” assessment, Groups 5 and 4 differed significantly from control following the “statistical” assessment.
- FSTRA 106, “Fertilization Success”, all data (with outliers) (Table 18a)
  - While Treatment Group 2 differed significantly from control following the “visual” assessment, no treatment groups differed significantly from control following the “statistical” assessment.
- FSTRA 106, “Male Body Weight (mg)”, all data (with outliers) (Table 18a)
  - While no treatment groups differed significantly from control following the “visual” assessment, Group 5 differed significantly from control following the “statistical” assessment.

References

OPPTS. 890.1100. Amphibian Metamorphosis (Frog), United States Environmental Protection Agency, Washington DC, EPA 740-C-09-002, October 2009.

OPPTS. 890.1350. Fish Short-Term Reproduction Assay, United States Environmental Protection Agency, Washington DC, EPA 740-C-09-007, October 2009.

Organization for Economic Co-operation and Development (OECD). 2006. Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application. Environmental Health and Safety Publications. Series on Testing and Assessment, No. 54, Paris, France.

Table 1a. AMA 386 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
HLL (mm) (Day 7)	Monotonic	NP	NP	NS	NP
HLL:SVL (Day 7)	Monotonic	NP	NP	NS	NP

1. Monotonicity was assessed by fitting a mixed effects ANOVA model on rank transformed data and determining linear and quadratic contrasts. The data were determined to be non-monotonic if the linear contrast was not significant and the quadratic contrast was significant. Otherwise the data were determined to be monotonic.
2. Shapiro-Wilk test for normality.
3. Levene's Test for homogeneity of variance.
4. Jonckheere-Terpstra step-down test was performed on monotonic concentration response data. Only statistically significant treatment trends were listed.
5. Control is group 1, and test concentration groups are groups 2 to 5. Only statistically significant pairwise comparisons were listed.
    - For non-monotonic, normally distributed, and homogeneous variance data, Dunnett's test was used in the pairwise comparisons to control.
    - For other types of data (e.g., non-monotonic, normally distributed, and heterogeneous variance data), Mann-Whitney-Wilcoxon test was used in the pairwise comparisons to control with Bonferroni-Holm multiple comparison adjustment.

NP Statistical test or comparison was not performed.

NS No statistically significant differences were found for Jonckheere-Terpstra step-down test, Dunnett's tests, or Mann-Whitney-Wilcoxon tests after Bonferroni-Holm adjustment.

HLL = hind limb length; SVL = snout-to-vent body length.

[There were no endpoints with monotonicity disagreement with potential statistical outliers removed.]

Table 2a. AMA 387 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
HLL:SVL (Day 7)	Monotonic	NP	NP	NS	NP
Body Weight (g) (Day 21) for NF Stage > 60	Monotonic	NP	NP	Group 5 (0.0035) Group 4 (0.0317)	NP

Table 2b. AMA 387 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement, with Potential Statistical Outliers Removed.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
HLL (mm) (Day 7)	Monotonic	NP	NP	NS	NP
Body Weight (g) (Day 21) for NF Stage > 60	Monotonic	NP	NP	Group 5 (0.0035) Group 4 (0.0317)	NP

1. Monotonicity was assessed by fitting a mixed effects ANOVA model on rank transformed data and determining linear and quadratic contrasts. The data were determined to be non-monotonic if the linear contrast was not significant and the quadratic contrast was significant. Otherwise the data were determined to be monotonic.
2. Shapiro-Wilk test for normality.
3. Levene's Test for homogeneity of variance.
4. Jonckheere-Terpstra step-down test was performed on monotonic concentration response data. Only statistically significant treatment trends were listed.
5. Control is group 1 and test concentration groups are groups 2 to 5. Only statistically significant pairwise comparisons were listed.
    - For non-monotonic, normally distributed, and homogeneous variance data, Dunnett's test was used in the pairwise comparisons to control.
    - For other types of data (e.g., non-monotonic, normally distributed, and heterogeneous variance data), Mann-Whitney-Wilcoxon test was used in the pairwise comparisons to control with Bonferroni-Holm multiple comparison adjustment.

NP Statistical test or comparison was not performed.

NS No statistically significant differences were found for Jonckheere-Terpstra step-down test, Dunnett's tests, or Mann-Whitney-Wilcoxon tests after Bonferroni-Holm adjustment.

HLL = hind limb length; SVL = snout-to-vent body length; NF = Nieuwkoop and Faber.

Table 3a. AMA 388 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
HLL (mm) (Day 21)	Monotonic	NP	NP	NS	NS
SVL (mm) (Day 7)	Monotonic	NP	NP	NS	NP
HLL:SVL (Day 7)	Monotonic	NP	NP	NS	NP
HLL:SVL (Day 21)	Monotonic	NP	NP	Group 5 (0.0418) Group 4 (0.0274)	NP

Table 3b. AMA 388 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement, with Potential Statistical Outliers Removed.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
HLL (mm) (Day 21)	Monotonic	NP	NP	NS	NP
HLL:SVL (Day 21)	Monotonic	NP	NP	Group 5 (0.0418) Group 4 (0.0274)	NP

1. Monotonicity was assessed by fitting a mixed effects ANOVA model on rank transformed data and determining linear and quadratic contrasts. The data were determined to be non-monotonic if the linear contrast was not significant and the quadratic contrast was significant. Otherwise the data were determined to be monotonic.
2. Shapiro-Wilk test for normality.
3. Levene's Test for homogeneity of variance.
4. Jonckheere-Terpstra step-down test was performed on monotonic concentration response data. Only statistically significant treatment trends were listed.
5. Control is group 1 and test concentration groups are groups 2 to 5. Only statistically significant pairwise comparisons were listed.
    - For non-monotonic, normally distributed, and homogeneous variance data, Dunnett's test was used in the pairwise comparisons to control.
    - For other types of data (e.g., non-monotonic, normally distributed, and heterogeneous variance data), Mann-Whitney-Wilcoxon test was used in the pairwise comparisons to control with Bonferroni-Holm multiple comparison adjustment.

NP Statistical test or comparison was not performed.

NS No statistically significant differences were found for Jonckheere-Terpstra step-down test, Dunnett's tests, or Mann-Whitney-Wilcoxon tests after Bonferroni-Holm adjustment.

HLL = hind limb length; SVL = snout-to-vent body length.

Table 4a. AMA 389 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
HLL:SVL (Day 7)	Monotonic	NP	NP	NS	NP

1. Monotonicity was assessed by fitting a mixed effects ANOVA model on rank transformed data and determining linear and quadratic contrasts. The data were determined to be non-monotonic if the linear contrast was not significant and the quadratic contrast was significant. Otherwise the data were determined to be monotonic.
2. Shapiro-Wilk test for normality.
3. Levene's Test for homogeneity of variance.
4. Jonckheere-Terpstra step-down test was performed on monotonic concentration response data. Only statistically significant treatment trends were listed.
5. Control is group 1 and test concentration groups are groups 2 to 5. Only statistically significant pairwise comparisons were listed.
    - For non-monotonic, normally distributed, and homogeneous variance data, Dunnett's test was used in the pairwise comparisons to control.
    - For other types of data (e.g., non-monotonic, normally distributed, and heterogeneous variance data), Mann-Whitney-Wilcoxon test was used in the pairwise comparisons to control with Bonferroni-Holm multiple comparison adjustment.

NP Statistical test or comparison was not performed.

NS No statistically significant differences were found for Jonckheere-Terpstra step-down test, Dunnett's tests, or Mann-Whitney-Wilcoxon tests after Bonferroni-Holm adjustment.

HLL = hind limb length; SVL = snout-to-vent body length.

[No potential statistical outliers were identified for this study.]

Table 5a. FSTRA 411 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
Female Blood Plasma VTG	Monotonic	NP	NP	Group 5 (0.0002) Group 4 (0.0021)	NP
Female Body Weight (g)	Monotonic	NP	NP	NS	NP

Table 5b. FSTRA 411 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement, with Potential Statistical Outliers Removed.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
Female Blood Plasma VTG	Monotonic	NP	NP	Group 5 (0.0002) Group 4 (0.0021)	NP
Female Body Weight (g)	Monotonic	NP	NP	NS	NP

1. Monotonicity was assessed by fitting a mixed effects ANOVA model on rank transformed data and determining linear and quadratic contrasts. The data were determined to be non-monotonic if the linear contrast was not significant and the quadratic contrast was significant. Otherwise the data were determined to be monotonic.
2. Shapiro-Wilk test for normality.
3. Levene's Test for homogeneity of variance.
4. Jonckheere-Terpstra step-down test was performed on monotonic concentration response data. Only statistically significant treatment trends were listed.
5. Control is group 1 and test concentration groups are groups 2 to 5. Only statistically significant pairwise comparisons were listed.
    - For non-monotonic, normally distributed, and homogeneous variance data, Dunnett's test was used in the pairwise comparisons to control.
    - For other types of data (e.g., non-monotonic, normally distributed, and heterogeneous variance data), Mann-Whitney-Wilcoxon test was used in the pairwise comparisons to control with Bonferroni-Holm multiple comparison adjustment.

NP Statistical test or comparison was not performed.

NS No statistically significant differences were found for Jonckheere-Terpstra step-down test, Dunnett's tests, or Mann-Whitney-Wilcoxon tests after Bonferroni-Holm adjustment.

[Note that Female/Male Testosterone and Female/Male Estradiol (steroid analyses) were not included in this statistical analysis task.]

Table 6a. FSTRA 6120 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
Fecundity	Monotonic	NP	NP	NS	NP
Fertilization Success	Monotonic	NP	NP	NS	NP
Female Blood Plasma VTG	Monotonic	NP	NP	NS	NP
Male Body Weight (g)	Monotonic	NP	NP	NS	NP
Survival	Monotonic	NP	NP	NS[6]	NP

Table 6b. FSTRA 6120 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement, with Potential Statistical Outliers Removed.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
Female Blood Plasma VTG	Monotonic	NP	NP	NS	NP

1. Monotonicity was assessed by fitting a mixed effects ANOVA model on rank transformed data and determining linear and quadratic contrasts. The data were determined to be non-monotonic if the linear contrast was not significant and the quadratic contrast was significant. Otherwise the data were determined to be monotonic.
2. Shapiro-Wilk test for normality.
3. Levene's Test for homogeneity of variance.
4. Jonckheere-Terpstra step-down test was performed on monotonic concentration response data. Only statistically significant treatment trends were listed.
5. Control is group 1 and test concentration groups are groups 2 to 5. Only statistically significant pairwise comparisons were listed.
   - For non-monotonic, normally distributed, and homogeneous variance data, Dunnett's test was used in the pairwise comparisons to control.
   - For other types of data (e.g., non-monotonic, normally distributed, and heterogeneous variance data), Mann-Whitney-Wilcoxon test was used in the pairwise comparisons to control with Bonferroni-Holm multiple comparison adjustment.
6. Step-down Cochran-Armitage test was performed on survival.

NP Statistical test or comparison was not performed.

NS No statistically significant differences were found for Jonckheere-Terpstra step-down test, Dunnett's tests, or Mann-Whitney-Wilcoxon tests after Bonferroni-Holm adjustment.

Table 7a. FSTRA 6123 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
Female Blood Plasma VTG	Monotonic	NP	NP	NS	NP
Male GSI	Non-Monotonic	Non-Normal	Homogeneous	NP	NS

Table 7b. FSTRA 6123 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement, with Potential Statistical Outliers Removed.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
Female Blood Plasma VTG	Monotonic	NP	NP	NS	NP

1. Monotonicity was assessed by fitting a mixed effects ANOVA model on rank transformed data and determining linear and quadratic contrasts. The data were determined to be non-monotonic if the linear contrast was not significant and the quadratic contrast was significant. Otherwise the data were determined to be monotonic.
2. Shapiro-Wilk test for normality.
3. Levene's Test for homogeneity of variance.
4. Jonckheere-Terpstra step-down test was performed on monotonic concentration response data. Only statistically significant treatment trends were listed.
5. Control is group 1 and test concentration groups are groups 2 to 5. Only statistically significant pairwise comparisons were listed.
    - For non-monotonic, normally distributed, and homogeneous variance data, Dunnett's test was used in the pairwise comparisons to control.
    - For other types of data (e.g., non-monotonic, normally distributed, and heterogeneous variance data), Mann-Whitney-Wilcoxon test was used in the pairwise comparisons to control with Bonferroni-Holm multiple comparison adjustment.

NP Statistical test or comparison was not performed.

NS No statistically significant differences were found for Jonckheere-Terpstra step-down test, Dunnett's tests, or Mann-Whitney-Wilcoxon tests after Bonferroni-Holm adjustment.

Table 8a. FSTRA 6126 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
Fertilization Success	Monotonic	NP	NP	Group 5 (0.0395) Group 4 (0.0399) Group 3 (0.0281)	NP
Female Blood Plasma VTG	Monotonic	NP	NP	NS	NP
Male GSI	Monotonic	NP	NP	NS	NP

1. Monotonicity was assessed by fitting a mixed effects ANOVA model on rank transformed data and determining linear and quadratic contrasts. The data were determined to be non-monotonic if the linear contrast was not significant and the quadratic contrast was significant. Otherwise the data were determined to be monotonic.
2. Shapiro-Wilk test for normality.
3. Levene's Test for homogeneity of variance.
4. Jonckheere-Terpstra step-down test was performed on monotonic concentration response data. Only statistically significant treatment trends were listed.
5. Control is group 1 and test concentration groups are groups 2 to 5. Only statistically significant pairwise comparisons were listed.
   - For non-monotonic, normally distributed, and homogeneous variance data, Dunnett's test was used in the pairwise comparisons to control.
   - For other types of data (e.g., non-monotonic, normally distributed, and heterogeneous variance data), Mann-Whitney-Wilcoxon test was used in the pairwise comparisons to control with Bonferroni-Holm multiple comparison adjustment.

NP Statistical test or comparison was not performed.

NS No statistically significant differences were found for Jonckheere-Terpstra step-down test, Dunnett's tests, or Mann-Whitney-Wilcoxon tests after Bonferroni-Holm adjustment.

[There were no endpoints with monotonicity disagreement with potential statistical outliers removed.]

Table 9a. FSTRA 6128 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
Fertilization Success	Monotonic	NP	NP	NS	NP
Female Blood Plasma VTG	Monotonic	NP	NP	NS	NP
Male Blood Plasma VTG	Monotonic	NP	NP	Group 5 (0.0237) Group 4 (0.0242) Group 3 (0.0149)	NP
Male GSI	Monotonic	NP	NP	NS	NP
Survival	Monotonic	NP	NP	NS[6]	NP

1. Monotonicity was assessed by fitting a mixed effects ANOVA model on rank transformed data and determining linear and quadratic contrasts. The data were determined to be non-monotonic if the linear contrast was not significant and the quadratic contrast was significant. Otherwise the data were determined to be monotonic.
2. Shapiro-Wilk test for normality.
3. Levene's Test for homogeneity of variance.
4. Jonckheere-Terpstra step-down test was performed on monotonic concentration response data. Only statistically significant treatment trends were listed.
5. Control is group 1 and test concentration groups are groups 2 to 5. Only statistically significant pairwise comparisons were listed.
   - For non-monotonic, normally distributed, and homogeneous variance data, Dunnett's test was used in the pairwise comparisons to control.
   - For other types of data (e.g., non-monotonic, normally distributed, and heterogeneous variance data), Mann-Whitney-Wilcoxon test was used in the pairwise comparisons to control with Bonferroni-Holm multiple comparison adjustment.
6. Step-down Cochran-Armitage test was performed on survival.

NP Statistical test or comparison was not performed.

NS No statistically significant differences were found for Jonckheere-Terpstra step-down test, Dunnett's tests, or Mann-Whitney-Wilcoxon tests after Bonferroni-Holm adjustment.

[No potential statistical outliers were identified for this study.]

Table 10a. FSTRA 6131 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
Female Blood Plasma VTG	Monotonic	NP	NP	NS	NP
Female GSI	Monotonic	NP	NP	NS	NP
Male Body Weight (g)	Monotonic	NP	NP	NS	NP

1. Monotonicity was assessed by fitting a mixed effects ANOVA model on rank transformed data and determining linear and quadratic contrasts. The data were determined to be non-monotonic if the linear contrast was not significant and the quadratic contrast was significant. Otherwise the data were determined to be monotonic.
2. Shapiro-Wilk test for normality.
3. Levene's Test for homogeneity of variance.
4. Jonckheere-Terpstra step-down test was performed on monotonic concentration response data. Only statistically significant treatment trends were listed.
5. Control is group 1 and test concentration groups are groups 2 to 5. Only statistically significant pairwise comparisons were listed.
   - For non-monotonic, normally distributed, and homogeneous variance data, Dunnett's test was used in the pairwise comparisons to control.
   - For other types of data (e.g., non-monotonic, normally distributed, and heterogeneous variance data), Mann-Whitney-Wilcoxon test was used in the pairwise comparisons to control with Bonferroni-Holm multiple comparison adjustment.

NP Statistical test or comparison was not performed.

NS No statistically significant differences were found for Jonckheere-Terpstra step-down test, Dunnett's tests, or Mann-Whitney-Wilcoxon tests after Bonferroni-Holm adjustment.

[No potential statistical outliers were identified for this study.]

Table 11a. FSTRA 404 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
Female Liver VTG mRNA	Monotonic	NP	NP	NS	NP
Male GSI	Monotonic	NP	NP	NS	NP
Female Body Weight (mg)	Monotonic	NP	NP	NS	NP
Male Body Weight (mg)	Monotonic	NP	NP	NS	NP
Female Body Length (mm)	Monotonic	NP	NP	NS	NP
Male Gonad Weight (mg)	Monotonic	NP	NP	NS	NP

Table 11b. FSTRA 404 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement, with Potential Statistical Outliers Removed.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
Female Body Weight (mg)	Monotonic	NP	NP	NS	NP
Male Body Weight (mg)	Monotonic	NP	NP	NS	NP
Female Gonad Weight (mg)	Monotonic	NP	NP	NS	NP

1. Monotonicity was assessed by fitting a mixed effects ANOVA model on rank transformed data and determining linear and quadratic contrasts. The data were determined to be non-monotonic if the linear contrast was not significant and the quadratic contrast was significant. Otherwise the data were determined to be monotonic.
2. Shapiro-Wilk test for normality.
3. Levene's Test for homogeneity of variance.
4. Jonckheere-Terpstra step-down test was performed on monotonic concentration response data. Only statistically significant treatment trends were listed.
5. Control is group 1 and test concentration groups are groups 2 to 5. Only statistically significant pairwise comparisons were listed.
    - For non-monotonic, normally distributed, and homogeneous variance data, Dunnett's test was used in the pairwise comparisons to control.
    - For other types of data (e.g., non-monotonic, normally distributed, and heterogeneous variance data), Mann-Whitney-Wilcoxon test was used in the pairwise comparisons to control with Bonferroni-Holm multiple comparison adjustment.

NP Statistical test or comparison was not performed.

NS No statistically significant differences were found for Jonckheere-Terpstra step-down test, Dunnett's tests, or Mann-Whitney-Wilcoxon tests after Bonferroni-Holm adjustment.

[Note Male Papillary Processing Score was not included in this statistical analysis task.]

Table 12a. FSTRA 405 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
Male Blood Plasma VTG (ng/mL)	Monotonic	NP	NP	Group 5 (0.0096)	NP
Male Liver VTG (ng/µL)	Monotonic	NP	NP	NS	NP
Female GSI (%)	Monotonic	NP	NP	Group 5 (0.0052)	NP
Female Body Weight (mg)	Monotonic	NP	NP	Group 5 (0.0260) Group 4 (0.0499)	NP
Male Body Weight (mg)	Monotonic	NP	NP	Group 5 (0.0395)	NP

Table 12b. FSTRA 405 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement, with Potential Statistical Outliers Removed.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
Male GSI (%)	Monotonic	NP	NP	Group 5 (0.0296)	NP

1. Monotonicity was assessed by fitting a mixed effects ANOVA model on rank transformed data and determining linear and quadratic contrasts. The data were determined to be non-monotonic if the linear contrast was not significant and the quadratic contrast was significant. Otherwise the data were determined to be monotonic.
2. Shapiro-Wilk test for normality.
3. Levene's Test for homogeneity of variance.
4. Jonckheere-Terpstra step-down test was performed on monotonic concentration response data. Only statistically significant treatment trends were listed.
5. Control is group 1 and test concentration groups are groups 2 to 5. Only statistically significant pairwise comparisons were listed.
    - For non-monotonic, normally distributed, and homogeneous variance data, Dunnett's test was used in the pairwise comparisons to control.
    - For other types of data (e.g., non-monotonic, normally distributed, and heterogeneous variance data), Mann-Whitney-Wilcoxon test was used in the pairwise comparisons to control with Bonferroni-Holm multiple comparison adjustment.

NP Statistical test or comparison was not performed.

NS No statistically significant differences were found for Jonckheere-Terpstra step-down test, Dunnett's tests, or Mann-Whitney-Wilcoxon tests after Bonferroni-Holm adjustment.

Table 13a. FSTRA 406 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
Female Liver VTG mRNA	Monotonic	NP	NP	NS	NP
Male Liver VTG mRNA	Monotonic	NP	NP	NS	NP
Female GSI	Monotonic	NP	NP	NS	NP
Male GSI	Monotonic	NP	NP	NS	NP
Female Body Weight (mg)	Monotonic	NP	NP	NS	NP
Male Body Weight (mg)	Monotonic	NP	NP	NS	NP
Female Gonad Weight (mg)	Monotonic	NP	NP	NS	NP
Male Gonad Weight (mg)	Monotonic	NP	NP	NS	NP

Table 13b. FSTRA 406 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement, with Potential Statistical Outliers Removed.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
Male GSI	Monotonic	NP	NP	NS	NP
Male Gonad Weight	Monotonic	NP	NP	NS	NP

1. Monotonicity was assessed by fitting a mixed effects ANOVA model on rank transformed data and determining linear and quadratic contrasts. The data were determined to be non-monotonic if the linear contrast was not significant and the quadratic contrast was significant. Otherwise the data were determined to be monotonic.
2. Shapiro-Wilk test for normality.
3. Levene's Test for homogeneity of variance.
4. Jonckheere-Terpstra step-down test was performed on monotonic concentration response data. Only statistically significant treatment trends were listed.
5. Control is group 1 and test concentration groups are groups 2 to 5. Only statistically significant pairwise comparisons were listed.
    - For non-monotonic, normally distributed, and homogeneous variance data, Dunnett's test was used in the pairwise comparisons to control.
    - For other types of data (e.g., non-monotonic, normally distributed, and heterogeneous variance data), Mann-Whitney-Wilcoxon test was used in the pairwise comparisons to control with Bonferroni-Holm multiple comparison adjustment.

NP Statistical test or comparison was not performed.

NS No statistically significant differences were found for Jonckheere-Terpstra step-down test, Dunnett's tests, or Mann-Whitney-Wilcoxon tests after Bonferroni-Holm adjustment.

[Note Male Papillary Processing Score was not included in this statistical analysis task.]

Table 14a. FSTRA 407 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
Male Liver VTG (ng/µL)	Non-Monotonic	Non-Normal	Homogeneous	NP	NS
Female GSI (%)	Monotonic	NP	NP	NS	NP

Table 14b. FSTRA 407 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement, with Potential Statistical Outliers Removed.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
Male Liver VTG (ng/µL)	Non-Monotonic	Non-Normal	Homogeneous	NP	NS

1. Monotonicity was assessed by fitting a mixed effects ANOVA model on rank transformed data and determining linear and quadratic contrasts. The data were determined to be non-monotonic if the linear contrast was not significant and the quadratic contrast was significant. Otherwise the data were determined to be monotonic.
2. Shapiro-Wilk test for normality.
3. Levene's Test for homogeneity of variance.
4. Jonckheere-Terpstra step-down test was performed on monotonic concentration response data. Only statistically significant treatment trends were listed.
5. Control is group 1 and test concentration groups are groups 2 to 5. Only statistically significant pairwise comparisons were listed.
    - For non-monotonic, normally distributed, and homogeneous variance data, Dunnett's test was used in the pairwise comparisons to control.
    - For other types of data (e.g., non-monotonic, normally distributed, and heterogeneous variance data), Mann-Whitney-Wilcoxon test was used in the pairwise comparisons to control with Bonferroni-Holm multiple comparison adjustment.

NP Statistical test or comparison was not performed.

NS No statistically significant differences were found for Jonckheere-Terpstra step-down test, Dunnett's tests, or Mann-Whitney-Wilcoxon tests after Bonferroni-Holm adjustment.

Table 15a. FSTRA 412 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
Male GSI	Monotonic	NP	NP	Group 5 (0.0142)	NP
Female Body Length (mm)	Monotonic	NP	NP	NS	NP
Male Body Length (mm)	Monotonic	NP	NP	Group 5 (0.0385)	NP
Female Gonad Weight (mg)	Non-Monotonic	Normal	Homogeneous	NP	Group 2 (0.0101) Group 3 (0.0258) Group 4 (0.0294)

1. Monotonicity was assessed by fitting a mixed effects ANOVA model on rank transformed data and determining linear and quadratic contrasts. The data were determined to be non-monotonic if the linear contrast was not significant and the quadratic contrast was significant. Otherwise the data were determined to be monotonic.
2. Shapiro-Wilk test for normality.
3. Levene's Test for homogeneity of variance.
4. Jonckheere-Terpstra step-down test was performed on monotonic concentration response data. Only statistically significant treatment trends were listed.
5. Control is group 1 and test concentration groups are groups 2 to 5. Only statistically significant pairwise comparisons were listed.
    - For non-monotonic, normally distributed, and homogeneous variance data, Dunnett's test was used in the pairwise comparisons to control.
    - For other types of data (e.g., non-monotonic, normally distributed, and heterogeneous variance data), Mann-Whitney-Wilcoxon test was used in the pairwise comparisons to control with Bonferroni-Holm multiple comparison adjustment.

NP Statistical test or comparison was not performed.

NS No statistically significant differences were found for Jonckheere-Terpstra step-down test, Dunnett's tests, or Mann-Whitney-Wilcoxon tests after Bonferroni-Holm adjustment.

[There were no endpoints with monotonicity disagreement with potential statistical outliers removed.]

[Note Male Papillary Processing Score was not included in this statistical analysis task.]

Table 16a. FSTRA 104 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
Female Liver VTG (ng/µL)	Monotonic	NP	NP	NS	NP
Male Liver VTG (ng/µL)	Monotonic	NP	NP	NS	NP
Male Body Length (mm)	Monotonic	NP	NP	NS	NP

Table 16b. FSTRA 104 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement, with Potential Statistical Outliers Removed.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
Female Liver VTG (ng/µL)	Monotonic	NP	NP	NS	NP

1. Monotonicity was assessed by fitting a mixed effects ANOVA model on rank transformed data and determining linear and quadratic contrasts. The data were determined to be non-monotonic if the linear contrast was not significant and the quadratic contrast was significant. Otherwise the data were determined to be monotonic.
2. Shapiro-Wilk test for normality.
3. Levene's Test for homogeneity of variance.
4. Jonckheere-Terpstra step-down test was performed on monotonic concentration response data. Only statistically significant treatment trends were listed.
5. Control is group 1 and test concentration groups are groups 2 to 5. Only statistically significant pairwise comparisons were listed.
    - For non-monotonic, normally distributed, and homogeneous variance data, Dunnett's test was used in the pairwise comparisons to control.
    - For other types of data (e.g., non-monotonic, normally distributed, and heterogeneous variance data), Mann-Whitney-Wilcoxon test was used in the pairwise comparisons to control with Bonferroni-Holm multiple comparison adjustment.

NP Statistical test or comparison was not performed.

NS No statistically significant differences were found for Jonckheere-Terpstra step-down test, Dunnett's tests, or Mann-Whitney-Wilcoxon tests after Bonferroni-Holm adjustment.

Table 17a. FSTRA 105 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
Fertilization Success	Monotonic	NP	NP	Group 5 (0.0284) Group 4 (0.0068)	NP
Female GSI (%)	Monotonic	NP	NP	NS	NP
Male Body Weight (mg)	Monotonic	NP	NP	NS	NP
Male Body Length (mm)	Monotonic	NP	NP	NS	NP
Female Gonad Weight (mg)	Monotonic	NP	NP	NS	NP
Male Gonad Weight (mg)	Monotonic	NP	NP	NS	NP

1. Monotonicity was assessed by fitting a mixed effects ANOVA model on rank transformed data and determining linear and quadratic contrasts. The data were determined to be non-monotonic if the linear contrast was not significant and the quadratic contrast was significant. Otherwise the data were determined to be monotonic.
2. Shapiro-Wilk test for normality.
3. Levene's Test for homogeneity of variance.
4. Jonckheere-Terpstra step-down test was performed on monotonic concentration response data. Only statistically significant treatment trends were listed.
5. Control is group 1 and test concentration groups are groups 2 to 5. Only statistically significant pairwise comparisons were listed.
   - For non-monotonic, normally distributed, and homogeneous variance data, Dunnett's test was used in the pairwise comparisons to control.
   - For other types of data (e.g., non-monotonic, normally distributed, and heterogeneous variance data), Mann-Whitney-Wilcoxon test was used in the pairwise comparisons to control with Bonferroni-Holm multiple comparison adjustment.

NP Statistical test or comparison was not performed.

NS No statistically significant differences were found for Jonckheere-Terpstra step-down test, Dunnett's tests, or Mann-Whitney-Wilcoxon tests after Bonferroni-Holm adjustment.

[There were no endpoints with monotonicity disagreement with potential statistical outliers removed.]

Table 18a. FSTRA 106 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
Fertilization Success	Monotonic	NP	NP	NS	NP
Female Blood Plasma VTG (µg/mL)	Monotonic	NP	NP	NS	NP
Male Blood Plasma VTG (ng/mL)	Monotonic	NP	NP	NS	NP
Female GSI (%)	Monotonic	NP	NP	NS	NP
Male GSI (%)	Monotonic	NP	NP	NS	NP
Female Body Weight (mg)	Monotonic	NP	NP	NS	NP
Male Body Weight (mg)	Monotonic	NP	NP	Group 5 (0.0500)	NP
Female Body Length (mm)	Monotonic	NP	NP	NS	NP
Female Gonad Weight (mg)	Monotonic	NP	NP	NS	NP
Male Gonad Weight (mg)	Monotonic	NP	NP	NS	NP

1. Monotonicity was assessed by fitting a mixed effects ANOVA model on rank transformed data and determining linear and quadratic contrasts. The data were determined to be non-monotonic if the linear contrast was not significant and the quadratic contrast was significant. Otherwise the data were determined to be monotonic.
2. Shapiro-Wilk test for normality.
3. Levene's Test for homogeneity of variance.
4. Jonckheere-Terpstra step-down test was performed on monotonic concentration response data. Only statistically significant treatment trends were listed.
5. Control is group 1 and test concentration groups are groups 2 to 5. Only statistically significant pairwise comparisons were listed.
    - For non-monotonic, normally distributed, and homogeneous variance data, Dunnett's test was used in the pairwise comparisons to control.
    - For other types of data (e.g., non-monotonic, normally distributed, and heterogeneous variance data), Mann-Whitney-Wilcoxon test was used in the pairwise comparisons to control with Bonferroni-Holm multiple comparison adjustment.

NP Statistical test or comparison was not performed.

NS No statistically significant differences were found for Jonckheere-Terpstra step-down test, Dunnett's tests, or Mann-Whitney-Wilcoxon tests after Bonferroni-Holm adjustment.

[There were no endpoints with monotonicity disagreement with potential statistical outliers removed.]

Table 19a. FSTRA 84660 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
Female Survival	Monotonic	NP	NP	Group 5 (<.0001)[6]	NP
Male Survival	Monotonic	NP	NP	Group 5 (0.0004)[6]	NP
Fertilization Success	Monotonic	NP	NP	NS	NP
Female Liver VTG [Total RNA (copies vtg/ng)]	Monotonic	NP	NP	NS	NP
Female GSI (%)	Monotonic	NP	NP	NS	NP
Male GSI (%)	Monotonic	NP	NP	NS	NP
Male Body Weight (mg)	Monotonic	NP	NP	NS	NP
Female Body Length (mm)	Monotonic	NP	NP	NS	NP
Female Gonad Weight (mg)	Monotonic	NP	NP	NS	NP
Male Gonad Weight (mg)	Monotonic	NP	NP	NS	NP

1. Monotonicity was assessed by fitting a mixed effects ANOVA model on rank transformed data and determining linear and quadratic contrasts. The data were determined to be non-monotonic if the linear contrast was not significant and the quadratic contrast was significant. Otherwise the data were determined to be monotonic.
2. Shapiro-Wilk test for normality.
3. Levene's Test for homogeneity of variance.
4. Jonckheere-Terpstra step-down test was performed on monotonic concentration response data. Only statistically significant treatment trends were listed.
5. Control is group 1 and test concentration groups are groups 2 to 5. Only statistically significant pairwise comparisons were listed.
    - For non-monotonic, normally distributed, and homogeneous variance data, Dunnett's test was used in the pairwise comparisons to control.
    - For other types of data (e.g., non-monotonic, normally distributed, and heterogeneous variance data), Mann-Whitney-Wilcoxon test was used in the pairwise comparisons to control with Bonferroni-Holm multiple comparison adjustment.
6. Step-down Cochran-Armitage test was performed on survival.

NP Statistical test or comparison was not performed.

NS No statistically significant differences were found for Jonckheere-Terpstra step-down test, Dunnett's tests, or Mann-Whitney-Wilcoxon tests after Bonferroni-Holm adjustment.

[There were no endpoints with monotonicity disagreement with potential statistical outliers removed.]

[Note Male Anal Fin Papillary Processes Count was not included in this statistical analysis task.]

Table 20a. FSTRA 84662 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
Female Liver VTG [Total RNA (copies vtg/ng)]	Monotonic	NP	NP	NS	NP
Male Liver VTG [Total RNA (copies vtg/ng)]	Monotonic	NP	NP	NS	NP
Female GSI (%)	Monotonic	NP	NP	NS	NP
Female Body Weight (mg)	Monotonic	NP	NP	NS	NP
Male Body Weight (mg)	Monotonic	NP	NP	NS	NP
Female Body Length (mm)	Monotonic	NP	NP	NS	NP
Female Gonad Weight (mg)	Monotonic	NP	NP	NS	NP

Table 20b. FSTRA 84662 Statistical Reanalysis Results for Endpoints with Monotonicity Disagreement, with Potential Statistical Outliers Removed.

Endpoint	Monotonicity Assessment[1]	Normality Test[2]	Homogeneity of Variance Test[3]	Jonckheere-Terpstra Test[4] (p-value)	Significant Pairwise Comparisons to Control[5] (p-value)
Male Liver VTG [Total RNA (copies vtg/ng)]	Monotonic	NP	NP	NS	NP
Male Body Weight (mg)	Monotonic	NP	NP	NS	NP

1. Monotonicity was assessed by fitting a mixed effects ANOVA model on rank transformed data and determining linear and quadratic contrasts. The data were determined to be non-monotonic if the linear contrast was not significant and the quadratic contrast was significant. Otherwise the data were determined to be monotonic.
2. Shapiro-Wilk test for normality.
3. Levene's Test for homogeneity of variance.
4. Jonckheere-Terpstra step-down test was performed on monotonic concentration response data. Only statistically significant treatment trends were listed.
5. Control is group 1 and test concentration groups are groups 2 to 5. Only statistically significant pairwise comparisons were listed.
    - For non-monotonic, normally distributed, and homogeneous variance data, Dunnett's test was used in the pairwise comparisons to control.
    - For other types of data (e.g., non-monotonic, normally distributed, and heterogeneous variance data), Mann-Whitney-Wilcoxon test was used in the pairwise comparisons to control with Bonferroni-Holm multiple comparison adjustment.

NP Statistical test or comparison was not performed.

NS No statistically significant differences were found for Jonckheere-Terpstra step-down test, Dunnett's tests, or Mann-Whitney-Wilcoxon tests after Bonferroni-Holm adjustment.

[Note Male Anal Fin Papillary Processes Count was not included in this statistical analysis task.]