STATISTICAL ANALYSIS OF SELENIUM TOXICITY DATA

-PEER REVIEW DRAFT-

August 31, 1999

Prepared for

U.S. Environmental Protection Agency
Health and Ecological Criteria Division
401 M St., SW
Washington, DC 20460

Contract No. 68-C7-0002
Work Assignment No. 4-31

Prepared by

Jie Tao, Penelope Kellar, and William Warren-Hicks
The Cadmus Group, Inc.
1920 Highway 54
Executive Park, Suite 100
Durham, NC 27713

DISCLAIMER

This report was prepared by the Cadmus Group Inc. under Contract No. 68-C7-0002 to the U.S. Environmental Protection Agency. The information in this document does not necessarily reflect the policy of the U.S. Environmental Protection Agency and no official endorsement should be inferred. Mention of Trade names or commercial products does not constitute endorsement or recommendation for use.

Table of Contents

List of Tables

List of Tables (continued)—

List of Figures

SECTION 1
BACKGROUND

The U.S. Environmental Protection Agency (EPA) is currently in the process of revising its aquatic life criteria for selenium. The selenium criteria were last published in 1987, and since then, additional data have become available on the effects of selenium on aquatic organisms. Included among the new data is a series of three studies on the effect of selenium on bluegills (*Lepomis macrochirus*) in outdoor experimental streams at the Monticello Ecological Research Station (MERS) in Monticello, Minnesota. Results from the first of the three studies were published in 1992[1], and they are not discussed here.

In this report, data from the second and third MERS selenium studies (Study II and Study III) are evaluated by The Cadmus Group, Inc., using a variety of statistical methods. EPA intends to consider the results from the statistical analyses presented in this report in its forthcoming manuscript on the second and third MERS studies. Section 2 of this report provides a brief summary of the experimental design and procedures used in Studies II and III. Additional detail on the experimental methodology will be provided by EPA in its forthcoming manuscript. Section 3 presents the statistical analysis of the effects of selenium on bluegill spawning and progeny. Section 4 presents the statistical analysis of the effects of selenium on bluegill spawning and progeny. Finally, details on the calculations, statistical output, graphical summaries, raw data, and the statistical programs used are provided in Appendices A through F.

[1] See Hermanutz, R.O., K.N. Allen, T.H. Rousch, and S. Hedtke. 1992. Effects of elevated selenium concentrations on bluegills (*Lepomis macrochirus*) in outdoor experimental streams. Environ. Toxicol. Chem. 11:217-224.

SECTION 2
EXPERIMENTAL DESIGN AND PROCEDURES

This section presents a brief summary of the experimental design and procedures used in Studies II and III for evaluating the effect of selenium on bluegills in outdoor streams at the Monticello Experimental Research Station. EPA intends to provide additional detail on the study design and experimental procedures in a forthcoming manuscript on Studies II and III. As discussed in Section 1, the results of a prior study (Study I) are not analyzed in this report.

Using data obtained from EPA, The Cadmus Group, Inc., (Cadmus) calculated many descriptive statistics. Some data were corrected for transcription errors before the analysis. EPA provided the description of the experimental methodology and guidance concerning some aspects of the statistical analyses, such as the use of PROC MIXED as opposed to PROC GLM. During the course of data evaluation, EPA and Cadmus considered such issues as the following: (1) methods for handling missing observations and data inconsistencies, (2) methods for handling repeated measures over time, (3) data transformations, and (4) issues regarding the calculations. The datasets used in our analyses are listed in Table 2-1. Table 2-2 provides a description of the basic experimental layout of the field studies conducted at the Monticello experimental stream sites. Table 2-3 presents descriptive statistics on the concentrations of selenium measured during each study.

The Monticello Ecological Research Station contains eight outdoor streams that were supplied with Mississippi River water or a mixture of river water and well water. Six of the eight streams were used in Study I, with nominal selenium concentrations of 0 µg/L, 10 µg/L, and 30 µg/L. All eight streams were used in Studies II and III. For Study II, dosing of the 10 µg/L selenium streams was continued. Dosing of the 30 µg/L streams was discontinued, and these streams were used to determine whether residual toxicity was caused by previous selenium dosing. The other four streams (2 unused, 2 controls in Study I) were randomly assigned to the nominal concentrations of 0 µg/L and 2.5 µg/L. In Study III, selenium was not added to any stream. Study III addresses only residual effects in recovering streams.

For both studies, adult bluegills were obtained from a south-central Minnesota farm pond. Eighty-five fish in Study II and 98 fish in Study III were randomly distributed without regard to sex in the upper reaches (i.e., sampling Stations 0-2) in each stream. In each stream, a random sample of the adults was transferred to the lower reach of the respective experimental stream (Station 6) for the reproduction portion of the studies; 26 per stream were transferred in Study II and 22 to 50 per stream in Study III. Each fish was weighed at the time of transfer, and two to four fish were randomly selected from each stream for measurement of selenium in selected tissues. After transfer, each Station 6 pool was checked daily for the presence of bluegill nests. All observed nests were sampled three times a week for the presence of embryos and larvae. If, after five passes of the sampling device, no embryos or larvae were present, the nest was considered inactive. In Study III, the sampling device was passed over the nest until an adequate sample was collected or the investigator determined the nest to be inactive. The numbers of live and dead embryos and larvae were recorded (hereinafter referred to as "Field Nest Data").

Table 2-1. Description of Datasets in the MERS Selenium Studies[a]

Dataset	File Name	Date Received	Description	Comments Concerning the Datasets
Study II				
Survival Data	s2surviv.sd2	October 1997	number of fish at the beginning and end of the two-phase study (survival and growth, reproduction) for each stream	
Growth Data	bgmay89.sd2 bgag8889.sd2	October 1997	weight and gender for each fish on the transfer date; weight, length, K factor, and gender for each fish at the end of study	
Field Nest Data	fldabnrm.sd2	October 1997, April 1998	number of live and dead embryos and larvae in each nest, age of the larvae, number of larvae in each subsample, and number of larvae with abnormalities in the subsample	Data received in October 1997 were replaced by the corrected version received in April 1998; duplicate spawning activities in a given nest were treated as independent events.
Egg Cup Data	cupdays.sd2	October 1997	number of live and dead embryos and larvae each day in each cup taken from randomly selected nests, and number of larvae with abnormalities each day in each cup	Discovered some measurement errors; for example, total number of larvae are not consistent throughout five-day experiment; as a result, the calculated percent hatch exceeds 100% for two cups; in this case, we truncated the values to 100% only for the arc-sine square-root transformation.
Study III				
Survival Data	s3surviv.sd2	October 1997	number of fish at the beginning and end of the two-phase study (survival and growth, reproduction) for each stream	
Growth Data	bgmay90.sd2 bgf8990.sd2	October 1997	weight and gender for each fish on the transfer date; weight, length, and gender for each fish at the end of study	
Field Nest Data	s3nest.sd2	October 1997	spawn number, number of larvae in each subsample, age of the larvae (mostly missing), and number of larvae with abnormalities in the subsample	
Egg Cup Data	s3eggcup.sd2	October 1997	initial number in each cup, number of live and dead embryos and larvae each day in each cup taken from randomly selected nests, and number of larvae with abnormalities each day in each cup	Discovered some measurement errors; for example, the initial number in each cup is sometimes not consistent with the calculated total number for that cup at Day 1; in this case, the calculated number was used in the analysis; total number of larvae is not consistent throughout the five-day experiment; as a result, the calculated percent hatch exceeds 100% for one cup, and in this case, we truncated the value to 100% only for the arc-sine square-root transformation.

[a] The calculations in Tables 3-1, 3-2, 4-1, and 4-2 are summarized in Appendix A. The complete dataset printouts are presented in Appendix E.

Table 2-2. Selenium Treatments for the Three Studies

Addition of Selenium			
Treatment	**Streams**	**Total Period of Dosing**	**Study**
30 µg/L	4 and 6	03/11/87 - 10/02/88 (571 days)	I
10 µg/L	3 and 8	03/11/87 - 11/07/89 (972 days)	I, II
2.5 µg/L	2 and 7	10/3/88 - 11/07/89 (400 days)	II
Studies of Bluegill			
Stream	**Study I[a]**	**Study II**	**Study III**
Beginning[b] Transfer[c] End[d]	09/01/87 05/16/88 08/22/88	10/88 05/89 08/89	11/89 05/90 07/90
1	Unused	Control	Control
2	Unused	2.5 µg/L	Recovering
3	10 µg/L	10 µg/L	Recovering
4	30 µg/L	Recovering	Recovering
5	Control	Control	Control
6	30 µg/L	Recovering	Recovering
7	Control	2.5 µg/L	Recovering
8	10 µg/L	10 µg/L	Recovering

[a] Previously reported (see Footnote 1, page 1), data are not evaluated in this report.
[b] Adult bluegills placed in Stations 0-2 for survival and growth study.
[c] Transfer of adult bluegills within each stream from Stations 0-2 to Station 6 for reproduction study.
[d] Adult bluegills removed from Station 6.

Table 2-3. Measured Concentrations of Selenium in Water During Study II

Intended Concentration	**2.5 µg/L**		**10 µg/L**	
Stream Number	2	7	3	8
Stations 1 and 3 combined				
Mean	2.48	2.67	8.87	9.65
Standard Deviation	0.28	0.45	1.46	1.81
Number of Measurements	29	30	27	30
Stations 5 and 7 combined				
Mean	2.67	2.55	10.40	10.83
Standard Deviation	0.43	0.39	1.15	1.07
Number of Measurements	28	27	27	27
Mean exposure concentration[a]	2.53	2.63	9.34	10.02

[a] Mean exposure concentration of adult bluegills was calculated as the time-weighted mean of the means of Stations 1 and 3 (221 days) and Stations 5 and 7 (99 days).

Samples of embryos were incubated in the laboratory to determine hatchability, larval survival, and incidence of larval anomalies. Randomly selected embryos were reared for several days in incubation cups (hereinafter referred to as "Egg Cup data"). Cup contents were removed and examined daily. Percent hatch and larval survival were recorded. Live larvae were examined for abnormalities.

The statistical analyses conducted on the Monticello field and laboratory data are described in the following sections.

SECTION 3
EFFECTS OF SELENIUM ON SURVIVAL AND GROWTH OF ADULT BLUEGILLS

This section presents the summary and analysis approaches for the survival and growth data for Studies II and III. Section 3.1 summarizes the survival and growth data. The analysis of variance (ANOVA) models applied to the survival and growth data are presented in Section 3.2

3.1 SUMMARY OF SURVIVAL AND GROWTH DATA

Table 3-1 summarizes the adult bluegill percent survival and growth data during Study II at the time of transfer to Station 6 (for the reproductive portion of the experiment) (Day 221) and at the end of the study (Day 320). Table 3-2 presents a similar summary for Study III at the time of transfer (Day 181) and at the end of the study (Day 265). An explanation of the calculation of the variables in the tables is presented in Tables A-1 and A-2 in Appendix A. To evaluate differences in growth between genders, Cadmus also summarized length, weight, and K factor (weight × 10/length^3) were summarized by gender, as presented in Tables 3-3 and 3-4.

Throughout this report, a standardized process was used to examine the results of an ANOVA model. First, an ANOVA model that is consistent with the experimental design was selected. The model was fit to the data. An F-statistic evaluating the relative fit of the model to the data was examined. If the p-value of the model F-statistic was less than 0.05 ($p<0.05$), the model was considered to be significant. For those models that were significant, individual parameters in the model were examined. Those parameters where the p-value of the parameter F-statistic was less than 0.05 were considered significant and subjected to a means separation test. Several means tests are available in the literature and most software packages. Each method has a slightly different interpretation, and the methods are not guaranteed to produce the same results. Dunnett's test against control and the Tukey's standardized range test were used for this analysis.[2] Those cases in which the two tests provide different results are noted in the narrative of the report.

3.2 ANOVA ON PERCENT SURVIVAL AND GROWTH OF ADULT BLUEGILLS

3.2.1 Survival

Percent survival at the transfer date and at the end of the study was analyzed using the following one-way ANOVA model to evaluate the effects of different selenium concentrations (0, 2.5, and 10 µg/L) on adult bluegills. The 30 µg/L treatment was excluded from this ANOVA for Study II, because this treatment represents the recovery from the previous study and, therefore, is not comparable to the continuous exposure regimes of the 2.5 and 10 µg/L treatments. Similarly, the 30 µg/L treatment was not comparable in Study III.

[2] Montgomery, D.C. 1991. "Design and Analysis of Experiments." Third Edition. John Wiley & Sons. pp. 73-80.

Table 3-1. Adult Percent Survival and Growth During Study II[a]

	Selenium Treatment							
	Control		2.5 µg/L		10 µg/L		Recovering 30 µg/L	
Stream	1	5	2	7	3	8	4	6
Day 221[b]								
% survival	47.1	49.4	32.9	42.4	34.1	43.5	34.1	51.8
Weight (g)[c]	103.0 (25.4)	98.9 (30.2)	101.8 (34.6)	101.3 (28.4)	100.0 (32.8)	92.2 (30.4)	98.1 (36.5)	101.5 (26.0)
Weight gain (g)[d]	24.8	20.7	23.6	23.1	21.8	14.0	19.9	23.3
Day 320[e]								
Cumulative % survival[f]	23.5	28.5	8.9	39.1	9.2	16.7	0	19.9
% survival from Day 221 to Day 320[b,e]	50.0	57.7	26.9	92.3	26.9	38.5	0	38.5
Weight (g)[c]	113.3 (24.3)	141.3 (26.1)	145.5 (24.9)	146.1 (34.2)	164.1 (37.0)	130.8 (19.5)	-[i]	156.7 (30.8)
Weight gain (g)[g]	35.1	63.1	67.3	67.9	85.9	52.6	-[i]	78.5
Length (mm)[c]	178.6 (11.4)	191.9 (12.4)	190.4 (8.5)	190.3 (10.4)	191.3 (15.4)	182.5 (8.9)	-[i]	189.4 (12.9)
K factor[c,h]	1.964 (0.178)	1.984 (0.121)	2.089 (0.144)	2.080 (0.179)	2.314 (0.123)	2.142 (0.143)	-[i]	2.284 (0.163)

[a] The initial averages (and standard deviations) were: weight = 78.2 (26.4) g, length = 164.5 (15.6) mm, and K factor = 1.711 (0.355).

[b] A subset of fish was transferred to Station 6 on Day 221 for the reproductive portion of the study.

[c] Numbers in parentheses are standard deviations.

[d] Weight gain from the beginning of the study, calculated by subtracting the average initial weight from the weight on Day 221.

[e] The study ended on Day 320.

[f] The product of percent survival on Day 221 and percent survival from Day 221 to Day 320.

[g] Weight gain from the beginning of the study, calculated by subtracting the average initial weight from the weight on Day 320.

[h] K factor = weight x 10 / length^3. Numbers in parentheses are standard deviations.

[i] All fish died by this time in Stream 4.

Table 3-2. Adult Percent Survival and Growth During Study III[a]

	Selenium Treatment							
	Control		Recovering 2.5 µg/L		Recovering 10 µg/L		Recovering 30 µg/L	
Stream	1	5	2	7	3	8	4	6
Day 181[b]								
% survival	33.7	34.7	53.1	27.6	32.7	24.5	33.7	28.6
Weight (g)[c]	113.6 (22.7)	115.1 (20.9)	102.6 (22.5)	111.4 (30.7)	114.9 (21.3)	102.1 (17.8)	107.3 (20.8)	112.9 (23.6)
Weight gain (g)[d]	26.6	27.5	15.0	23.8	27.3	14.5	19.7	25.3
% survival from Day 181 to Day 265[b,e]	29.0	31.3	34.0	48.0	30.0	36.4	61.3	-[i]
Day 265[e]								
% survival[f]	9.8	10.8	18.0	13.2	9.8	8.9	20.6	-[i]
Weight (g)[c]	119.6 (36.9)	156.2 (24.1)	123.7 (26.7)	142.9 (34.8)	164.6 (40.3)	162.6 (16.8)	130.3 (25.2)	-[i]
Weight gain (g)[g]	32.0	68.6	36.1	55.3	77.0	75.0	42.7	-[i]
Length (mm)[c]	178.0 (12.8)	191.2 (8.8)	179.1 (11.6)	184.8 (15.7)	194.7 (14.2)	173.5 (32.7)	183.7 (10.2)	-[i]
K factor[c, h]	2.059 (0.202)	2.216 (0.112)	2.120 (0.095)	2.232 (0.166)	2.207 (0.290)	4.161 (3.878)	2.077 (0.119)	-[i]

[a] The initial averages (and standard deviations) were: weight = 87.6 (21.3) g, length = 167.5 (11.0) mm, and K factor = 1.841 (0.273).

[b] A subset of fish was transferred to Station 6 on Day 181 for the reproductive portion of the study.

[c] Numbers in parentheses are standard deviations.

[d] Weight gain from the beginning of the study, calculated by subtracting the average initial weight from the weight on Day 181.

[e] The study ended on Day 265.

[f] The product of percent survival on Day 181 and percent survival from Day 181 to Day 265.

[g] Weight gain from the beginning of the study, calculated by subtracting the average initial weight from the weight on Day 265.

[h] K factor = weight x 10 / length^3. Numbers in parentheses are standard deviations.

[i] Stream 6 was removed from this study after Day 181.

Table 3-3. Adult Growth by Gender During Study II[a]

	Selenium Treatment							
	Control		2.5 µg/L		10 µg/L		Recovering 30 µg/L	
Stream	**1**	**5**	**2**	**7**	**3**	**8**	**4**	**6**
Day 221[b]								
Weight (g)								
Female	96.9 (24.1)	98.9 (27.1)	99.4 (33.1)	106.8 (22.3)	106.0 (31.7)	93.4 (32.3)	89.4 (21.7)	100.1 (22.6)
Indeterminate	81.4 (27.1)	69.7 (24.5)	71.3 (13.4)	65.8 (19.6)	60.0 (15.0)	59.3 (5.1)	67.0 (7.7)	80.1 (19.4)
Male	117.0 (19.4)	110.7 (29.7)	124.8 (35.4)	107.5 (31.4)	107.4 (29.2)	96.7 (28.8)	139.9 (29.1)	113.5 (27.7)
Day 320[c]								
Weight (g)								
Female	109.1 (23.6)	125.3 (14.6)	128.8 (18.7)	126.1 (30.1)	173.3 (29.7)	132.0 (22.6)	-[d]	152.4 (38.5)
Male	136.1 (16.7)	149.3 (27.5)	167.7 (6.0)	158.1 (31.6)	160.5 (42.1)	130.5 (20.3)		161.0 (24.7)
Length (mm)								
Female	177.0 (11.7)	186.0 (8.5)	185.3 (6.6)	184.4 (9.9)	199.5 (14.8)	182.0 (18.4)	-[d]	188.2 (18.4)
Male	187.5 (2.1)	194.8 (13.4)	197.3 (5.0)	193.9 (9.2)	188.0 (15.9)	182.6 (7.4)		190.6 (5.4)
K factor								
Female	1.946 (0.180)	1.945 (0.122)	2.017 (0.141)	1.977 (0.161)	2.177 (0.112)	2.200 (0.288)	-[d]	2.255 (0.133)
Male	2.061 (0.183)	2.003 (0.123)	2.185 (0.092)	2.142 (0.164)	2.369 (0.080)	2.127 (0.112)		2.312 (0.199)

[a] The initial averages (and standard deviations) were: weight = 78.2 (26.4) g, length = 164.5 (15.6) mm, and K factor = 1711 (0.355). Numbers in parentheses are standard deviations.

[b] The fish were transferred on Day 221.

[c] The study ended on Day 320.

[d] All fish died by this time in Stream 4.

Table 3-4. Adult Growth by Gender During Study III[a]

Stream	Selenium Treatment							
	Control		Recovering 2.5 µg/L		Recovering 10 µg/L		Recovering 30 µg/L	
	1	5	2	7	3	8	4	6
Day 181[b]								
Weight (g)								
Female	108.8 (19.1)	102.1 (18.0)	99.8 (17.6)	107.2 (22.8)	107.5 (18.2)	97.1 (10.7)	100.2 (15.8)	105.8 (19.7)
Indeterminate			57.7 (5.5)	76.4 (18.6)				
Male	126.3 (27.6)	129.6 (13.1)	114.2 (21.0)	129.6 (27.2)	131.1 (19.0)	110.6 (24.1)	115.0 (23.1)	130.8 (24.1)
Day 265[c]								
Weight (g)								
Female	104.9 (25.3)	136.3 (32.1)	94.9 (8.3)	113.4 (12.5)	150.9 (41.9)	156.2 (14.9)	111.8 (10.3)	- [d]
Male	171.2 (14.4)	164.7 (15.6)	139.3 (18.4)	172.5 (20.2)	175.6 (39.9)	173.2 (16.5)	150.9 (20.1)	
Length (mm)								
Female	173.0 (9.0)	184.0 (12.1)	166.3 (4.6)	173.2 (8.7)	192.8 (18.2)	176.6 (22.6)	176.9 (6.3)	- [d]
Male	195.5 (6.4)	194.3 (5.4)	186.0 (7.4)	196.3 (12.1)	196.2 (12.3)	168.3 (51.4)	191.3 (8.1)	
K factor	1.993	2.159	2.059	2.181	2.102	3.111	2.017	
Female	(0.177)	(0.076)	(0.095)	(0.108)	(0.419)	(1.546)	(0.103)	- [d]
Male	2.288	2.241	2.154	2.285	2.290	5.911	2.144	
	(0.031)	(0.120)	(0.080)	(0.205)	(0.131)	(6.364)	(0.102)	

[a] All Study III streams were recovering from prior selenium exposures in Study II (2.5 and 10 µg/L) and Study I (30 µg/L). The initial averages (and standard deviations) were: weight = 87.6 (21.3) g, length = 167.5 (11.0) mm, and K factor = 1.841 (0.273). Numbers in parentheses are standard deviations.

[b] A subset of fish was transferred to Station 6 on Day 181 for the reproductive portion of this study.

[c] The study ended on Day 265.

[d] Stream 6 was removed from this study after Day 181.

The models used for statistical analyses were:

$$\text{Study II:} \quad P\text{-}221_{ij},\ P\text{-}320_{ij} = \mu + T_i + \varepsilon_{ij}$$
$$\text{Study III:} \quad P\text{-}181_{ij},\ P\text{-}265_{ij} = \mu + T_i + \varepsilon_{ij}$$

where,

$P\text{-}221_{ij}$ = percent survival from Day 1 to Day 221, for the i^{th} treatment and j^{th} stream;

$P\text{-}320_{ij}$ = percent survival from Day 1 to Day 320; calculated as the product of the percent survival from Day 1 to Day 221 and percent survival from Day 221 to Day 320, for the i^{th} treatment and j^{th} stream;

$P\text{-}181_{ij}$ = percent survival from Day 1 to Day 181, for the i^{th} treatment and j^{th} stream;

$P\text{-}265_{ij}$ = percent survival from Day 1 to Day 265; calculated as the product of the percent survival from Day 1 to Day 181 and percent survival from Day 181 to Day 265;

μ = overall mean;

T_i = treatment effect, i = 1 to 3; and

ε_{ij} = random error.

For the purpose of hypothesis testing, the error term is assumed to follow a normal distribution with independent realizations of the data. Because the dataset contains only six observations spanning three treatments, the assumption of normal independent errors may not hold. Figures 3-1 and 3-2 present plots of the model residuals against treatment. Examination of the residual plots shows that the residual variance is not consistent for each treatment. Particularly, the residual variance is largest for the 2.5 µg/L selenium treatment. We transformed the response variable (percent survival) using the arc-sine square-root transformation and reran the ANOVA model. The results of these runs are presented in Table 3-5.

Table 3-5. ANOVA Results on Percent Survival: Model Significance (p-value)

Approach	Study II		Study III	
	Day 1-221	Day 1-320	Day 1-181	Day 1-265
Arc-sine Square-root of the Response Variables	0.263	0.604	0.618	0.086

As shown in Table 3-5, p-values of the ANOVA model are greater than 0.05, which indicates no significant differences in percent survival at $\alpha = 0.05$ with varying concentrations of selenium for either Study II or Study III. The p-value is defined as the probability of observing a sample outcome more contradictory to H_o (no difference in response variables with different selenium concentrations) than the observed sample result. The smaller the p-value, the heavier the weight of the sample evidence against H_o.

Figure 3-1. Residuals from the ANOVA Model on Percent Survival—Study II

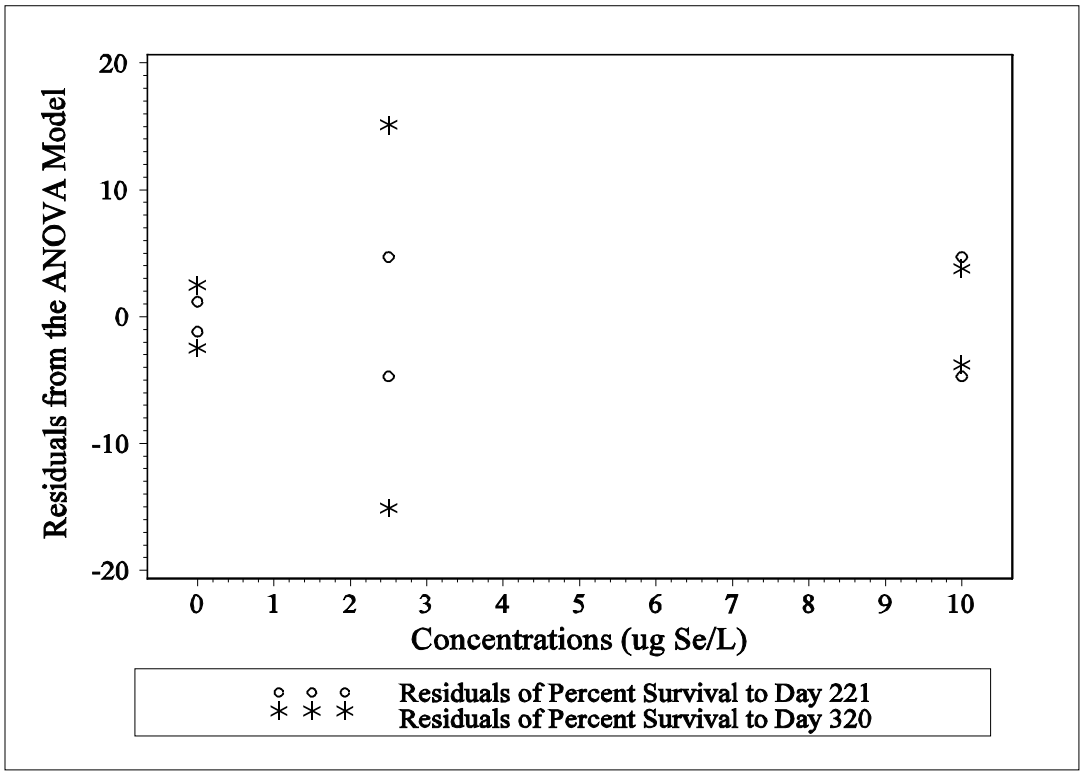
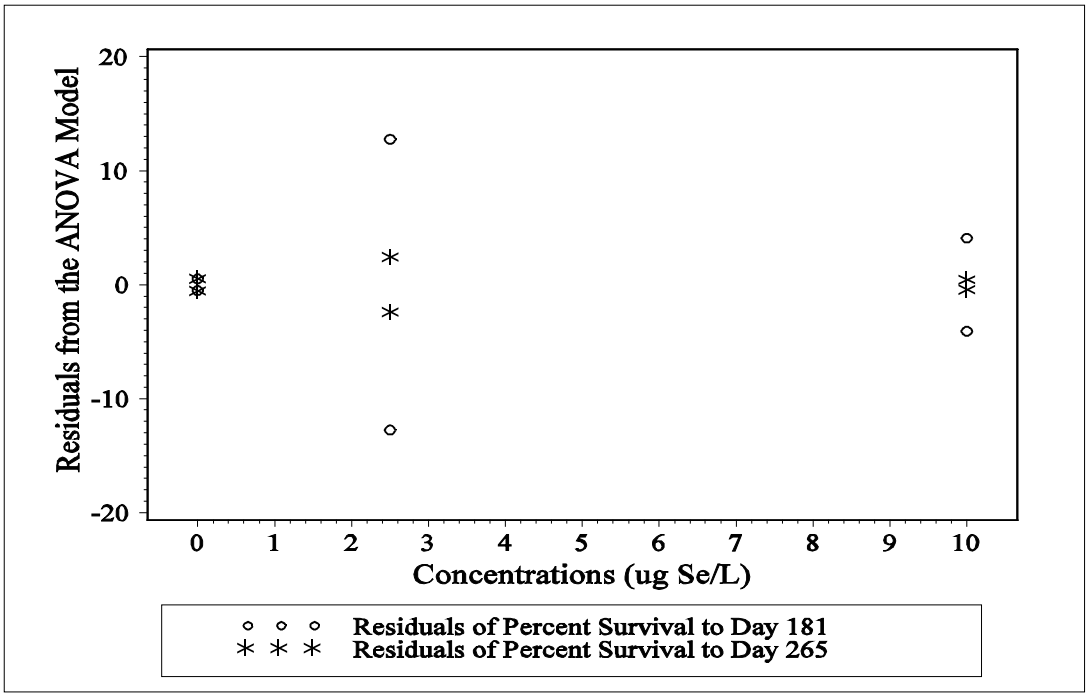


Figure 3-2. Residuals from the ANOVA Model on Percent Survival—Study III



Although the residual plots show some violations of the error assumptions, the ANOVA results are little affected by these violations. The largest impact on the ANOVA results is probably due to the small overall sample size and the small number of within-treatment replicates.

We calculated the power of the ANOVA model using standard statistical methods (Steel and Torrie, 1960[3]). The power of the test is illustrated in Figures 3-3 through 3-6 for selected Type I error rates. A key issue in the use of any statistical method, including ANOVA, is the number of experimental units required for a specified decision criterion. In the ANOVA models where percent survival is the response variable, the experimental unit or replicate is the stream. The null hypothesis inherent in the model is that mean percent survival is the same among all selenium treatments. The ANOVA model is fit to several datasets with six streams and three selenium treatments. Figures 3-3 through 3-6 indicate the number of streams required to meet specified levels of power and Type I error for tests of hypotheses on fish survival (note: the calculations are model specific, and extrapolating the results to different endpoints or ANOVA models may not be appropriate). Power is the probability of detecting a real difference of 10% among treatment means, and the Type I error (0.05, 0.10, 0.20) represents the probability of erroneously detecting a treatment difference when none exists. The streams are assumed to be equally allocated among three treatments. It is desirable to have a high chance of classifying the treatment means as equal, when they are (i.e., high power, as indicated in the upper regions of the abscissa). Examination of the plots shows that six streams provide reasonably good power for all study-specific datasets (power ranges between 0.6 and 1.0)

3.2.2 Growth

While weight gain for individual fish is not available in the datasets, we calculated weight gain during the period between transfer and the end of the study for each gender in each stream. These gender-stream-specific data were evaluated using the following model:

$$W_{ijk} = \mu + T_i + G_j + (TG)_{ij} + S_{k(i)} + \varepsilon_{ijk}$$

where,

W_{ijk} = weight gain between the transfer and the end of the study, i = 1 to 3, j=1 to 2, k = 1 to 2;

μ = overall mean;

T_i = treatment effect, i = 1 to 3;

G_j = gender effect, j = 1 to 2;

$(TG)_{ij}$ = interaction between the treatment and gender, i = 1 to 3, j = 1 to 2;

$S_{k(i)}$ = stream effect, k = 1 to 2, nested within treatment, considered as a random effect; and

ε_{ijk} = random error.

The results indicate that no effect is significant at $\alpha = 0.05$ for weight gain in either Study II or Study III.

[3] Steel, R.G.D., and J.H. Torrie. 1960. Principles and Procedures of Statistics, McGraw-Hill, New York, NY.

Figure 3-3. Power of the ANOVA Model on Percent Survival—Study II Days 1-221

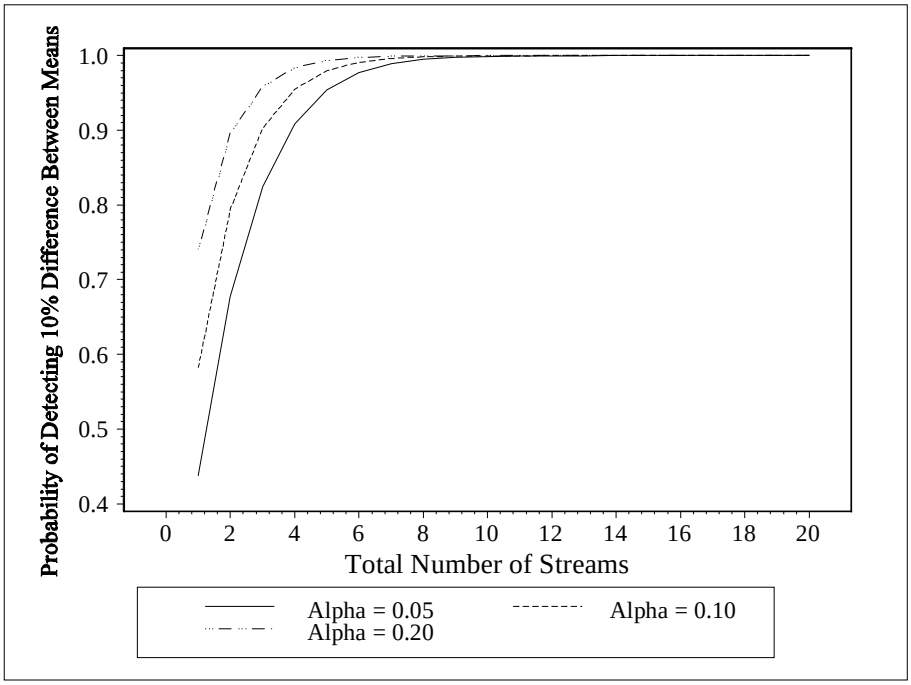


Figure 3-4. Power of the ANOVA Model on Percent Survival—Study II Days 1-320

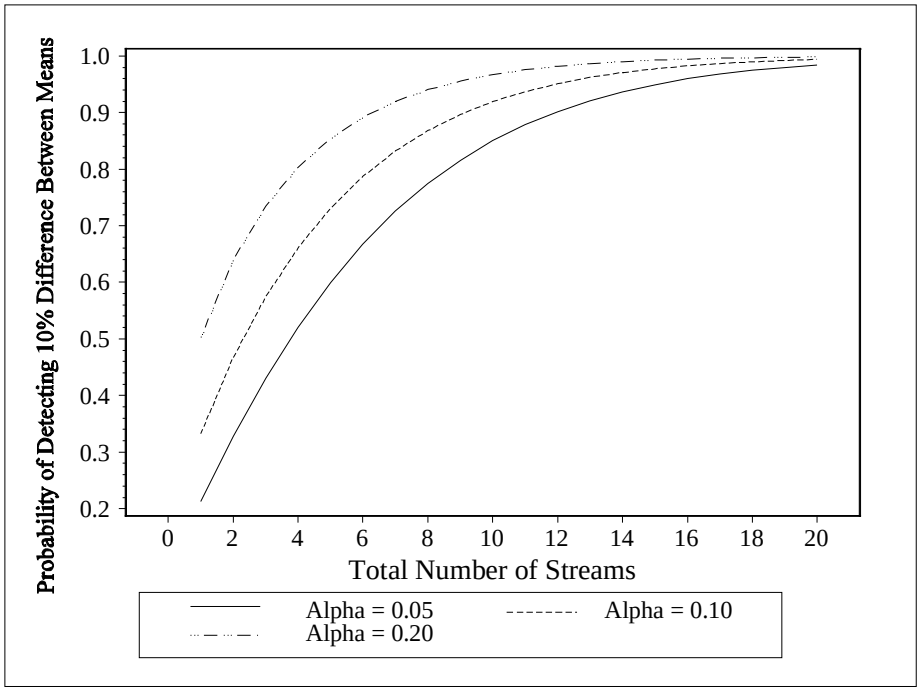


Figure 3-5. Power of the ANOVA Model on Percent Survival— Study III Days 1-181

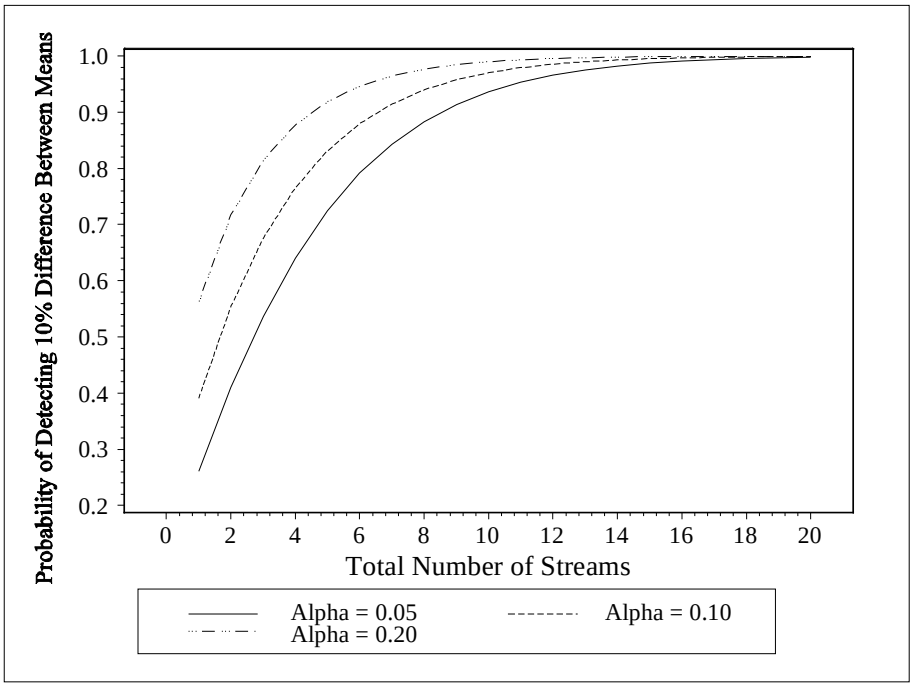
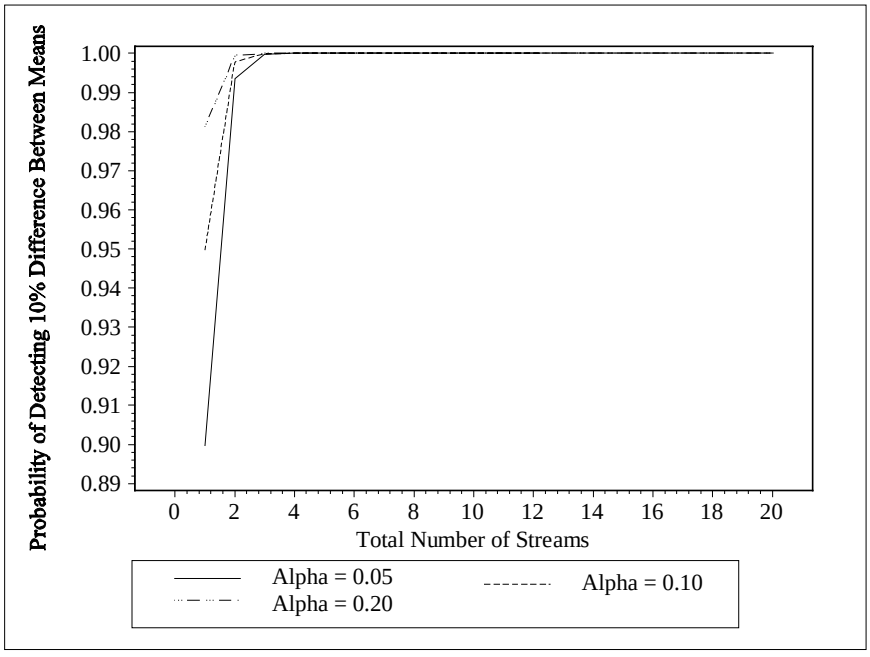


Figure 3-6. Power of the ANOVA Model on Percent Survival— Study III Days 1-265



SECTION 4
EFFECTS OF SELENIUM ON SPAWNING ACTIVITY AND PROGENY OF ADULT BLUEGILLS

After fish were transferred to Station 6 in each stream, each pool was checked daily for the presence of bluegill nests. All observed nests were marked with labeled stakes and sampled every Monday, Wednesday, and Friday for the presence of embryos and larvae. The numbers of live and dead embryos and larvae were recorded. Dead embryos were identified by their opaqueness. Samples of live larvae were observed under a dissecting microscope to identify morphological anomalies. This set of data is referred to as "Field Nest" data in the following text.

Samples of embryos were incubated in the laboratory to determine embryo hatchability, larval survival, and incidence of larval anomalies. Randomly selected embryos were reared for several days in incubation cups in both Studies II and III. The embryos were exposed to a proportional diluter that contained the same river water and the same nominal selenium concentrations as those in the respective test streams. Cup contents were removed and examined daily; live embryos and larvae were returned after the cup was cleaned. Percent hatch and larval survival were recorded. Live larvae were examined for abnormalities. This set of data is referred to as "Egg Cup" data in the following text.

The Field Nest data and Egg Cup data for both studies are summarized in Section 4.1. Section 4.2 describes the experimental design, the ANOVA model, and results for these datasets.

4.1 SUMMARY OF FIELD NEST AND EGG CUP DATA

Tables 4-1 and 4-2 summarize the adult bluegill spawning activity and effects on progeny under the influence of differing selenium concentrations for Study II and Study III, respectively. The number of active nests, number of embryos collected, number of larvae collected, and number of samples containing larvae were summarized for each stream using the Field Nest Data. In addition, the percentage of dead embryos, percentage of dead larvae, and percentage of abnormalities, among live larvae were also calculated and summarized for each stream. The egg cup data were used to calculate percent hatch, percent survival to the third day (to eliminate the starvation effect later in the experiment), percent abnormalities and percent healthy among live larvae for each stream. Numbers in parentheses are the standard deviations of the parameter. Tables A-3 and A-4 in Appendix A explain the calculations for each variable in the tables.

The average percent abnormalities among live larvae on each day at each selenium concentration for Field Nest Data and egg cup data for Study II and Study III are summarized and presented in Tables 4-3 through 4-6. In addition, a series of plots was generated to depict the percent abnormalities among live larvae on different days at different dosing levels for the Field Nest Data and egg cup data for Studies II and III. These figures are presented in Appendix C.

4.2 ANOVA ON FIELD NEST AND EGG CUP DATA

To examine the effects of selenium on adult bluegill spawning activity and progeny, candidate ANOVA models were evaluated and those models most consistent with the experimental design were selected. Because the datasets are highly unbalanced and random effects (e.g., stream, nest) are present in the models, PROC MIXED provided by SAS® was used for the analyses presented in this section. Detailed information concerning PROC MIXED can be found in "SAS/STAT Software:

Changes and Enhancement through Release 6.12." Additional information on PROC MIXED can be found in Latour et al. (1994),[4] Wolfinger et al. (1994),[5] and Schwarz (1993).[6]

Table 4-1. Spawning Activity and Effects on Progeny During Study II

	Selenium Treatment							
	Control		2.5 µg/L		10 µg/L		Recovering 30 µg/L	
Stream	**1**	**5**	**2**	**7**	**3**	**8**	**4**	**6**
Field Nest Data								
# of active nests[a]	6	9	1	5	2	3	0	8
# of embryos collected[a]	2,458	1,329	0	1,462	672	931	---	646
% dead embryos[a]	0.94	0	---	0	0	0.32	---	0
# larvae collected[a]	3,252	3,435	2,497	4,717	5,376	750	--	6,788
% dead larvae[a]	0.03	1.05	0.20	0.08	0.50	0.40	---	7.79
# of samples containing larvae (n)	7	13	3	8	9	4	---	16
% edema[b]	0 (0-0)	0 (0-3.3)	4.1 (4.1-4.1)	0 (0-10.7)	81.4 (66.7-96.2)	50.0 (0-91.2)	---	27.3 (0-91.3)
% lordosis[b]	0 (0-0)	0 (0-0)	25.0 (25.0-25.0)	0 (0-3.2)	5.0 (3.3-6.7)	14.7 (0-23.3)	---	0 (0-6.7)
% hemorrhaging[b]	0 (0-0)	0 (0-2.8)	77.6 (77.6-77.6)	52.0 (0-100.0)	55.5 (23.1-87.9)	26.7 (20.6-57.1)	---	17.1 (0-22.7)
Egg Cup Data								
# of trials[c]	6	5	0	4	3	2	---	6
% hatch[d]	93.0 (6.4)	96.4 (3.1)	---	81.4 (11.9)	83.3 (23.1)	91.1 (9.1)	---	92.9 (12.7)
% survival to third day[e]	75.2 (14.8)	71.5 (22.1)	---	71.6 (7.4)	57.7 (32.1)	57.1 (25.7)	---	79.0 (14.2)
% edema[f]	0 (0-3.7)	0 (0-10.0)	---	0 (0-20.0)	100.0 (0-100.0)	100.0 (74.1-100)	---	17.4 (0-94.3)
% lordosis[f]	0 (0-8.7)	0 (0-3.8)	---	0 (0-32.1)	11.1 (0-51.1)	18.2 (0-40.7)	---	0.0 (0-37.1)
% hemorrhaging[f]	0 (0-4.3)	0 (0-10.0)	---	3.6 (0-81.4)	49.3 (0-100.0)	41.1 (0-83.3)	---	11.5 (0-45.7)
% healthy[g]	97.8 (91.3-100)	97.9 (90-100)	---	92.2 (18.6-100)	0 (0-100)	0 (0-25.9)	---	70.7 (5.7-100)

[a] Cumulative for the stream, i.e., one value per stream.
[b] Among live larvae; the median and range (in parentheses) of the maximum incidence per set of nest observations are given.
[c] A trial was set up whenever sufficient larvae were collected.
[d] Cumulative percent hatch for each cup; the mean and standard deviation (in parentheses) are given.
[e] Mean percent survival to third day after first larva hatched; the standard deviation is given in parentheses.
[f] Among larvae that survived up to third day after first larva hatched; the median and range (in parentheses) of the maximum incidence are given.
[g] Among live larvae that survived up to third day after first larvae hatched; assumes the observations of multiple abnormality types always co-occurred in the same organism, this may overestimate the actual % healthy when this assumption is violated. The median and range (in parentheses) of the percent healthy per cup are given.

[4] Latour D., K. Latour, and R.D. Wolfinger. 1994. Getting started with PROC MIXED. SAS Institute, Inc., Cary, NC.

[5] Wolfinger R., R. Tobias, and J. Sall. 1991. Mixed models: A future direction. Proceedings of the Sixteenth Annual SAS Users Group Conference, SAS Institute, Inc., Cary, NC. pp. 1380-1388.

[6] Schwarz, C.J. 1993. The mixed-model ANOVA: The truth, the computer packages, the books. The American Statistician 47(1):48-59.

Table 4-2. Spawning Activity and Effects on Progeny During Study III[a]

	Selenium Treatment							
	Control		Recovering 2.5 µg/L		Recovering 10 µg/L		Recovering 30 µg/L	
Stream	**1**	**5**	**2**	**7**	**3**	**8**	**4**	**6**
Field Nest Data								
# of active nests[b]	4	3	4	2	3	4	7	---
# of samples containing larvae (n)	6	3	5	2	9	6	13	---
% edema[c]	0 (0-0)	0 (0-0)	0 (0-0)	0 (0-0)	0 (0-0)	0 (0-0)	0 (0-4.7)	---
% lordosis[c]	0 (0-0)	0 (0-0)	0 (0-0)	0 (0-0)	0 (0-0)	0 (0-0)	0 (0-2.3)	---
% hemorrhaging[c]	0 (0-0)	0 (0-0)	2.44 (0-10.7)	0 (0-0)	0 (0-0)	0 (0-2.0)	2.38 (0-12.5)	---
Egg Cup Data								
# of trials[d]	2	3	3	7	---	3	5	---
% hatch[e]	85.3 (3.8)	76.9 (20.1)	90.0 (6.0)	88.0 (12.5)	---	78.9 (13.3)	92.5 (13.7)	---
% survival to third day[f]	62.9 (12.5)	68.0 (19.0)	71.3 (27.3)	72.2 (11.2)	---	63.4 (3.7)	81.1 (21.2)	---
% edema[g]	0 (0-0)	0 (0-0)	0 (0-0)	0 (0-0)	---	0 (0-0)	---	---
% lordosis[g]	0 (0-0)	0 (0-0)	0 (0-0)	0 (0-0)	---	0 (0-0)	---	---
% hemorrhaging[g]	0 (0-0)	0 (0-0)	0 (0-0)	0 (0-0)	---	0 (0-3.03)	---	---
% healthy[h]	100 (100-100)	100 (100-100)	100 (100-100)	100 (100-100)	---	100 (97-100)	---	---

[a] All Study III streams were recovering from prior selenium exposures in Study II (2.5 and 10 µg/L) or Study I (30 µg/L).

[b] Cumulative for the stream, i.e., one value per stream.

[c] Among live larvae; the median and range (in parentheses) of the maximum incidence per set of nest observations are given.

[d] A trial was set up whenever sufficient larvae were collected.

[e] Cumulative percent hatch for each cup; the mean and standard deviation (in parentheses) are given.

[f] Mean percent survival to third day after first larva hatched; the standard deviation is given in parentheses.

[g] Among larvae that survived up to third day after first larva hatched; the median and range (in parentheses) of the maximum incidence are given.

[h] Among larvae that survived up to third day after first larvae hatched; assumes the observations of multiple abnormality types always co-occurred in the same organism, this may overestimate the actual % healthy when this assumption is violated. The median and range (in parentheses) of the minimum percent healthy per cup are given.

Table 4-3. Summary of Average Percent Abnormalities for Study II, Field Nest Data

Age	Selenium Treatment															
	Control				2.50 µg/L				10.00 µg/L				Recovering 30.00 µg/L			
	# of obs.	% edema	% lordosis	% hemor.	# of obs.	% edema	% lordosis	% hemor.	# of obs.	% edema	% lordosis	% hemor.	# of obs.	% edema	% lordosis	% hemor.
1	...	...	...	...	1	0	0	11.1	3	24.4	11.1	8.9	1	0	0	0
2	4	0.8	0	0.7	3	1.4	2.4	59.2	6	35.1	0.5	35.5	1	0	0	22.7
3	6	0	0	0	4	0.4	0	26.0	2	78.9	7.4	41.8	5	25.5	0.9	12.4
4	2	0	0	0	1	10.7	0	14.3	...	...	...	...	2	9.3	3.3	5.0
5	2	0	0	0	1	0	25.0	0	...	...	...	...	...	...	...	...
6	5	0	0	0	...	...	...	...	...	...	...	...	1	72.7	0	13.6
7	1	0	0	0	...	...	...	...	...	...	...	...	1	80.0	2.9	17.1

Table 4-4. Summary of Average Percent Abnormalities for Study II, Egg Cup Data

Day	Selenium Treatment																			
	Control					2.50 µg/L					10.00 µg/L					Recovering 30.00 µg/L				
	# of obs.	% edema	% lor-dosis	% hemor.	% healthy[a]	# of obs.	% edema	% lor-dosis	% hemor.	% healthy[a]	# of obs.	% edema	% lor-dosis	% hemor.	% healthy[a]	# of obs.	% edema	% lor-dosis	% hemor.	% healthy[a]
1	22	0	0.4	0.2	99.4	6	0	0	25.6	74.4	12	15.2	7.3	22.8	69.9	6	2.5	2.2	6.0	92.5
2	22	0.5	0.5	0.6	98.9	12	0.4	1.2	1.6	97.4	11	57.7	9.4	27.6	35.1	11	3.1	1.5	3.2	94.4
3	22	0.6	0.2	1.0	99.0	12	2.7	3.1	5.1	91.1	11	97.6	10.0	35.1	2.4	11	33.5	6.4	15.4	64.3
4	22	1.2	0.2	1.5	98.1	12	8.9	1.5	3.2	90.4	10	90.0	15.0	18.4	0.0	11	70.5	1.0	23.4	28.6
5	16	0	0.9	0.0	99.1	11	1.0	4.9	3.1	93.0	4	100.0	19.3	6.3	0.0	9	83.6	6.2	11.1	16.4

[a] % healthy = [# of live larvae - max(# of edema, # of lordosis, # of hemor.)]/# of live larvae * 100.

Table 4-5. Summary of Average Percent Abnormalities for Study III, Field Nest Data[a]

	Selenium Treatment															
	Control				Recovering 2.50 µg/L				Recovering 10.00 µg/L				Recovering 30.00 µg/L			
Age	# of obs.	% edema	% lordosis	% hemor.	# of obs.	% edema	% lordosis	% hemor.	# of obs.	% edema	% lordosis	% hemor.	# of obs.	% edema	% lordosis	% hemor.
0.5	...	...	...	...	1	0	0	0	1	0	0	0	...	...	...	...
1	...	...	...	...	1	0	0	10.7	...	...	...	...	...	...	...	...
1.5		...	...	...	...	...	...	...	...	...	...	...	1	0	0	0
2		...	...	...	1	0	0	0	...	...	...	...	4	1.2	0	4.1
3	1	0	0	0	...	...	...	...	...	...	...	...	1	0	0	0
3.5	...	...	...	...	1	0	0	0	...	...	...	...	...	...	...	...
4	...	...	...	...	...	...	...	...	...	...	...	...	1	0	0	0
4.5	...	...	...	...	1	0	0	0	...	...	...	...	...	...	...	...

[a] All Study III streams were recovering from prior selenium exposures in Study II (2.5 and 10 µg/L) or Study I (30 µg/L).

Table 4-6. Summary of Average Percent Abnormalities for Study III, Egg Cup Data[a]

	Selenium Treatment																			
	Control					Recovering 2.50 µg/L					Recovering 10.00 µg/L					Recovering 30.00 µg/L				
Day	# of obs.	% edema	% lor-dosis	% hemor.	% healthy[b]	# of obs.	% edema	% lor-dosis	% hemor.	% healthy[b]	# of obs.	% edema	% lor-dosis	% hemor.	% healthy[b]	# of obs.	% edema	% lor-dosis	% hemor.	% healthy[b]
1	1	0	0	0	100.0	7	0	0	0	100.0	3	0	0	0.7	100.0	2	0	0	0	100.0
2	5	0	0	0	100.0	10	0	0	0	100.0	3	0	0	0	100.0	5	0	0	0	100.0
3	5	0	0	0	100.0	10	0	0	0	100.0	2	0	0	0	100.0	5	0	0	0	100.0
4	5	0	0.6	0.6	99.4	9	0	0	0	100.0	3	0	0	0	100.0	5	0	0	0	100.0
5	5	0	0	0	100.0	10	0	0.4	0	99.6	3	0	0	0	100.0	5	0	0	0	100.0

[a] All Study III streams were recovering from prior selenium exposures in Study II (2.5 and 10 µg/L) or Study I (30 µg/L).

[b] % healthy = [# of live larvae - max (# of edema, # of lordosis, # of hemor.)]/# of live larvae * 100.

As a part of the analysis, tests for normality and homogeneity of variance were performed on the appropriate data. As appropriate, data transformations were applied to adjust for non-normal model errors.

Two types of models were fit to the data. First, we fit a standard ANOVA model that does not account for time-based effects associated with the repeated sampling of nests or egg cups during the course of the experiment. The response variables were the maximum incidences (e.g., maximum % edema, maximum % hemorrhaging) per nest or cup. For egg cup data, the minimum % healthy was also included as one of the response variables. Note that the value of % healthy was calculated for the live larvae only, and under the assumption that observations of multiple abnormality types always occurred in combination on the same group of "affected" organisms. Therefore, this may result in an overestimation of the actual % healthy when such an assumption is violated. Second, we fit a form of the ANOVA model associated with repeated-measures analysis. This model adjusts the model results for any time-dependent correlations that exist among the data. Each model is described below.

In addition, for Study II, none of the information associated with 30 µg/L selenium treatments was included in the analytical dataset to which the ANOVAs were fit. We excluded this treatment because it represents the recovery from the previous study and therefore is not comparable to the continuous exposure regimes of the 2.5 and 10 µg/L treatments.

4.2.1 Study II Field Nest Data

Experimental Design

The experimental design for the Study II Field Nest data, ignoring time-series effects, is summarized in Table 4-7.

Table 4-7. Experimental Design for Field Nest Data—Study II

Selenium Treatment	Stream	Maximum % Abnormalities from Each Spawn[a]
Control	1	C2, C3, D1, D2, D3, D4
	5	B1, B11, B3, B4, B5, B7, C2, C4
2.50 µg/L	2	D1
	7	B6, C2, C3, C4, C8
10.0 µg/L	3	A11, A12, A21, A22
	8	B1, B3, C1
Recovering 30.0 µg/L	4	No Data
	6	A1, A12, A13-1, A13-2, A2-1, A2-2, A3, A4, A5, A8

[a] Table entries (e.g., C2) represent the spawns that have the values for the response variables.

In this formulation, a single value for each spawn in each nest is the data evaluated by the ANOVA. The table indicates the nest number for each spawn. Where multiple spawns occurred in a nest, the spawns are indexed. Note that the recovering 30 µg/L treatment was not evaluated in the ANOVA. Only the maximum percent abnormalities from each spawn are used in the analysis.

ANOVA Analysis

The following mixed model was performed on the Field Nest Data, with stream considered as the random effect:

$$R_{ijk} = \mu + C_i + S_{j(i)} + \varepsilon_{ijk}$$

where,

R_{ijk} = multivariate response: % edema, % lordosis, and % hemorrhaging (all maximum incidence per nest);

μ = overall mean;

C_i = treatment effect, i = 1 to 3 treatments;

$S_{j(i)}$ = stream effect, j = 1 to 2 streams, nested within treatment, considered as a random effect; and

ε_{ijk} = random error.

A test of normality on the model residuals (ε_{ijk}) and a homogeneity of variance test among treatment levels were performed. An arc-sine square-root transformation was applied to the data when model assumptions of normality and/or homogeneity of variance were violated. In cases where an arc-sine square-root transformation was inadequate, a ranking transformation was substituted. The results are summarized in Table 4-8.

Table 4-8. Tests of Normality and Homogeneity of Variance—Study II Field Nest Data

	Normality			Homogeneity of Variance		
	Max % edema	**Max % lordosis**	**Max % hemor.**	**Max % edema**	**Max % lordosis**	**Max % hemor.**
p-value	0.0001*	0.0001*	0.0103*	<0.0001*	**	<0.0001*

* Significant p-value at $\alpha = 0.05$.
** Non-calculable, given the values in the dataset.

Table 4-8 indicates that the data generally do not satisfy the normality and homogeneity of variance assumptions and thus require transformation. Therefore, an arc-sine square-root transformation was applied to the data. The results of repeated testing are summarized in Table 4-9.

Table 4-9. Tests of Normality and Homogeneity of Variance—Study II Field Nest Data (Arc-sine Square-root Transformed Data)

	Normality			Homogeneity of Variance		
	Max % edema	**Max % lordosis**	**Max % hemor.**	**Max % edema**	**Max % lordosis**	**Max % hemor.**
p-value	0.0004*	0.0013*	0.0032*	<0.0001*	**	<0.0001*

* Significant p-value at $\alpha = 0.05$.
** Non-calculable, given the values in the dataset.

Again, the data did not pass the tests. Because both the raw data and the transformed data failed the normality and homogeneity of variance tests, the ANOVA was also performed on ranks (Iman,

1982[7]). Here, we ranked the response variables from lowest to highest across treatments and reran the ANOVA. Fitting an ANOVA model to ranks may not be appropriate. First, the model residuals are non-normal. Second, the residuals are, by definition, truncated within the range of the lowest and highest rank. Therefore, applying standard hypothesis testing techniques, which required the assumption of normally distributed residuals with zero mean, may not be appropriate. Generally, analysis of ranked data proceeds with nonparametric approaches, thereby negating the distributional assumptions inherent in parametric techniques. However, because nonparametric methods are not available for mixed-model ANOVA designs (the model that is most consistent with the experimental design of this study), the results of the ANOVA were presented on both raw and ranked data. The results of performing the ANOVA on both the raw data and the ranks are summarized in Table 4-10.

Table 4-10. ANOVA Results From PROC MIXED—Study II Field Nest Data

Parameter	Test of Parameter Significance (p-value, raw data)			Test of Parameter Significance (p-value, ranks)		
	Max % edema	Max % lordosis	Max % hemor.	Max % edema	Max % lordosis	Max % hemor.
Treatment	0.0233*	0.5510	0.0568	0.0271*	0.1907	0.0129*
Stream(treatment), random	**	0.2669	**	**	0.6313	**

* Significant p-value at $\alpha = 0.05$.
** Missing p-values resulted from the variance components estimates being zero.

The missing p-values in Table 4-10 and in the following tables from PROC MIXED result from the fact that the variance component estimates are zero, and therefore test statistics cannot be computed. The zero variance estimates may arise for several reasons. For example, the variability in the data may be large enough to produce a negative estimate (negative values were restricted to zeros), even though the true value of the variance component is positive; data may contain outliers and a different model for interpreting the data may be appropriate. Although alternative models might better address the missing p-value issue, such alternative models were not used, because they were judged to be incompatible with the experimental design.

Note that the test on the random effect, stream(treatment), provided by PROC MIXED is based on large sample asymptotic theory[8] and therefore may not be appropriate in this application. The random effect results should therefore be interpreted with caution. Table 4-10 indicates that ANOVAs on the raw and ranked data provide generally consistent results with respect to determination of significant parameters. Selenium concentration appears to be an important component of the maximum percent incidence determinations for % edema and perhaps % hemorrhaging, but not for % lordosis. To examine which treatments differ, the means test from PROC MIXED, which adjusts the standard error of the treatment for the random effects in the model, was performed. The results are presented in Table 4-11. Dunnett's multiple comparison against control was performed on this dataset and all subsequent datasets.

The results indicate that, for maximum % edema, the selenium treatment mean of 10 µg/L differed from the control mean, using both the raw data and the rank transformed data. For maximum %

[7] Iman, R.L. 1982. Some Aspects of the Rank Transform in Analysis of Variance Problems. Seventh Annual SAS Users Group International Conference.

[8] Self, S.G., and K.Y. Liang. 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J. Amer. Stat. Assoc. 82:605-610.

hemorrhaging, the selenium treatment mean of 2.5 µg/L differed from the control mean for the rank transformed data (treatment effects were marginally not significant at 2.5 µg/L for the raw data). In addition, the mean of 10 µg/L differed from the control mean for ranked data only for maximum % hemorrhaging.

Table 4-11. Means Test (Dunnett's) From PROC MIXED—Study II Field Nest Data

Selenium Treatments	Test of the Difference (p-value, raw data)			Test of the Difference (p-value, ranks)		
	Max % edema	Max % lordosis	Max % hemor.	Max % edema	Max % lordosis	Max % hemor.
2.5 - 0.0 µg/L	0.962	0.478	0.050*, [a]	0.358	0.285	0.022*
10.0 - 0.0 µg/L	0.020*	0.687	0.133	0.020*	0.166	0.014*

* Significant p-value at $\alpha = 0.05$.

[a] Treatment effects were marginally not significant at $\alpha = 0.05$, see Table 4-10.

Repeated-Measures ANOVA

For the Field Nest Data, each nest was sampled two or three times a week using a plastic tube to determine the presence of embryos and larvae. Table 4-12 presents the experimental design for the Field Nest Data using time-based samples. If the nest is considered the sampling unit, then a repeated-measures analysis is appropriate. The table shows that the number of times a nest is sampled is not consistent over the experiment. Due to a large number of missing values in the dataset for ages other than two and three days, only these data were retained in the dataset for the repeated-measures analysis. The repeated-measures ANOVA model corresponding to the experimental design can be written as follows:

$$R_{ijklm} = \mu + C_i + S_{j(i)} + N_{k(j)} + A_l + (CA)_{il} + \varepsilon_{ijklm}$$

where,

R_{ijklm} = multivariate response: % edema, % lordosis, and % hemorrhaging;

μ = overall mean;

C_i = treatment effect, i = 1 to 3 treatments;

$S_{j(i)}$ = stream effect, j = 1 to 2 streams, nested within treatment, considered as a random effect;

$N_{k(j)}$ = spawn effect, nested within stream, considered as a random effect, and the subject for the repeated-measures analysis;

A_l = age effect, k = 1 to 7 days;

$(CA)_{il}$ = interaction between treatment and age, i = 1 to 3, k = 1 to 7; and

ε_{ijklm} = random error.

Before the analysis, normality and homogeneity of variance were tested. The results are summarized in Table 4-13. Because the raw data failed the test for normality, the arc-sine square-root transformation was applied to the data, and tests of normality and homogeneity of variance were then performed on the transformed data. These results are also included in Table 4-13.

Table 4-12. Repeated-measures Design for Field Nest Data—Study II

Selenium Treatment	Stream	Nest	% Abnormalities at Age (Days)						
			1	2	3	4	5	6	7
Control	1	D1			x			x	
		D2			x			x	
		D3		x			x		
		D4		x					
	5	B1				x			
		B11		x				x	
		B3			x			x	
		B4				x			x
		B5			x			x	
		B7					x		
		C3			x				
		C4		x	x				
2.5 µg/L	2	D1		x	x		x		
	7	B6			x				
		C3		x	x	x			
		C4	x	x					
		C8			x				
10 µg/L	3	A1-1		x	x				
		A1-2	x	x					
		A2-1	x	x					
		A2-2		x					
	8	B1			x				
		B3		x					
		C1	x	x					
Recovering 30 µg/L	4	No data							
	6	A12		x					
		A13-1			x	x			
		A13-2			x				
		A2-1				x			
		A2-2			x				
		A3			x			x	
		A4			x				
		A5				x			x
		A8	x						

Table 4-13. Tests of Normality and Homogeneity of Variance—Study II Field Nest Data (Repeated-measures Analysis)

p-value using	Normality			Homogeneity of Variance		
	% edema	% lordosis	% hemor.	% edema	% lordosis	% hemor.
Raw data	0.0001*	0.0001*	0.0570	<0.0001*	**	<0.0001*
Arc-sine transformed data	0.0005*	0.0001*	0.0478*	<0.0001*	**	<0.0001*

* Significant p-value at $\alpha = 0.05$.

** Missing p-values resulted from all zero values of % lordosis for the control. As a result, the standard deviation cannot be computed, and the test cannot be performed.

Because both the raw data and the transformed data failed the normality test, the ANOVA was performed on ranked data as well. The ranks were constructed independently for ages of 2 and 3 days. The results for the ANOVAs are presented in Table 4-14.

Table 4-14. Repeated-measures ANOVA From PROC MIXED—Study II Field Nest Data

Parameters	p-value (raw data)			p-value (ranks)		
	% edema	% lordosis	% hemor.	% edema	% lordosis	% hemor.
Treatment	0.0294*	0.1518	0.1211	0.2023	0.1788	0.0443*
Age	0.1617	0.2038	0.5031	0.4264	0.6817	0.5881
Age × Treatment	0.1389	0.0140*	0.4758	0.3318	0.0215*	0.4965
Stream(treatment), random	**	0.8353	**	0.5790	**	**

* Significant p-value at $\alpha = 0.05$.

** Missing p-values resulted from the variance components estimates being zeroes.

Table 4-14 indicates that the interaction between age and treatment is significant for % lordosis, using both the raw data and rank transformed data; or, in other words, selenium effects differ for different ages. As a result, the means test on the age × treatment interaction for % lordosis should be examined. The results ("Tests of Effect Slices" in SAS® output and Table 5-1) show that 10 µg/L selenium had a significant effect on % lordosis at age 3 (raw data). At 2.5 µg/L, significant effects on % lordosis also occurred at age 2 (ranked data). Regarding treatment effects, repeated measures analysis indicates that selenium had a significant effect at 10 µg/L on % edema (raw data) and % hemorrhaging (ranked data). Age does not appear to have a significant effect on any of the three measures. The complete output from PROC MIXED is presented in Appendix B. The results of the means test (Dunnett's) for the treatment effect from PROC MIXED are summarized in Table 4-15.

Table 4-15. Means Test (Dunnett's) for Repeated-measures ANOVA From PROC MIXED—Study II Field Nest Data

Selenium Treatment	Test of the difference (p-value, raw data)			Test of the difference (p-value, ranks)		
	% edema	% lordosis	% hemor.	% edema	% lordosis	% hemor.
2.5 - 0.0 µg/L	0.999	0.638[a]	0.121	0.681	0.167[a]	0.075
10.0 - 0.0 µg/L	0.027*	0.118[a]	0.163	0.159	0.252[a]	0.041*

* Significant p-value at $\alpha = 0.05$.

[a] Significant p-value was found at $\alpha = 0.05$ for treatment × interaction (see Table 4-14).

4.2.2 Study II Egg Cup Data

ANOVA Analysis

The experimental design for Study II Egg Cup data, ignoring time effects, is summarized in Table 4-16. Again, the information associated with the recovering 30 µg/L selenium treatment was not used in the analysis. Only incidence rates to the third day of sampling were evaluated. The data analysis was truncated at three days because the starvation effect will more likely be confounded with the

selenium effect after the third day of the experiment. Based on the experimental design, the following ANOVA model was implemented:

$$R_{ijkl} = \mu + C_i + S_{j(i)} + N_{k(j)} + \varepsilon_{ijkl}$$

where,

R_{ijkl} = % hatch,
% survival (to the third day),
% edema (maximum incidence to the third day),
% lordosis (maximum incidence to the third day),
% hemorrhaging (maximum incidence to the third day),
% healthy[9] (minimum incidence to the third day);

μ = overall mean;

C_i = treatment effect, i = 1 to 3 treatments;

$S_{j(i)}$ = stream effect, j = 1 to 2 streams, nested within treatment, considered as a random effect;

$N_{k(j)}$ = spawn effect, nested within stream, considered as a random effect; and

ε_{ijkl} = random error.

Tests for normality on model residuals and homogeneity of variance among treatments on the response variables were performed on the dataset. The results are summarized in Table 4-17.

Because the data failed the normality test for most of the response variables, and failed the homogeneity of variance test for all the variables, the arc-sine square-root transformation was applied to the data, and tests of normality on model residuals and homogeneity of variance on the transformed data were performed. The results are summarized in Table 4-18.

This transformation resulted in an improved normality test on model residuals and homogeneity of variance test for only one out of six response variables, i.e., maximum % hemorrhaging and maximum % edema, respectively. The dataset was then rank transformed, and the ANOVA was performed on both the raw data and the rank transformed data. The results are summarized in Table 4-19.

Examination of Table 4-19 shows a treatment effect for maximum % edema and minimum % healthy using both raw and ranked data. Means testing on these parameters show a significant difference between the mean control effect and the mean 10 µg/L selenium effect. The p-values of the means test are provided in Table 4-20.

Repeated-measures ANOVA

For the egg cup data, each cup was sampled for five days to observe the abnormalities among live larvae. To account for any time-dependent correlation, a repeated-measures ANOVA was performed. The experimental design for the repeated-measures analysis is summarized in Table 4-21. Before performing any analysis, the egg cup data were examined for the occurrence of zero incidence

[9] % healthy = [# of live larvae - max(# of edema, # of lordosis, # of hemor.)] /# of live larvae * 100 for each observation. This calculation assumes the observations of multiple abnormality types always co-occurred in the same group of "affected" organisms; hence the value may be overestimated when a violation of the assumption occurs.

values, as summarized in Table 4-22. Accompanying figures showing the zero percent abnormalities are also presented in Appendix D. The analysis is performed on two datasets: the full time-series dataset and the partial time-series dataset.

Table 4-16. Experimental Design for Study II Egg Cup Data

Selenium Treatment	Stream	Spawn (nest)	% Survival to the Third Day, Maximum % Abnormalities to the Third Day[a]
Control	1	1 (D1)	S1, S2
		2 (D2)	S1, S2
		3 (D3)	S1, S2
		4 (D4)	S1, S2
		5 (C2)	S1, S2
		6 (C3)	S1, S2
	5	1 (B3)	S1, S2
		2 (B7)	S1, S2
		3 (C2)	S1, S2
		4 (C3)	S1, S2
		5 (C4)	S1, S2
2.5 µg/L	2	No Data	
	7	1 (C3)	S1, S2, S3, S4, S5, S6
		3 (C4)	S1, S2
		4 (C2)	S1, S2
		5 (C8)	S1, S2
10 µg/L	3	1 (A1)	S1, S2, S3, S4
		2 (A1)	S1, S2 (only 1 day's data available)
		3 (A2)	S1, S2
	8	1 (B1)	S1, S2
		2 (B3)	S1, S2
Recovering 30 µg/L	4	No Data	
	6	1 (A4)	S1, S2
		2 (A3)	S1
		3 (A8)	S1, S2
		4 (A11)	S1, S2
		5 (A13)	S1, S2
		6 (A13)	S1, S2

[a] Sx: Multivariate response observed for sample x (or cup x)

Table 4-17. Tests of Normality and Homogeneity of Variance—Study II Egg Cup Data

	Normality					
	% hatch	% survival	Max % edema	Max % lordosis	Max % hemor.	Min % healthy
p-value	0.2096	0.8156	0.0001*	0.0001*	0.0008*	0.0001*

Homogeneity of Variance						
p-value	<0.0001*	0.0004*	<0.0001*	<0.0001*	<0.0001*	<0.0001*

* Significant p-value at $\alpha = 0.05$.

Table 4-18. Tests of Normality and Homogeneity of Variance—Study II Egg Cup Data (Arc-sine Square-root Transformed Data)

Normality						
	% hatch	% survival	Max % edema	Max % lordosis	Max % hemor.	Min % healthy
p-value	0.9239	0.8939	0.0001*	0.0001*	0.2270	0.0040*
Homogeneity of Variance						
p-value	0.0015*	0.0004*	0.0638	0.0011*	<0.0001*	<0.0001*

* Significant p-value at $\alpha = 0.05$.

Table 4-19. ANOVA Results From PROC MIXED—Study II Egg Cup Data

Factors	% hatch	% survival	Max % edema	Max % lordosis	Max % hemor.	Min % healthy
p-value (raw data)						
Treatment	0.2770	0.2629	0.0012*	0.0597	0.0841	0.0129*
Stream(treat), random	**	**	**	**	**	**
Spawn(stream), random	0.0118*	0.0305*	0.0993	**	0.0068*	0.0148*
p-value (rank transformed data)						
Treatment	0.4061	0.3852	0.0362*	0.0548	0.0647	0.0491*
Stream(treat), random	**	**	**	**	**	**
Spawn(stream), random	0.0200*	0.0502	0.0514	**	0.0289*	0.0344*

* Significant p-value at $\alpha = 0.05$.

** Missing p-values resulted from the variance components estimates being zeroes.

Table 4-20. Means Test (Dunnett's) From PROC MIXED—Study II Egg Cup Data

Selenium Treatment	% hatch	% survival	Max % edema	Max % lordosis	Max % hemor.	Min % healthy
Test of the difference (p-value, raw data)						
2.5 - 0.0 µg/L	0.504	0.982	0.562	0.609	0.297	0.132
10.0 - 0.0 µg/L	0.248	0.230	0.001*	0.050*	0.070	0.010*
Test of the difference (p-value, ranks)						
2.5 - 0.0 µg/L	0.417	0.890	0.547	0.508	0.195	0.223

Table 4-20. Means Test (Dunnett's) From PROC MIXED—Study II Egg Cup Data

Selenium Treatment	% hatch	% survival	Max % edema	Max % lordosis	Max % hemor.	Min % healthy
10.0 - 0.0 µg/L	0.501	0.332	0.030*	0.045*[a]	0.055	0.040*

* Significant p-value at $\alpha = 0.05$.
[a] Treatment effects were marginally nonsignificant at $\alpha = 0.05$ (see Table 4-19).

Table 4-21. Study II Egg Cup Data—Repeated-measures Design

Selenium Treatment	Stream	Spawn (nest)	Cup (or Sample)	Multivariate Response on Day				
				1	2	3	4	5
Control	1	1 (D1)	1	x	x	x	x	x
			2	x	x	x	x	x
		2 (D2)	1	x	x	x	x	x
			2	x	x	x	x	x
		3 (D3)	1	x	x	x	x	x
			2	x	x	x	x	x
		4 (D4)	1	x	x	x	x	x
			2	x	x	x	x	x
		5 (C2)	1	x	x	x	x	x
			2	x	x	x	x	x
		6 (C3)	1	x	x	x	x	x
			2	x	x	x	x	x
	5	1 (B3)	1	x	x	x	x	x
			2	x	x	x	x	x
		2 (B7)	1	x	x	x	x	
			2	x	x	x	x	x
		3 (C2)	1	x	x	x	x	
			2	x	x	x	x	
		4 (C3)	1	x	x	x	x	x
			2	x	x	x	x	
		5 (C4)	1	x	x	x	x	
			2	x	x	x	x	
2.5 µg/L	2	No Data						
	7	1 (C3)	1		x	x	x	x
			2		x	x	x	x
			3		x	x	x	x
			4		x	x	x	x
			5	x	x	x	x	x
			6	x	x	x	x	x
		3 (C4)	1	x	x	x	x	x
			2	x	x	x	x	x

Table 4-21. Study II Egg Cup Data—Repeated-measures Design

Selenium Treatment	Stream	Spawn (nest)	Cup (or Sample)	Multivariate Response on Day				
				1	2	3	4	5
		4 (C2)	1	x	x	x	x	x
			2	x	x	x	x	
		5 (C8)	1		x	x	x	x
			2		x	x	x	x

Table 4-21. Study II Egg Cup Data—Repeated-measures Design

Selenium Treatment	Stream	Spawn (nest)	Cup (or Sample)	Multivariate Response on Day				
				1	2	3	4	5
10 µg/L	3	1 (A1)	1	x	x	x	x	x
			2	x	x	x	x	x
			3	x	x	x	x	
			4	x	x	x	x	
		2 (A1)	1	x	x	x	x	
			2	x				
		3 (A2)	1	x	x	x	x	
			2	x	x	x	x	
	8	1 (B1)	1	x	x	x	x	x
			2	x	x	x	x	x
		2 (B3)	1	x	x	x		
			2	x	x	x	x	
Recovering 30 µg/L	4	No Data						
	6	1 (A4)	1	x	x	x	x	x
			2	x	x	x	x	x
		2 (A3)	1		x	x	x	x
			No Data					
		3 (A8)	1	x	x	x	x	x
			2		x	x	x	x
		4 (A11)	1	x	x	x	x	
			2	x	x	x	x	
		5 (A13)	1		x	x	x	x
			2	x	x	x	x	x
		6 (A13)	1		x	x	x	x
			2		x	x	x	x

Table 4-22. Summary of Zero Abnormalities for Study II, Egg Cup Data

Day	Total # of Obs.	Total # (Percent) of Zero Abnormalities		
		Edema	Hemorrhaging	Lordosis
1	46	41 (89.1%)	34 (73.9%)	40 (87.0%)
2	56	37 (66.1%)	35 (62.5%)	37 (66.1%)
3	56	32 (57.1%)	28 (50.0%)	41 (73.2%)
4	55	30 (54.5%)	28 (50.9%)	43 (78.2%)
5	40	26 (65.0%)	28 (70.0%)	32 (80.0%)

Full time-series dataset

Due to the relatively high incidence of zero abnormalities at Day 1 for % edema and % lordosis (see Table 4-22), these observations were eliminated from the repeated-measures ANOVA. The resulting dataset is referred to as a "full time-series" dataset. The following model based on the experimental design presented in Table 4-21 was applied to this dataset:

$$R_{ijklmn} = \mu + C_i + S_{j(i)} + N_{k(j)} + Cp_{l(k)} + A_m + (CA)_{im} + \varepsilon_{ijklmn}$$

where,

R_{ijklmn} = multivariate response: % edema, % lordosis, % hemorrhaging, and % healthy;

μ = overall mean;

C_i = treatment effect, i = 1 to 3 treatments;

$S_{j(i)}$ = stream effect, j = 1 to 2 streams, nested within treatment, considered as a random effect;

$N_{k(j)}$ = spawn effect, nested within streams, considered as a random effect;

$Cp_{l(k)}$ = cup effect, nested within spawns, considered as a random effect;

A_m = time effect, m = 1 to 5 days for % hemorrhaging and % healthy, m = 1 to 4 days for % edema and % lordosis;

$(CA)_{im}$ = interaction between treatment and time, i = 1 to 3, k = 1 to 5 (or 1 to 4); and

ε_{ijklmn} = random error.

Again, tests for normality on model residuals and homogeneity of variance on the response variables were performed on this dataset. The results are summarized in Table 4-23. Because the data generally failed the normality test for % edema, % lordosis, and % hemorrhaging, the arc-sine square-root transformation was applied to the data, and tests of normality on model residuals and homogeneity of variance on the response variables were performed on the transformed data. The results are also summarized in Table 4-23.

Because both the raw data and the transformed data failed the normality and the homogeneity of variance tests, the ANOVA was also performed on the ranked response variables. The results are summarized in Table 4-24.

Examination of Table 4-24 shows that the interaction between age (day) and treatment for the full time-series data sets is significant for % edema (raw data), % hemorrhaging (raw and ranked), and % healthy (raw and ranked), indicating that selenium effects differ for different ages. As a result, means tests were conducted on the day x treatment interaction for these variables. The results ("Tests of Effect Slices" in SAS® output and Table 5-2) show that 10 µg/L selenium had a significant effect on % edema (days 2-5, raw), % hemorrhaging (days 1-3, raw; days 1-4, ranked), and % healthy (days 1-5, raw and ranked). At 2.5 µg/L, selenium had a significant effect on % hemorrhaging (day 1, raw) and % healthy (day 1, raw). Detailed information can be found in Appendix B "PROC MIXED Output from SAS®." For measures where no significant interaction between age and

Table 4-23. Tests of Normality and Homogeneity of Variance—Study II Egg Cup Data (Repeated-measures Analysis Full Time-series Dataset)

p-value using	Normality				Homogeneity of Variance			
	% edema	% lordosis	% hemor.	% healthy	% edema	% lordosis	% hemor.	% healthy
Raw data	0.0001*	0.0001*	0.0001*	0.0001*	<0.0001*	<0.0001*	<0.0001*	<0.0001*
Arc-sine square-root transformed data	0.0001*	0.0001*	0.0001*	0.0001*	<0.0001*	<0.0001*	<0.0001*	<0.0001*

* Significant p-value at $\alpha = 0.05$.

Table 4-24. Repeated-measures ANOVA From PROC MIXED—Study II Egg Cup Data (Full Time-series Dataset)

Factor	p-value (raw data)				p-value (rank transformed data)			
	% edema	% lordosis	% hemor.	% healthy	% edema	% lordosis	% hemor.	% healthy
Treatment	0.0559	0.2974	0.0510	0.0080*	0.0375*	0.0391*	0.0303*	0.0130*
Day	0.0001*	0.3980	0.0174*	0.0001*	0.0001*	0.0041*	0.0041*	0.0002*
Day × Treatment	0.0001*	0.7361	0.0064*	0.0001*	0.0804	0.0856	0.0104*	0.0302*
Stream(treat), random	0.3653	0.3770	**	**	0.7730	**	0.8831	**
Spawn(stream), random	0.1691	**	0.1920	0.0238*	0.0515	**	0.5898	0.3136

* Significant p-value at $\alpha = 0.05$.

** Missing p-values resulted from the variance components estimates being zeroes.

treatment was found, repeated measures analysis and subsequent means testing indicate that selenium had a significant effect at 10 µg/L on % edema (ranked) and % lordosis (ranked). Table 4-25 contains the complete results of the means testing. Age (day) effects were also significant on % edema and % lordosis for the ranked data.

Table 4-25. Means Test (Dunnett's) for Repeated-Measures ANOVA From PROC MIXED—Study II Egg Cup Data (Full Time-Series Dataset)

Selenium Treatment	Test of the Difference (p-value, raw data)				Test of the Difference (p-value, ranks)			
	% edema	% lordosis	% hemor.	% healthy	% edema	% lordosis	% hemor.	% healthy
2.5 - 0.0 µg/L	0.977[a]	0.952	0.256[a]	0.211[a]	0.684	0.269	0.171[a]	0.122[a]
10.0 - 0.0 µg/L	0.051[a]	0.264	0.042*,[a]	0.007*,[a]	0.032*	0.032*	0.025*,[a]	0.011*,[a]

* Significant p-value at $\alpha = 0.05$.

[a] Significant p-value was found at $\alpha = 0.05$ for treatment x day interaction (see Table 4-24).

Partial time-series dataset

To address the concern that, during the last days of the experiment, a starvation effect could affect fish response and thus confound the selenium effect, a second ANOVA model was implemented.

This model, termed the partial time-series repeated-measures ANOVA, was performed on the egg cup data after eliminating observations for Days 4 and 5 from the full time-series dataset. The form of the ANOVA model is the same as that described above.

The tests for normality on model residuals and the homogeneity of variance among treatments on the response variables for this partial time-series dataset were performed. The results are summarized in Table 4-26. Because the data generally failed the normality and homogeneity of variance tests, an arc-sine square-root transformation was applied to the data, and tests of normality and homogeneity of variance were performed on the transformed data. The results are also summarized in Table 4-26.

Table 4-26. Tests of Normality and Homogeneity of Variance—Study II Egg Cup Data (Repeated-measures Analysis—Partial Time-series Dataset)

	Normality				Homogeneity of Variance			
p-value using	% edema	% lordosis	% hemor.	% healthy	% edema	% lordosis	% hemor.	% healthy
Raw data	0.0001*	0.0001*	0.0001*	0.0001*	<0.0001*	<0.0001*	<0.0001*	<0.0001*
Arc-sine square-root transformed data	0.0001*	0.0001*	0.0001*	0.0008*	<0.0001*	<0.0001*	<0.0001*	<0.0001*

* Significant p-value at $\alpha = 0.05$.

Because both the raw data and the transformed data failed the normality and the homogeneity of variance tests, the ANOVA was also performed on the ranked response variables. The results are summarized in Table 4-27.

Table 4-27. Repeated-measures ANOVA From PROC MIXED—Study II Egg Cup Data (Partial Time-series Dataset)

	p-value (raw data)				p-value (rank transformed data)			
Factor	% edema	% lordosis	% hemor.	% healthy	% edema	% lordosis	% hemor.	% healthy
Treatment	0.0945	0.0618	0.0580	0.0227*	0.0201*	0.1427	0.0384*	0.0244*
Day	0.0001*	0.6258	0.2804	0.0001*	0.7070	0.8787	0.0736	0.6089
Day × Treatment	0.0001*	0.7862	0.0259*	0.0001*	0.4508	0.4639	0.0071*	0.0154*
Stream(treat), random	0.3504	**	**	**	**	0.4679	**	**
Spawn(stream), random	0.8744	**	0.1218	0.0189*	0.1198	**	0.0600	0.1102

* Significant p-value at $\alpha = 0.05$.
** Missing p-values resulted from the variance components estimates being zeroes.

Examination of Table 4-27 shows that the interaction between age (day) and treatment for the partial time-series data sets is significant for % edema (raw data), % hemorrhaging (raw and ranked), and % healthy (raw and ranked), indicating that selenium effects differ for different ages. These results are consistent with those of the full time-series dataset (Table 4-25). Means tests were conducted on the day x treatment interaction for these variables. The results ("Tests of Effect Slices" in SAS®

output and Table 5-2) show that 10 µg/L selenium had a significant effect on % edema (days 2 and 3, raw), % hemorrhaging (days 1-3, raw and ranked), and % healthy (days 1-3, raw and ranked). At 2.5 µg/L, selenium had a significant effect on % hemorrhaging (day 1, raw) and % healthy (day 1, raw). Detailed information can be found in Appendix B "PROC MIXED Output from SAS®." For measures where no significant interaction between age and treatment was found, repeated measures analysis and subsequent means testing indicate that selenium had a significant effect at 10 µg/L on % edema (ranked) and % lordosis (ranked). Table 4-28 contains the complete results of the means testing. No statistically significant age (day) effects were found for measures shown to have no significant age and treatment interaction.

Table 4-28. Means Test (Dunnett's) for Repeated-measures ANOVA From PROC MIXED—Study II Egg Cup Data (Partial Time-series Dataset)

Selenium Treatment	Test of the Difference (p-value, raw data)				Test of the Difference (p-value, ranks)			
	% edema	% lordosis	% hemor.	% healthy	% edema	% lordosis	% hemor.	% healthy
2.5 - 0.0	0.998[a]	0.563	0.252[a]	0.300[a]	0.723	0.918	0.259[a]	0.235[a]
10.0 - 0.0	0.089[a]	0.051	0.048*,[a]	0.018*,[a]	0.017*	0.126	0.031*,[a]	0.020*,[a]

* Significant p-value at $\alpha = 0.05$.

[a] Significant p-value was found at $\alpha = 0.05$ for treatment × day interaction (see Table 4-27).

4.2.3 Study III Field Nest Data

ANOVA Analysis

The experimental design for the Study III Field Nest Data, ignoring time effects, is summarized in Table 4-29. As before, observations associated with the 30 µg/L selenium treatment were not evaluated.

Table 4-29. Experimental Design for Field Nest Data—Study III

Selenium Treatment	Stream	Maximum % Abnormalities from Each Spawn
Control	1	1, 2, 3, 4
	5	1, 2, 3
2.5 µg/L	2	1, 2, 3, 4
	7	1, 2
10.0 µg/L	3	1, 2, 3
	8	1, 2, 3, 4
30.0 µg/L	4	1, 2, 3, 4, 5, 6, 7
	6	No Data

The following mixed model was performed on the Study III Field Nest Data, with stream considered as the random effect:

$$R_{ijk} = \mu + C_i + S_{j(i)} + \varepsilon_{ijk}$$

where,

R_{ijk} = multivariate response: % edema, % lordosis, and % hemorrhaging (all maximum incidence per nest);

μ = overall mean;

C_i = treatment effect, i = 1 to 3 treatments;

$S_{j(i)}$ = stream effect, j = 1 to 2 streams, nested within treatment, considered as a random effect; and

ε_{ijk} = random error.

As before, the normality of the model residuals and the homogeneity of variance among treatment levels for the dependent variables were tested. The results are summarized in Table 4-30. The arc-sine square-root transformation was also applied to the dataset, and the tests for normality and homogeneity of variance on the transformed data were performed and also summarized in Table 4-30.

Table 4-30. Tests of Normality and Homogeneity of Variance—Study III Field Nest Data

p-value using	Normality			Homogeneity of Variance		
	Max % edema	Max % lordosis	Max % hemor.	Max % edema	Max % lordosis	Max % hemor.
Raw data	**	**	0.0001*	**	**	***
Arc-sine square-root transformed data	**	**	0.0001*	**	**	***

* Significant p-value at $\alpha = 0.05$.
** Missing p-values are due to all zero values in the dataset.
*** Not calculable, given the values in the dataset.

The unavailability of test results (missing p-values) for maximum % edema and % lordosis results from the fact that the values for these variables are all zeroes for treatments 0, 2.5, and 10 µg/L in the dataset. Because both the raw data and the transformed data failed the normality test for maximum % hemorrhaging, the ANOVA was performed on the ranked data. The ANOVA results on both the raw data and the rank transformed data are summarized in Table 4-31.

Table 4-31. ANOVA Results From PROC MIXED—Study III Field Nest Data

Parameter	Test of Parameter Significance (p-value, raw data)			Test of Parameter Significance (p-value, ranks)		
	Max % edema	Max % lordosis	Max % hemor.	Max % edema	Max % lordosis	Max % hemor.
Treatment	NA	NA	0.3360	NA	NA	0.3703
Stream(treatment), random	NA	NA	**	NA	NA	0.9384

* Significant p-value at $\alpha = 0.05$.
** Missing p-values resulted from the variance components estimates being zeroes.
NA means not available, because all values were zeroes.

Because the treatment is not significant for maximum % hemorrhaging, the means test was not performed.

Repeated-Measures ANOVA
Due to a large number of missing values for "age" in the dataset, the repeated-measures ANOVA was not performed.

4.2.4 Study III Egg Cup Data

ANOVA Analysis
The experimental design for Study III Egg Cup data, ignoring time effects, is summarized in Table 4-32. Again, observations associated with the 30 µg/L selenium treatment are not included in the analysis. The following multivariate ANOVA model was performed on this dataset based on the experimental design:

$$R_{ijk} = \mu + C_i + S_{j(i)} + \varepsilon_{ijk}$$

where,

R_{ijk} = % hatch,
% survival (to the third day),
% edema (maximum incidence to the third day),
% lordosis (maximum incidence to the third day),
% hemorrhaging (maximum incidence to the third day),
% healthy (minimum incidence to the third day);

μ = overall mean;

C_i = treatment effect, i = 1 to 3 treatments;

$S_{j(i)}$ = stream effect, j = 1 to 2 streams, nested within treatment, considered as a random effect; and

ε_{ijk} = random error.

Again, tests for normality of model residuals and homogeneity of variance of the response variables were performed on this dataset. The results are summarized in Table 4-33.

The missing p-values in the above table result from the occurrence of all zero values for maximum % edema and % lordosis per spawn in the dataset. Although two variables failed normality tests (max % hemorrhaging, %healthy) and only one variable failed the homogeneity of variance tests (max % hemorrhaging), the arc-sine square-root transformation was applied to all variables in the dataset, and tests of normality and homogeneity of variance were performed on the model using the transformed data. The results are summarized in Table 4-34.

Although only one variable (maximum % hemorrhaging) failed the normality and homogeneity of variance tests in both the raw data and transformed datasets, ANOVAs on both the raw data and the rank transformed data were performed. The results are summarized in Table 4-35. Because the treatment effects are not significant for the response variables, the means test were not performed.

Repeated-Measures ANOVA
Due to predominantly zero values for the response variables in this dataset, the repeated-measures ANOVA was not performed.

Table 4-32. Experimental Design for Egg Cup Data—Study III

Selenium Treatment	Stream	Spawn (nest)	Maximum Incidence to the Third Day[a]
Control	1	1 (B5)	S1
		2 (B1)	S1
	5	1 (A1)	S1
		2 (B1)	S1
		3 (B1)	S2
Recovering 2.5 µg/L	2	1 (D2)	S1
		2 (D6)	S1
		3 (A1)	S1
	7	1 (A3)	S1
		2 (A3)	S2
		3 (A4)	S1
		4 (B2)	S1
		5 (A3)	S1
		6 (B3)	S1
		7 (B4)	S1
Recovering 10 µg/L	8	1 (A1)	S1
		2 (A2)	S1
		3 (A5)	S1
	3	No Data	
Recovering 30 µg/L	4	1 (C1)	S1
		2 (C3)	S1
		3 (C4)	S1
		4 (C6)	S1
		5 (C7)	S1
	6	No Data	

[a] S1: multivariate responses for sample 1 (or cup 1).

Table 4-33. Tests of Normality and Homogeneity of Variance—Study III Egg Cup Data

Normality						
	% hatch	**% survival**	**Max % edema**	**Max % lordosis**	**Max % hemor.**	**% healthy**
p-value	0.4170	0.4474	**	**	0.0001*	0.0001*
Homogeneity of Variance						
p-value	0.6849	0.1543	**	**	<0.0001*	**

* Significant p-value at $\alpha = 0.05$.
** Missing p-values are due to all zero values in the dataset.

Table 4-34. Tests of Normality and Homogeneity of Variance—Study III Egg Cup Data (Arc-sine Square-root Transformed Data)

	% hatch	% survival	Max % edema	Max % lordosis	Max % hemor.	% healthy
Normality						
p-value	0.4781	0.6998	**	**	0.0001*	0.0001*
Homogeneity of Variance						
p-value	0.9789	0.1324	**	**	<0.0001*	**

* Significant p-value at $\alpha = 0.05$.
** Missing p-values are due to all zero values in the dataset.

Table 4-35. ANOVA Results From PROC MIXED—Study III Egg Cup Data

Factors	% hatch	% survival	Max % edema	Max % lordosis	Max % hemor.	% healthy
p-value (raw data)						
Treatment	0.4628	0.6525	NA	NA	0.2424	NA
Stream(treat), random	**	**	NA	NA	**	NA
Spawn(stream), random	0.0065*	0.0064*	NA	NA	0.3840	NA
p-value (rank transformed data)						
Treatment	0.4662	0.5861	NA	NA	0.2424	NA
Stream(treat), random	**	**	NA	NA	**	NA
Spawn(stream), random	0.0080*	0.0079*	NA	NA	0.0276*	NA

* Significant p-value at $\alpha = 0.05$.
** Missing p-values resulted from the negative variance components estimates.
NA means not available because all values were zeroes.

SECTION 5
SUMMARY AND CONCLUSIONS

The statistical findings from Studies II and III are summarized below. Additional summaries of the statistical results of selenium effects on progeny in Study II are provided in Tables 5-1 and 5-2, respectively.

Bluegill Survival and Growth

- This study found that selenium concentrations of 2.5 and 10 µg/L did not significantly influence survival or growth of juvenile and adult bluegills.

Adult Spawning Activity and Effects on Progeny (Field Nest Data)

Study II Results

- ***Effects of 2.5 µg/L Selenium.*** ANOVA results on maximum % abnormalities from the Field Nest Study II indicate that 2.5 µg/L selenium resulted in statistically significant effects on larvae for one of the three measures (maximum % hemorrhaging) based on ranked-transformed data ($p<0.05$, Table 5-1). Repeated measures analysis indicates a significant interaction between age and treatment effects for % lordosis. Subsequent means testing indicates that 2.5 µg/L selenium resulted in statistically significant effects on % lordosis for two-day old larvae using ranked-transformed data.

- ***Effects of 10 µg/L Selenium.*** ANOVA results on maximum % abnormalities from the Field Nest Study II indicate that 10 µg/L selenium resulted in statistically significant effects on larvae for two of the three measures (maximum % edema—raw and ranked data; maximum % hemorrhaging—raw data, Table 5-1). Repeated measures analysis and subsequent means testing indicate that 10 µg/L selenium resulted in statistically significant effects on % edema (raw data) and % hemorrhaging (ranked data). Repeated measures analysis indicates a significant interaction between age and treatment effects for % lordosis. Subsequent means testing indicates that 10 µg/L resulted in significant effects on % lordosis for three-day old larvae using raw data. No statistically significant effects were observed on % lordosis at 10 µg/L for two-day old larvae (ranked data) despite such effects occurring at 2.5 µg/L (see above).

Study III Results

- No statistically significant effect of selenium was found on any of the three abnormality measures on larvae in streams recovering from prior selenium additions in Study II.

Adult Spawning Activity and Effects on Progeny (Egg Cup Data)
Study II Results

- ***Effects of 2.5 µg/L Selenium.*** ANOVA results on maximum % abnormalities from the Egg Cup Study II indicate that 2.5 µg/L selenium resulted in no statistically significant effects on larvae for any of the four abnormality measures (Table 5-2). Repeated measures analysis indicates a significant interaction between age and treatment effects for three of the four measures (% edema—raw data; % hemorrhaging—raw and ranked data; % healthy—raw and ranked data). Subsequent means testing indicates that 2.5 µg/L selenium resulted in statistically significant effects on % hemorrhaging and % healthy for one-day old larvae using the raw data. Analysis of the full and partial time-series data sets showed consistent results for age-dependent effects at 2.5 µg/L.

- ***Effects of 10 µg/L Selenium.*** ANOVA results on maximum % abnormalities from the Egg Cup Study II indicate that 10 µg/L selenium resulted in statistically significant effects on larvae for two of the four abnormality measures (maximum % edema—raw and ranked data; maximum % healthy—raw and ranked data, Table 5-2). Repeated measures analysis and subsequent means testing using the full time-series data indicate that 10 µg/L selenium had a statistically significant effect on % edema (two- through five-day old larvae—raw data), % hemorrhaging (one- through three-day old larvae—raw data; one- through four-day old larvae—ranked data), and % healthy (one- through five-day old larvae—raw and ranked data). For abnormality measures that did not show a significant age and treatment interaction, repeated measures analysis indicates that statistically significant effects occurred at 10 µg/L for % edema (ranked data) using both the full and partial time-series data and % lordosis (ranked data) using the full time series data. For days common to the two time series, analysis of the full and partial time-series data sets generally showed results with respect to age-dependent effects at 10 µg/L.

- No statistically significant effect of selenium was found on % hatch or % survival of larvae in egg cups in Study II.

Study III Results

- No statistically significant effect of selenium was found on any of the four % abnormality measures on larvae in streams recovering from prior selenium additions in Study III. Similarly, no statistically significant effect of selenium was found on % hatch or % survival of larvae in the recovering streams of Study III.

Table 5-1. Summary of Statistical Results from Study II Field Nest Data[a]

Statistical Method (Factor)	% edema		% lordosis		% hemorrhaging	
	Raw Data	Ranked Data	Raw Data	Ranked Data	Raw Data	Ranked Data
ANOVA[b] (Treatment)	**S, 10 µg/L** (p = 0.020)	**S, 10 µg/L** (p = 0.020)	NS	NS	NS (p = 0.057)	**S, 2.5 µg/L** (p = 0.022) **S, 10 µg/L** (p = 0.014)
Repeated-Measures (Treatment)	**S, 10 µg/L** (p = 0.027)	NS	-[c]	-[c]	NS	**S, 10 µg/L** (p = 0.041)
Repeated-Measures (Treatment × Age)	NS	NS	**S, 10 µg/L age 3** (p = 0.035)[d]	**S, 2.5 µg/L age 2** (p = 0.031)[e]	NS	NS

[a] S = statistically significant ($p \leq 0.05$). When statistically significant effects are reported, p-values are from Dunnett's means test unless otherwise noted.
NS = not statistically significant ($p > 0.05$. p-values were reported if $p \leq 0.10$).

[b] Unless otherwise noted, Tukey's and Dunnett's showed consistent overall indication of statistical significance.

[c] ANOVA was performed on maximum % abnormality per nest.

[d] Statistically significant p-value was found at $\alpha = 0.05$ for age × treatment interactions.

[e] This p-value is from Tukey's means test, Dunnett's test results are not reported.

Tukey's test of the means was marginally not significant (p = 0.069).

Table 5-2. Summary of Statistical Results from Study II Egg Cup Data[a]

Statistical Method (Factor)	% edema		% lordosis		% hemorrhaging		% healthy	
	Raw Data	Ranked Data	Raw Data	Ranked Data	Raw Data	Ranked Data	Raw Data	Ranked Data
ANOVA[b] (Treatment)	**S (10)** (p = 0.001)	**S (10)** (p = 0.030)	NS* (p = 0.060)	NS* (p = 0.055)	NS* (p = 0.084)	NS* (p = 0.065)	**S (10)** (p = 0.010)	**S (10)** (p = 0.040)
Repeated-Measures[c] (Full, Treatment)	- [f]	**S (10)** (p = 0.032)	NS	**S (10)** (p = 0.032)	- [f]	- [f]	- [f]	- [f]
Repeated-Measures[c] (Full, Treat. × Day)	**S (10, Day 2)** (p = 0.008) **S (10, Days 3, 4, 5)** (p = 0.000[e])	NS* (p = 0.080)	NS	NS* (p = 0.086)	**S (2.5, Day 1)** (p = 0.005) **S (10, Day 1)** (p = 0.001) **S (10, Day 2)** (p = 0.0004[e]) **S (10, Day 3)** (p = 0.000[e])	**S (10, Day 1)** (p = 0.010) **S (10, Days 2 & 3)** (p = 0.000[e]) **S (10, Day 4)** (p = 0.030[e])	**S (2.5, Day 1)** (p = 0.004) **S (10, Day 1)** (p = 0.000) **S (10, Days 2, 3, 4, 5)** (p = 0.000[e])	**S (10, Day 1)** (p = 0.000) **S (10, Days 2, 3, 4)** (p = 0.000[e]) **S (10, Day 5)** (p = 0.006[e])
Repeated-Measures[d] (Partial, Treatment)	- [f]	**S (10)** (p = 0.017)	NS* (p = 0.062)	NS	- [f]	- [f]	- [f]	- [f]
Repeated-Measures[d] (Partial, Treat. × Day)	**S (10, Day 2)** (p = 0.024) **S (10, Day 3)** (p = 0.0001[e])	NS	NS	NS	**S (2.5, Day 1)** (p = 0.016) **S (10, Day 1)** (p = 0.008) **S (10, Day 2)** (p = 0.004[e]) **S (10, Day 3)** (p = 0.0001[e])	**S (10, Day 1)** (p = 0.0122) **S (10, Days 2, 3)** (p = 0.000[e])	**S (2.5, Day 1)** (p = 0.039) **S (10, Day 1)** (p = 0.009) **S (10, Days 2, 3)** (p = 0.000[e])	**S (10, Day 1)** (p = 0.0002) **S (10, Days 2, 3)** (p = 0.000[e])

[a] S = statistically significant ($p \leq 0.05$). p-values are from Dunnett's means test unless otherwise noted.
NS = not statistically significant ($p > 0.10$).
NS* indicates $0.05 < p \leq 0.10$.
Numbers in parentheses (e.g., 10) indicate treatments where statistically significant difference was observed relative to control.
Unless otherwise noted, Tukey's and Dunnett's showed consistent overall indication of statistical significance.

[b] ANOVA was performed on maximum % abnormality and minimum % healthy per cup.

[c] Repeated-measures analysis on "full" time-series (Day 1 data were excluded for % edema and % lordosis).

[d] Repeated-measures analysis on "partial" time-series (Days 1, 4, and 5 data were excluded for % edema and % lordosis; Days 4 and 5 were excluded for % hemorrhaging and % healthy).

[e] This p-value is from Tukey's test, Dunnett's tests were not reported. All the ages indicated in parentheses have this same p-value.

[f] Statistically significant p-value was found at $\alpha = 0.05$ for treatment × day interactions.