APPENDIX D

Maximum likelihood estimation using PESC and Elphick et al. (2011) data conducted by Dr. Carl Schwarz.

A statistical examination of the effect of water hardness

on the dose-response of fresh water aquatic species to sulphate.

Carl James Schwarz

Department of Statistics and Actuarial Science

Simon Fraser University

Burnaby, BC V5A 1S6

cschwarz@stat.sfu.ca

1. Introduction

The British Columbia Ministry of Environment (MOE) received the results from two separate studies on the toxicological effects of sulphate at varying water hardness values using various freshwater species of aquatic organisms. These studies were conducted by Nautilis Environmental and Environment Canada.

In the Environment Canada study, three water hardness values were examined using various freshwater species of aquatic organisms. The tests were done at a low water hardness (50 mg/L), a medium water hardness (100 mg/L), and a high water hardness (250 mg/L). Details of the experimental prot0col are found in Meays et al. (submitted). We use the data from Meays et al. (submitted) to assess if there is evidence of an effect of water hardness on the dose-response relationship between sulphate and the various endpoints measured.

Raw data were provided as an Excel workbook, the raw output sheets (in pdf format) from the analyses done by Meays et al. (submitted) using the CETIS software, and additional pdf file from the

Saskatchewan Research Council who performed some of the work under sub-contract from Environment Canada on some species.

In the Nautilus Environmental study, there were between one and four hardness levels (ranging from 15 to 320 mg/L) and a variety of freshwater species of aquatic organisms. We use the data from Elphick et al. (2010) to also assess if there is evidence of an effect of water hardness on the dose-response relationship between sulphate and the various endpoints measured. Only those organisms where at least two levels of water hardness were studies are used in this paper.

The raw data was extracted from copies of the raw output sheets (in pdf format) from the analyzes done by Elphick et al. (2010) using the CETIS software. Only the organisms where at least two water hardness levels were tested were used.

It is assumed that all the data presented are valid and no examination of the raw data for outliers or other anomalous points has been done.

In the rest of the paper, the prefix EC and NA will refer to the Environment Canada and Nautilus studies respectively.

2. Methods

The sampling protocol for each aquatic organism is presented in detail in Meays et al (submitted) and Elphick (2010). A brief summary is presented in Table 1. All tests were performed at various levels of hardness of water and usually five or six nominal concentrations of sulphate. In the Environment Canada studies, the actual sulphate concentration was measured at the start and the end of the experiment and the average of the two values was used as the actual sulphate concentration. In the Nautilus studies, the nominal sulphate levels as recorded on the CETIS sheets were used directly. In Elphick et al. (2010, Table 2), a comparison of the measured vs. nominal sulphate levels showed a relatively good agreement. Most experiments also had a control (nominal zero concentration) of sulphate.

There are two general types of responses in this series of experiments.

2.1 Probit Models:

First, are the mortality responses. For these experiments, Probit models (Bliss, 1934) will be used. The basic Probit model assumes that the number of deaths follows a binomial distribution where the probability of mortality is "linked" to a linear function through the normal distribution. For example, consider the Probit model for a fixed hardness level – the statistical model is:

$$Dead_{ij} \sim Binomial(BatchSize_{ij}, p_i)$$
$$p_i = \Phi\left(\beta_0 + \beta_i \log\left(D_{ij}\right)\right)$$

where $Dead_{ij}$ is the number of dead organisms observed in the j^{th} batch out of the initial $BatchSize_{ij}$ units on tests at dose level (sulphate) D_i; β_0, β_1 are the intercept and slope in the Probit model; and Φ is the cumulative normal distribution. [The original papers on Probit analysis added 5 to the linear functions to avoid negative numbers in hand computations, but this is no longer required when using computers.] The parameters are estimated using maximum likelihood (e.g. via Proc Probit in SAS). Estimates of the LCxx values (i.e. at what concentration will a fraction xx or organism die) can be found once estimates of the slope and intercept are found by solving the equation

$$LCxx/100 = \Phi\left(\hat{\beta}_0 + \hat{\beta}_i \log\left(D_{ij}\right)\right)$$

Maximum likelihood estimates are asymptotically the best possible estimates and extract the maximum amount of information from the data. Estimates of precision (i.e. standard errors) can be found automatically for the parameters of the likelihood equations and by the delta method (Taylor series expansion) for the LCxx values.

The formulation above assumes that the probability of death will decline to zero as the sulphate dose declines to 0. Probit models have been developed to deal with non-zero natural responses. In the original papers, the observed mortality at control doses was treated as a fixed known natural response and the probit analysis applied only to mortalities above this level. This approach ignored the uncertainty in the estimate and the resulting estimates and standard errors from the remainder of the fit did not account for this. A more modern approach is to let the natural response rate be another parameter to be estimated in the model along with the slope and intercept of the Probit function. Again consider the Probit model for a fixed hardness level – the statistical model is:

$$Dead_{ij} \sim Binomial(BatchSize_{ij}, p_i)$$
$$p_i = NR + (1 - NR)\Phi\left(\beta_0 + \beta_i \log\left(D_{ij}\right)\right)$$

where NR is the natural response (mortality) at no (the control batches) sulphate, i.e. the fraction of units expected to die in the absence of an effect of sulphate. The parameters are again estimated using maximum likelihood (e.g. Proc Probit in SAS).

Note that in models with a very small dose-response effect, there is some ambiguity in the parameterization. This is because it is very hard then to distinguish between a natural response, or a model with a slope close to 0 as both will give similar fits to the data. In cases like this, it may be better drop the natural response terms.

Because of the natural response, estimation of the LCxx values must be done with care. For example, the LC25 values refer to the dose that results in a 25% mortality of the organism that *survive the natural response*. Suppose that the estimated natural response is 13%. Consequently, only 87% of the organisms would survive in the absence of sulphates. The LC25 refers to the additional 25% of 87%=22% mortality above the natural response for a total mortality of 12% + 22% = 35%. The estimated LC25 value is found by now solving:

$$.35 = .13 + .87\Phi\left(\hat{\beta}_0 + \hat{\beta}_1 \log(D)\right)$$

which again leads to

$$.25 = \Phi\left(\hat{\beta}_0 + \hat{\beta}_1 \log(D)\right)$$

i.e. you cannot simply read on a plot of mortality by dose at the .25 value on the *Y* axis.

A goodness-of-fit statistic of the Probit model (both with and without a natural response) to the data is found by comparing the observed and expected counts:

$$X^2 = \sum \frac{\left(Dead_{ij} - BatchSize_{ij}\hat{p}_{ij}\right)^2}{BatchSize_{ij}\hat{p}_{ij}} + \sum \frac{\left(Alive_{ij} - BatchSize_{ij}\left(1-\hat{p}_{ij}\right)\right)^2}{BatchSize_{ij}\left(1-\hat{p}_{ij}\right)}$$

where $\hat{p}_{ij}$ is the predicted probability of death for each batch. If the assumptions of the model are satisfied, this statistics should follow a χ^2_{df} distribution where the *df* is found appropriately. If the X^2 statistic is extreme, it indicates a lack-of-fit. There are two common reasons for lack-of-fit. First, the model itself can be wrong (e.g. the response is not linear on the Probit scale), or the structural model is valid (i.e. the response is linear on the Probit scale), but the data are more variable than expected from a binomial response. The latter is termed overdispersion. For example, consider the sample proportion of organisms that die in batches of 30 organisms where the underlying mortality rate is 30%. Statistical theory indicates that under the binomial model, the average number that would die would be 9 = 30(.3), but the actual number that could die would range from 4 to 14. If the observed number that dies ranged from 1 to 17, this would indicate overdispersion, even though the average number that dies is still be 9. Typically causes of overdispersion are non-independence in the fate of the organism. For example, if all

the organisims are placed in the same test tube, a local contaminant could reduce/increased the survival rate of this batch from the projected 30%.

The consequence of overdispersion is that estimates remain unbiased, but the reported standard errors (and p-values derived from them) are understated, i.e. the results appear to be more precise than they really are.

There are two methods for correcting for overdispersion. In the quasi-likelihood approach, an overdispersion correction factor is found as the ratio of the goodness-of-fit statistic to the df

$$\hat{c} = \frac{\mathrm{X}^2}{df}.$$

Quasi-likelihood theory indicates that the estimates will be unchanged under overdispersion, but the standard errors of all estimates should be inflated by a factor of $\sqrt{\hat{c}}$ to account for the extra variation in the data. This is the approach taken by CETIS.

More modern approaches incorporate the overdispersion directly in the model through the random effect probit models (Gibbons et al, 1994; Gibbons and Hedeker, 1994). In the random effect model, latent (unobserved) random noise is added to the probit function:

$$\begin{aligned} Dead_{ij} &\sim Binomial(BatchSize_{ij}, p_i) \\ p_i &= NR + (1-NR)\Phi\left(\beta_0 + \beta_i \log\left(D_{ij}\right) + \varepsilon_{ij}\right) \\ \varepsilon_{ij} &\sim N\left(0, \sigma^2\right) \end{aligned}$$

for non-control doses of sulphate, and

$$\begin{aligned} Dead_{ij} &\sim Binomial(BatchSize_{ij}, p_i) \\ p_i &= \Phi\left(\Phi^{-1}\left(NR\right) + \varepsilon_{ij}\right) \\ \varepsilon_{ij} &\sim N\left(0, \sigma^2\right) \end{aligned}$$

for control doses of sulphate, where ε_{ij} is a latent random effect that comes from a normal distribution with mean 0 and variance σ^2, i.e. adding extra variation in the mortality rate at a specified dose. So even if the expected mortality rate at a particular dose is 30%, the random effect (applied at the batch level) could vary this higher or lower. This model can also be fit using maximum likelihood (e.g. Proc Nlmixed in SAS). Estimates from the fitted model automatically incorporate the effects of the excess random variation in their standard errors.

The primary goal of this paper is to investigate the effect of hardness levels on the dose-response curve. We accomplish this by fitting two (or more) models to the combined data from the three hardness levels. In first model (the **Separate response** model), a separate probit curve is fit to each hardness level. So, if the basic probit model is used with 3 hardness levels, this model will require 6 parameters (an intercept and a slope for each hardness level). This can be done in a single model fit rather than (the equivalent) running three separate models (one for each dose). In the second model (the **Common response**) model, the data are pooled over all hardness levels and single probit model is fit. This model has 2 parameters. In some cases, additional models were run where individual probit curves were fit (one for each hardness level), but the curve were constrained to have a common LC10, LC25, or LC50 values (e.g. refer to Jeske et al., 2009). For these models, penalized maximum likelihood was used where the penalty is computed as the difference among the LCxx values from the individual curves which declines to zero as the condition is satisfied.

2.2 Isotonic regression models:

There is no common model suitable for modeling weight, reproduction, frond number, or other non-binomials endpoints. The CETIS software has a wide suite of potential models (e.g. the Gompertz) but in the majority of the cases here, the CETIS software uses a linear interpolation method (ICPIN). This is also known as isotonic regression (Barlow et al, 1972). The basic premise is that the response variable should decline with increasing sulphate levels. However, because of sampling fluctuation, the observed curve may not show the monotonic decline.

We also used isotonic regression (Barlow et al., 1972) to fit models to this data. Basically, isotonic regression works from left to right through the data. If the mean response at the next *X* value is higher than the current fitted *Y* value, then the previous data and the new *Y* are pooled, a new mean is computed, and algorithm moves to the next *X* value. This is a "non-parametric" method, but can be shown to be the maximum likelihood approach under monotonicity. The *R* function *isoreg()* can be used to fit these models. The likelihood, assuming that the distribution of data values is normally distributed at a particular dose level, can be found from a transformation of the sum-of-squares of the residuals from the fit.

Estimates of the ICxx values are found by linear interpolation on the log(dose) scale. Standard errors (and confidence limits) are found using a bootstrap method. Several hundred bootstrap samples were generated with replacement from the observed data. For each bootstrap sample, the isotonic regression model was fit and the estimate of the ICxx value determined. The 2.5^{th} and 97.5^{th} percentile of the

bootstrap estimates were used as the 95% confidence intervals for the parameter. Note that it is impossible to estimate any ICxx value that exceeds the largest dose observed in the experiment and in these cases, no estimate is reported. Similarly, in some cases, the isotonic regression line is completely flat and no estimate of the ICxx values can be computed.

This method can also be used with mortality data if there is evidence of a structural lack of fit in the probit model (i.e. the response is not linear on the probit scale). The isotonic model treats a natural response as simply another set of data values. In these cases, the LCxx values from isotonic regression are not directly comparable to those from the maximum likelihood probit approach with a natural response. In the isotonic method, no natural response is assumed and so the LCxx value includes the natural response in total mortality. Consequently, a simple comparison of LCxx values based on the CETIS output sheets should be done carefully as they may not be comparable.

2.3 Model selection:

We use the information theoretic paradigm for model selection using the Akaike Information Criterion (AIC; Anderson, 2008; Burnham and Anderson, 2002). In this paradigm, one recognizes that all models are but approximations to reality. For example, while the common response model fits a single probit curve for all hardness levels, it is not theoretically justified that the dose-response curves are really identical among the three hardness levels. Rather, the data may not be sufficiently strong (e.g. too few data points; too much natural variation) to distinguish between the curve. So rather than being forced to choose among the common or separate model, in the AIC paradigm, the relative support among the models is computed. This can be used to rank the models in terms of their support.

The Akaike Information Criterion corrected for small sample sizes (AICc, Anderson, 2008) was used to rank the models. The AICc measures the tradeoff between model complexity (measured by the number of parameters) and model fit (measured by the likelihood value). More complex models (i.e. more parameters, such as the separate response curve model) will always fit the data better (have a higher likelihood value), but the improvement in fit may be overshadowed by the increase in complexity (number of parameters). Smaller values of AICc indicate a model that is better in the fit/complexity tradeoff. The actual value of the AICc is not interpretable, but given an *a priori* model set, the difference in AICc between the best fitting model and the other models is a measure of how similar two models are in the fit-complexity tradeoff. Differences in AICc of more than about 4 or 5 indicate that the model with the larger AICc is vastly inferior to the model with the smaller AICc. The differences in AICc can be converted into model weights, which measure the relative support of each model in the model set. Cases where the data are sparse or highly variable would result in the competing models having roughly

equal weight. Cases where the data are rich, or the differences extreme in the fit, would result in some models having the majority of the fit.

A common misconception is that AICc indicates the "correct" model. This is not true – AICc only ranks models in the fit/complexity tradeoff given the data at hand. With more data, more complex models can be justified and the model ranking would shift towards more complex models. With sparse data, AICc will tend to favor simpler models as an adequate description of the data. AICc only looks at the models in the set – it could be possible that the best model in the set still does not adequately fit the data. Consequently, it is still important to assess model fit via visual inspection of the fit or goodness-of-fit statistics before interpreting the results of the model ranking.

3. Results.

For each species tested, an initial plot of the raw data is presented. This will often indicate which modeling approach is required (e.g. the basic probit, the probit with a natural response, the probit model with overdispersion, isotonic regression). Then at least two models are fit (the common and separate models) and ranked using AICc. Fits of the two models are presented to visually assess if there is structural lack of fit. Estimates under the two models of the parameters of the model are presented.

3.1 EC Rainbow Trout Eggs

A plot of the raw data is shown in Figure EC-RT-1. The plot indicates the presence of a non-zero natural response. It appears that even at low concentrations of sulfate, there is a non-zero mortality rate. As well, there is strong evidence of overdispersion. For example, based on batches of 30 eggs and a (true) average mortality rate of 0.40, the observed mortality rates would tend to vary between 0.30 and 0.50, but the observed range in Figure EC-RT-1 is much wider. Consequently, the probit model with random effects will be used.

Six models were fit to assess the effect of water hardness and sulphate levels on mortality (Table RT-1). The most general model had a separate dose-response curve for each hardness level while the most simple model fit a single dose response curve to all hardness levels. Intermediate models forced the individual curves to match at the LC10, LC25, or LC50 points; or had a common natural response value with separate dose-response curves. The fitted values from selected models are shown in Figures RT-2 to RT-4. In models where the LCxx values are forced to be equal, the actual mortality at the common LCxx dose is not equal because of potentially differing natural response values.

Note that the estimates reported by CETIS do not correspond directly to the estimates for the separate curves reported in Table RT-1 for several reasons. First, even though the models are the "same" (i.e. a separate curve for each hardness level), the maximum likelihood estimate uses all of the hardness level data to estimate the common overdispersion parameter while CETIS estimates an overdispersion separately for each hardness level. The CETIS program does not model overdispersion directly – rather it fits a model that assumes no overdispersion and then adjusts estimates based on the goodness-of-fit. While asymptotically (i.e. in large samples) equivalent to maximum likelihood, the results may differ in small samples. Because CETIS does not treat overdispersion directly in the model, "outliers" in the control doses have an inordinate effect on the estimate of the natural response. For example, in the low hardness testing, the three control replicates has natural mortality rates of 13%, 13%, and 53% for an average of 28%. CETIS estimated the natural response for the low hardness set at 35% while the maximum likelihood approach estimated the natural response at only 14%. This has implication then in estimating the LCxx values as the non-natural response curve estimated by CETIS is not as steep as in the maximum likelihood approaches.

According to the AICc criteria, the model with a single dose-response curve for all hardness values has overwhelming support (model weight of .84) among the models fit. There is little evidence for an effect of hardness on dose-response curve for sulphates.

Note that there is still some evidence of lack-of-fit for the best probit model. For example, refer to Figure EC-RT-2. Notice that at the lowest hardness level, the three replicates tend to be all either above or below the fitted probit line. This indicates that some other random effect that operates on the set of batches may be present. For example, the three batches corresponding to a single sulphate dose level were all prepared together (rather than randomizing the preparation) and a local containment influenced all three batches.

3.2 EC Chinook Eggs

A plot of the raw data is shown in Figure EC-CH-1. The mortality rate is very low and relatively constant over doses of sulphate for all but the hardest water. This will make it virtually impossible to disentangle the natural response rate and a dose response curve with a slope equal to zero. Consequently, the basic probit models with no natural response will be used. The raw data seems to indicate that observed mortality decreases with sulphate dose under medium hardness and so the probit slope will be constrained to be positive. Lastly, in the low and medium hardness levels, mortality never approached a 10% overall mortality, so extrapolating to LC10, LC25, and LC50 should not be done. There was no evidence of overdispersion, and none was modeled.

Because of the lack of observable dose-response for the low and medium hardness tests, only two models were fit to assess the effect of water hardness and sulphate levels on mortality (Table CH-1). The fitted values from selected models are shown in Figures CH-2 to CH-3. The most general model had a separate dose-response curve for each hardness level while the most simple model fit a single dose response curve to all hardness levels.

The AICc model selection criteria gave strong evidence that the dose-response curves are not equal across all hardness levels – this is evident from the initial plot of the raw data where the dose-response curve showed evident curvature for the highest hardness level.

The estimated slopes for the low and medium hardness levels were either constrained to be zero or just above zero indicating little evidence of a dose-response curve for these two hardness levels. Consequently extrapolation of the curve to estimate the LCxx values was not done. Extrapolation to estimate the LCxx values for the highest hardness is also not advisable except for LC10 as the other values are so far outside the range of observed mortalities.

3.3 EC Hyalella

3.3.1 Mortality:

A plot of the raw data is shown in Figure EC-HY-Mort-1. There was no evidence of overdispersion and none was modeled. There was no evidence of a dose-response relationship for tests done at the medium level of hardness

Two models were fit to the data. In the first model, separate probit relationships with a natural response were fit for the low and high hardness levels; only a constant natural mortality could be fit for the medium hardness data as no dose-response relationship was found. In the second model, a common probit model was fit. No models were fit where the LCxx values were in common as these are typically so far outside the range of observed suphate levels that the extrapolation is meaningless.

Results are presented in Table HY-Mort-1, and summary plots in Figures HY-Mort-1 and HY-Mort-2. The AICc model selection criteria provided strong evidence that the dose-response relationship is not homogeneous across hardness levels. Extrapolation to LC25 and LC50 values are far outside the range of experimental data that they were not estimated.

3.3.2 Growth:

A plot of the raw data is shown in Figure EC-HY-Weight-1. The dose-response curve does not have a standard shape. Indeed CETIS fit simple isotonic (i.e. must not increase over time) curves to the medium and hard water tests, and a Gompertz curve for the low water hardness tests. We used isotonic regression for all hardness levels.

Two models were fit to the data. In the first model, a separate isotonic relationship was fit for each hardness level. In the second model, the data were pooled over all hardness levels. Results are presented in Table HY-weight-1 and a summary plot in Figure EC-HY-weight-1.

Estimates of the ICxx values indicate that they are not comparable. It is difficult to compute the IC50 value because it is often outside the range of doses used in the experiments.

The AICc model selection criteria provided strong evidence in favor of the model where the dose-response curve is homogeneous across hardness levels.

3.4 EC Freshwater Mussels

A plot of the raw data is found in Figure EC-MY-1. The small sample sizes in each batch (3 or 4) make the reported mortality rates very discrete with only a few possible values. These small sample sizes will make it impossible to estimate any natural response and so only probit models with no natural response were fit.

Five models were fit to the data. The most general model had a separate probit line for each hardness level; the simplest model had a single probit response over all hardness levels; intermediate models constrained the LC10, LC25, or LC50 points to be equal among the three individual curves. Results are presented in Table MY-1, and summary plots in Figures MY-2 to Figure MY-4. Results for models where the LC25 and LC50 were constrained to be equal are not shown because these points are so far outside the range of observed doses making the extrapolation dubious.

The AICc model selection criteria provided strong evidence that the dose-response relationship is not homogeneous across hardness levels. Extrapolation to LC25 and LC50 values are far outside the range of experimental data that they were not estimated.

3.5 EC Tadpole

No analysis was attempted. In low hardness trial, only 1/90 tadpoles died; in medium hardness trials only 6/90 tadpoles died; in high hardness trials only 2/90 tadpoles died. No modeling of the dose-response relationship for mortality was done because of the very low mortality rates observed.

The change in weight over the 28 days of the trials was also recorded. No analysis was done because the mean weight increased with increasing doses of sulphate rather than decreased as expected. An isotonic regression in this case would not be sensible.

3.6 EC Fat Head Minnow

3.6.1 Mortality:

A plot of the raw data is shown in Figure EC-FM-Mort-1. There appears to be little or no natural response in the control groups. There was no evidence of overdispersion. Consequently, probit models with no natural response were fit.

Five models were fit to the data. The most general model had a separate probit line for each hardness level; the simplest model had a single probit response over all hardness levels; intermediate models constrained the LC10, LC25, or LC50 points to be equal among the three individual curves. Results are presented in Table FM-Mort-1, and summary plots in Figures FM-Mort-2 to FM-Mort-4. Results for models where the LC25 and LC50 were constrained to be equal are not shown because these points are so far outside the range of observed doses making the extrapolation dubious.

The AICc model selection criteria provided strong evidence that the dose-response relationship is not homogeneous across hardness levels. Extrapolation to LC25 and LC50 values are far outside the range of experimental data that they were not estimated.

3.6.2 Weight.

A plot of the raw data is shown in Figure EC-FM-weight-1 and the fitted isotonic curves in Figure EC-FM-weight-2. Results of the model fitting are presented in Table FM-weight-1.

The AICc model selection criteria provided strong evidence that the dose-response relationship is not homogeneous across hardness levels. Extrapolation to LC25 and LC50 values for the highest hardness level is not possible as the isotonic curve is flat.

3.7 EC Lemna

3.7.1 Final Weight.

A plot of the raw data is shown in Figure EC-LM-Weight-1. There appears to be little evidence of an inhibitory effect. We also used isotonic regression (Barlow et al., 1972) to fit models to this data. Two models were fit to the data. In the first model, a separate isotonic relationship was fit for each hardness level. In the second model, the data were pooled over all hardness levels. Results are presented in Table LM-weight-1 and a summary plot in Figure EC-LM-weight-2.

Computation of the ICxx values could not be done because the fitted isotonic curves are flat for each hardness, or the estimated ICxx value is beyond the largest observed dose for the combined data.

The AICc model selection criteria provided strong evidence in favor of the model where the dose-response curve is homogeneous across hardness levels.

3.7.2 Frond increase.

A plot of the raw data is shown in Figure EC-LM-Frond-1. There appears to be little evidence of an inhibitory effect. We also used isotonic regression (Barlow et al., 1972) to fit models to this data. Two models were fit to the data. In the first model, a separate isotonic relationship was fit for each hardness level. In the second model, the data were pooled over all hardness levels. Results are presented in Table LM-Frond-1 and a summary plot in Figure EC-LM-Frond-2. Computation of the ICxx values was usually not possible as it often exceeded the largest observed dose in the experiment.

The AICc model selection criteria provided strong evidence in favor of the model where the dose-response curve is homogeneous across hardness levels.

3.8 NA Daphnia

3.8.1 Mortality.

A plot of the raw data is shown in Figure NA-DA-Mort-1. There appears to be little or no natural response in the control groups. There was no evidence of overdispersion. Consequently, probit models with no natural response were fit.

Five models were fit to the data. The most general model had a separate probit line for each hardness level; the simplest model had a single probit response over all hardness levels; intermediate models constrained the LC10, LC25, or LC50 points to be equal among the three individual curves. Results are presented in Table MY-Mort-1, and summary plots in Figures MY-Mort-2 and MY-Mort-3.

The AICc model selection criteria provided strong evidence that the dose-response relationship is not homogeneous across hardness levels.

3.8.2 Reproduction.

A plot of the raw data is shown in Figure NA-DA-Repro-1. The inhibitory effect of sulphate appears to be ameliorated at higher hardness levels. We also used isotonic regression (Barlow et al., 1972) to fit models to this data. Two models were fit to the data. In the first model, a separate isotonic relationship was fit for each hardness level. In the second model, the data were pooled over all hardness levels. Results are presented in Table NA-DA-Repro-1 and a summary plot in Figure NA-DA-Repro-2. The estimated ICxx values indicate that these appear to differ across the various hardness levels in the experiment.

The AICc model selection criteria provided strong evidence in favor of the model where the dose-response curve is different across hardness levels.

3.9 NA Rotifer

A plot of the raw data is shown in Figure NA-RO-Repro-1. The inhibitory effect of sulphate appears to differ across hardness levels. We also used isotonic regression (Barlow et al., 1972) to fit models to this data. Two models were fit to the data. In the first model, a separate isotonic relationship was fit for each hardness level. In the second model, the data were pooled over all hardness levels. Results are

presented in Table NA-RO-Repro-1 and a summary plot in Figure NA-RO-Repro-2. The estimated ICxx values indicate that these appear to differ across the various hardness levels in the experiment.

The AICc model selection criteria provided strong evidence in favor of the model where the dose-response curve is different across hardness levels.

3.10 NA Fat Head Minnow

3.10.1 Mortality:

A plot of the raw data is shown in Figure NA-FM-Mort-1. There appears to be little or no natural response in the control groups. There was no evidence of overdispersion. Consequently, probit models with no natural response were fit.

Five models were fit to the data. The most general model had a separate probit line for each hardness level; the simplest model had a single probit response over all hardness levels; intermediate models constrained the LC10, LC25, or LC50 points to be equal among the three individual curves. Results are presented in Table NA-FM-Mort-1, and summary plots in Figures NA-FM-Mort-2 and NA-FM-Mort-3.

The AICc model selection criteria provided strong evidence that the dose-response relationship is not homogeneous across hardness levels.

3.10.2 Weight.

A plot of the raw data is shown in Figure NA-FM-Weight-1. The inhibitory effect of sulphate appears to differ across hardness levels especially at higher hardness levels. We used isotonic regression (Barlow et al., 1972) to fit models to this data. Two models were fit to the data. In the first model, a separate isotonic relationship was fit for each hardness level. In the second model, the data were pooled over all hardness levels. Results are presented in Table NA-FM-Weight-1 and a summary plot in Figure NA-FM-Weight-2. The estimates of the IC values indicate that the corresponding population values do not appear to be common across hardness levels.

The AICc model selection criteria provided strong evidence in favor of the model where the dose-response curve is different across hardness levels.

3.11 NA Tadpoles

3.11.1 Mortality:

A plot of the raw data is shown in Figure NA-TA-Mort-1. There were only two hardness levels tested in this experiment, both at lower levels of hardness. There appears to be little or no natural response in the control groups. There was no evidence of overdispersion. Consequently, probit models with no natural response were fit.

Five models were fit to the data. The most general model had a separate probit line for each hardness level; the simplest model had a single probit response over all hardness levels; intermediate models constrained the LC10, LC25, or LC50 points to be equal among the three individual curves. Results are presented in Table NA-TA-Mort-1, and summary plots in Figures NA-TA-Mort-2 and NA-TA-Mort-3.

The AICc model selection criteria provided strong evidence that the dose-response relationship is homogeneous across hardness levels. This is not surprising as both tests were at low levels of hardness.

3.11.2 Weight.

A plot of the raw data is shown in Figure NA-TA-Weight-1. The inhibitory effect of sulphate appears to differ across hardness levels. We used isotonic regression (Barlow et al., 1972) to fit models to this data. Two models were fit to the data. In the first model, a separate isotonic relationship was fit for each hardness level. In the second model, the data were pooled over all hardness levels. Results are presented in Table NA-TA-Weight-1 and a summary plot in Figure NA-TA-Weight-2. The estimates of the IC values indicate that the corresponding population values do not appear to be common across hardness levels.

The AICc model selection criteria provided strong evidence in favor of the model where the dose-response curve is different across hardness levels.

3.12 NA Algae

A plot of the raw data is shown in Figure NA-Al-Repro-1. The inhibitory effect of sulphate appears appears to be roughly the same except for the unexpected non-inhibitory effect at intermediate levels of sulphate at the highest hardness level. We used isotonic regression (Barlow et al., 1972) to fit models to this data. Two models were fit to the data. In the first model, a separate isotonic relationship was fit for each hardness level. In the second model, the data were pooled over all hardness levels. Results are presented in Table NA-AL-Weight-1 and a summary plot in Figure NA-TA-Weight-2. The estimates of the ICxx values do not strongly indicate differences in the corresponding population values as the confidence intervals often overlap.

The AICc model selection criteria provided moderate evidence in favor of the model where the dose-response curve is homogeneous across hardness levels.

4. Summary

The results of the modeling exercise are decidedly mixed as shown below! Support for a common model across hardness levels varies from 0.0 to 1.0 and can vary also within the same aquatic species depending on the endpoint. Models with a common LC10, LC25, or LC50 either could not be fit (because the common endpoint was well beyond the range of the observe data) or had little support (because the individual hardness curves are so disparate).

Aquatic species	Model weight for common response across hardness levels
EC Rainbow Trout egg mortality	0.84
EC Chinook egg mortality	0.00
EC Hyalella mortality	0.00
EC Hyalella weight	0.06
EC Mussels mortality	0.28[#]
EC Bullfrog tadpoles	Insufficient data
EC Fat head minnows mortality	0.00
EC Fat head minnows weight	0.00
EC Lemna weight	0.86
EC Lemna frond increase	1.00
NA Daphnia mortality	0.00
NA Daphnia reproduction	0.00
NA Rotifer reproduction	0.05
NA Fat head minnow mortality	0.00
NA Fat head minnow weight	0.00
NA Tadpole mortality	0.34
NA Tadpole weight	0.02
NA Algae reproduction	0.76

[#] Model with common LC10 had weight 0.23.

In many cases, the inability to distinguish between dose-response curves for different hardness levels may be due to low power because of insufficient numbers tested or high variability in the responses. If it

is important to detect (and adjust) for differing sulphate levels, a power analysis should be performed to indicate what biological differences are detectable with a specified sample sizes before the experiment is conducted to avoid disappointment.

Estimates of the ECxx/ICxx are presented in Table-Summary along with 95% confidence intervals for the parameter.. These can differ (often considerably) from the results in the two reports for a variety of reasons. First, all mortality studies in this report used a probit model and all non-mortality endpoints were modeled suing isotonic regression, but the software used in the other two reports selected a wide range of models often with no justification. In cases where overdispersion was present, the models in this report used a random effects model while the other two reports used quasi-likelihood corrections.

In cases where there is strong evidence that the dose-response curve varies both as a function of hardness and sulphate, a omnibus model where both sulphate and hardness are present in the linear term of the probit model may be a more suitable way to predict the LCxx values for various combinations of sulphate and water hardness.

References

Anderson, D. (2008) Model Based Inference in the Life Sciences: A Primer on Evidence. Springer: New York.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. (1972) Statistical inference under order restrictions; Wiley, London.

Bliss C.I. (1934). The method of probits. Science 79 (2037): 38–39.

Burnham K.P. & Anderson D.R. (2002) Model selection and multi-model inference: a practical information-theoretic approach. Springer, New York.

Elphick, J. Davies, M., Gilron, G., Canaria, E.C., Lo, B. and Bailey, H. C. (2010). An aquatic toxicological evaluation of sulphate: the case for considering hardness as a modifying factor in setting water quality guidelines. Environmental Toxicology and Chemistry. Online early. DOI: 10.1002/etc.363.

Gibbons. R. D. and Hedeker. D. (1994). Application of random effects probit regression models. Journal of Clincal and Consulting Psychology 62, 285-296.

Gibbons. R. D. Hedeker. D.; Charles. S. C. and Frisch. P. (1994). A random effects probit model for predicting medical malpractice claims. Journal of the American Statistical Association 89, 760-767.

Gilmour. A. R.; Anderson. R. D. and Rae, A. L. (1985). The analysis of binomial data by generalized linear mixed model. Biometrtka 72, 593-599.

Jeske, D.R. Xu, H.K., Blessinger, T., Jensen, P. Trumble, J. (2009) testing for the Equality of EC50 Values in the Presence of Unequal Slopes With Application to Toxicity of Selenium Types. Journal of Agricultural, Biological and Environmental Statistics 14, 469-483.

Meays, C. et al. (submitted). Toxicological assessment of sulfate at three water hardness values using various freshwater aquatic organisms. Submitted to Environmental Management.

Organisms

Table 1. Summary of sampling protocols for the experiments conducted.		
Environment Canada Studies		
Aquatic species	Response	Sampling protocol at each combination of water hardness and sulphate levels
Rainbow Trout	Survival of eggs to 21 days	Triplicate batches of 30 eggs were incubated and the number of mortalities from each batch was recorded.
Chinook	Survival of eggs to 28 days.	Triplicate batches of 30 eggs were incubated and the number of mortalities from each batch was recorded.
Hyalella	Survival and growth of organisms to 28 days.	Quintuplicate batches (except for 10 batches in the case of control doses of sulphate in soft water) of 15 Hyalella were incubated and the number of mortalities from each batch was recorded. The mean weight of each batch of the organisms at the end of the experiment was measured.
Mussels	Survival and growth of organisms to 28 days.	Triplicate batches of 3, 3, or 4 mussels were incubated and the number of mortalities in each batch was recorded. Wet weight and the beginning and end of the experiment was measured.
Bullfrog tadpoles	Survival and growth to 28 days.	Triplicate batches of 5 tadpoles were incubated and the number of mortalities in each batch was recorded. The change in weight over the 28 days was also recorded.
Fat head minnows	Survival and growth to 7 days.	Quadruplicate batches of 10 minnows were incubated and the number of mortalities in each batch was recorded. The final mean weight in each batch was also recorded.
Lemna	Frond growth and increase in weight	Quadruplicate replicates of Lemna were incubated and the number of new fronds and final weight were recorded for each surviving organism.
Nautilus Studies		
Daphnia	Survival for 6 days and reproduction	10 individual organisms were incubated and the status (dead/alive) and reproductive output was recorded.
Rotifer	Reproduction after 49 hours.	8 individual organisms were incubated and the population growth was recorded.
Fat head minnows	Survival and growth	Triplicate batches of 10 minnows were incubated and the number of mortalities was recorded. The final mean

	to 7 days.	weight in each batch was also recorded.
Bullfrog tadpoles	Survival and growth to 28 days.	Triplicate batches of 5 tadpoles were incubated and the number of mortalities in each batch was recorded. The final biomass was also recorded.
Algae	Cell yield	Four to 10 batches of 10,000 cells were incubated and the percentage increase in the number of cells was recorded.

Table RT-1. Summary of model fits[#] for the EC Rainbow Trout Eggs mortality experiment.

	Single curve for all hardness levels		Common NR with separate curves		Separate curve for each hardness but equal LC10		Separate curve for each hardness but equal LC25		Separate curve for each hardness but equal LC50		Separate curve for each hardness	
	Est	**SE**	**Est**	**SE**	**Est**	**SE**	**Est**	**SE**	**Est**	**SE**	**Est**	**SE**
$AICc$	345.1		350.0		350.9		352.1		353.2		356.2	
$\Delta AICc$[2]	0.00		4.95		5.80		6.97		8.09		11.07	
Model weight[3]	0.84		0.07		0.05		0.03		0.01		0.00	
β_0 **50**	-4.46	0.99	-4.70	1.37	-5.32	1.24	-5.69	1.53	-4.33	1.91	-4.95	1.50
β_0 **100**	.	.	-4.20	1.61	-4.26	1.07	-4.01	1.46	-4.24	1.53	-4.11	1.63
β_0 **250**	.	.	-4.93	1.84	-3.78	0.97	-3.51	1.60	-5.67	1.95	-5.17	2.08
β_1 **50**	0.63	0.15	0.73	0.22	0.81	0.20	0.84	0.23	0.61	0.28	0.76	0.23
β_1 **100**	.	.	0.58	0.25	0.60	0.18	0.56	0.23	0.60	0.22	0.57	0.25
β_1 **250**	.	.	0.66	0.28	0.50	0.17	0.48	0.25	0.80	0.27	0.69	0.30
LC10 50	152	64	110	60	149	61	185	82	148	130	127	74
LC10 100			148	106			129	83	142	107	138	108
LC10 250			245	140			107	95	243	145	275	182
LC25 50	398	104	255	91	316	97	381	122	401	184	284	114
LC25 100			418	171	413	128			392	163	399	180
LC25 250			612	209	503	180			522	196	661	281
LC50 50	1159	301	647	186	729	219	847	255	1214	313	695	225
LC50 100			1329	647	1284	615	1269	609			1295	652
LC50 250			1693	794	1941	1318	1564	992			1752	852

Table RT-1. Summary of model fits[#] for the EC Rainbow Trout Eggs mortality experiment.

	Single curve for all hardness levels		Common NR with separate curves		Separate curve for each hardness but equal LC10		Separate curve for each hardness but equal LC25		Separate curve for each hardness but equal LC50		Separate curve for each hardness	
	Est	**SE**	**Est**	**SE**	**Est**	**SE**	**Est**	**SE**	**Est**	**SE**	**Est**	**SE**
NR[4] 50	0.10	0.03	0.10	0.03	0.14	0.06	0.17	0.08	0.17	0.08	0.13	0.06
NR 100					0.09	0.04	0.09	0.03	0.09	0.04	0.09	0.04
NR 250					0.09	0.05	0.08	0.04	0.09	0.05	0.11	0.06
σ^2	0.56	0.15	0.48	0.13	0.49	0.14	0.49	0.13	0.52	0.15	0.50	0.14

All models were fit using maximum likelihood using Proc NLMIXED of SAS 9.2. Non-linear constraints (e.g. models where the LC50 were constrained to be equal among the three curves) were fit using penalized likelihood where the penalty function went to zero as the constraint was satisfied.

[1] Akaike information criterion corrected for small sample sizes.

[2] Difference in AICc from best model.

[3] Model weights measure the support for each model.

[4] Natural Response, i.e. mortality at dose 0.

Table CH-1. Summary of model fits[#] for the EC Chinook Eggs mortality experiment.				
	Single Curve		Separate curve for each hardness	
	Est	**SE**	**Est**	**SE**
$AICc$	188.4		164.8	
$\Delta AICc$[2]	23.6		0.00	
Model weight[3]	0.00		1.00	
β_0 **50**	-3.44	0.41	-2.21	0.57
β_0 **100**			-1.98	0.12
β_0 **250**			-6.95	0.99
β_1 **50**	0.26	0.06	0.04	0.09
β_1 **100**			0.00[1]	.
β_1 **250**			0.79	0.14
LC10 50	4373	2149	*	*
LC10 100		2149	*	*
LC10 250		2149	1247	151
LC25 50	*	*	*	*
LC25 100	*	*	*	*
LC25 250	*	*	*	*
LC50 50	*	*	*	*
LC50 100	*	*	*	*
LC50 250	*	*	*	*

[#]All models were fit using maximum likelihood. Boundary non-negativity constraints were applied to the slopes of the probit model for all hardness concentrations.

[1] Slope constrained to be non-negative so no estimate of precision possible.

* Estimates of LCxx are too far outside range of doses tested in the experiment and would require extrapolation well beyond the range that is sensible.

Table HY-Mort-1. Summary of model fits[#] for the EC Hyalella mortality experiment.				
	Separate curve for each hardness		Single curve for all hardness levels	
	Est	**SE**	**Est**	**SE**
$AICc$	300.0		314.0	
$\Delta AICc$[2]	0.00		14.0	
Model weight[3]	1.00		0.00	
NR 50	0.09	0.01	0.08	0.01
NR 100	0.07	0.01		
NR 250	0.05	0.02		
β_0 **50**	-11.26	3.57	-10.00[1]	NA[1]
β_0 **100**	**	**		
β_0 **250**	-3.67	2.68		
β_1 **50**	1.38	0.47	1.13	0.02
β_1 **100**	**	**		
β_1 **250**	0.26	0.36		
LC10 50	1400	232	2235	344
LC10 100	**	**		
LC10 250	8848	24576		
LC25 50	*	*	*	*
LC25 100	*	*	*	*
LC25 250	*	*	*	*
LC50 50	*	*	*	*
LC50 100	*	*	*	*
LC50 250	*	*	*	*

[#]All models were fit using maximum likelihood. Boundary non-negativity constraints were applied to the slopes of the probit model for all hardness concentrations.

[1] Intercept constrained to be not less than -10, so estimates of precision cannot be computed.

* Estimates of LCxx are too far outside range of doses tested in the experiment and would require extrapolation well beyond the range that is sensible.

** No dose-response curve for medium hardness could be fit and only a natural mortality was fit.

Table HY-Weight-1. Summary of model fits for the EC Hyalella weight experiment.				
	Separate curve for each hardness		Single curve for all hardness levels	
	Est	**SE**	**Est**	**SE**
AICc	-37.6		-32.0	
$\Delta AICc$	0.0		5.6	
Model weight	0.94		0.06	
IC10 50	1325	292	653	214
IC10 100	648	237		
IC10 250	533	190		
IC25 50	1880	306	1281	175
IC25 100	913	172		
IC25 250	1130	305		
IC50 50	>2150[1]		>2150[1]	
IC50 100	>2080[1]			
IC50 250	1809	149		

[1] ICxx value is greater than maximum dose use in the experiment.

Table MY-1. Summary of model fits[1] for the EC Mussel mortality trials						
	Single Curve for all hardness levels		Separate curve for each hardness but equal LC10		Separate curve for each hardness	
	Est	**SE**	**Est**	**SE**	**Est**	**SE**
$AICc$	102.3		102.7		101.2	
$\Delta AICc$	1.1		1.4		0.0	
Model weight[3]	0.28		0.23		0.47	
β_0 **50**	-2.61	0.69	-3.53	1.05	-3.51	1.04
β_0 **100**			-1.16	0.05	-1.51	1.22
β_0 **250**			-1.50	0.07	-2.21	1.63
β_1 **50**	0.25	0.11	0.46	0.17	0.46	0.16
β_1 **100**			-0.03	0.01	0.02	0.20
β_1 **250**			0.04	0.02	0.15	0.25
LC10 50[2]	222	138	131	85	130	84
LC10 100					>10,000	
LC10 250					583	886

[1] All models fit using maximum likelihood. Constraints on the LCxx were imposed using a penalized likelihood approach.

[2] Estimates of LC25 and LC50 are not presented because they require extrapolation well beyond the doses observed in the experiment.

[3] Two additional models are not shown. The model with constrained LC25 points had a model weight of 0.0. The model with constrained LC50 points had a model weight of 0.02.

Table FM-1. Summary of model fits[1] for the EC Fat Head Minnow mortality trials				
	Single Curve for all hardness levels		Separate curve for each hardness	
	Est	**SE**	**Est**	**SE**
$AICc$	395.1		180.1	
$\Delta AICc$	215.0		0.0	
Model weight[3]	0.0		1.0	
β_0 **50**	-7.07	0.57	-7.78	0.83
β_0 **100**			-15.22	1.98
β_0 **250**			-66.58	3249.53
β_1 **50**	0.90	0.08	1.14	0.12
β_1 **100**			2.03	0.27
β_1 **250**			8.11	407.97
LC10 50	623	57	292	41
LC10 100			972	91
LC10 250			3130	13171
LC25 50	1224	86	497	52
LC25 100			1311	97
LC25 250			3373	26888

[1] All models fit using maximum likelihood.

[2] Estimates of LC50 are not presented because they require extrapolation well beyond the doses observed in the experiment.

[3] Three additional models are not shown. Models with constrained LC10, LC25, and LC50 all had model weights of 0.0.

Table FM-Weight-1. Summary of model fits for the EC Fat Head Minnow weight experiment.				
	Separate curve for each hardness		Single curve for all hardness levels	
	Est	**SE**	**Est**	**SE**
$AICc$	-14.6		84.2	
$\Delta AICc$	0.0		98.8	
Model weight	1.0		0.0	
IC10 50	805	231	1923	533
IC10 100	1092	156		
IC10 250	NC[2]			
IC25 50	893	82	2468	427
IC25 100	1481	118		
IC25 250	NC[2]			
IC50 50	1062	198	2879	139
IC50 100	1860	198		
IC50 250	NC[2]			

[2] Not computable as isotonic line is completely flat.

Table LM-Weight-1. Summary of model fits for the EC Lemna weight experiment.				
	Separate curve for each hardness		Single curve for all hardness levels	
	Est	**SE**	**Est**	**SE**
$AICc$	391.4		387.8	
$\Delta AICc$	3.6		0.0	
Model weight	0.14		0.86	
IC10 50	NC[2]		>2290[1]	
IC10 100	NC			
IC10 250	NC			
IC25 50	NC		>2290[1]	
IC25 100	NC			
IC25 250	NC			
IC50 50	NC		>2290[1]	
IC50 100	NC			
IC50 250	NC			

[1] ICxx value is greater than maximum dose use in the experiment.

[2] Not computable as isotonic line is completely flat.

Table LM-Frond-1. Summary of model fits for the EC Lemna frond increase experiment.				
	Separate curve for each hardness		Single curve for all hardness levels	
	Est	**SE**	**Est**	**SE**
$AICc$	692.9		678.5	
$\Delta AICc$	14.3		0.0	
Model weight[3]	0.0		1.0	
IC10 50	1106	493	2200	407
IC10 100	1759	469		
IC10 250	>2190[1]			
IC25 50	>2290[1]		>2290[1]	
IC25 100	>2085[1]			
IC25 250	>2190[1]			
IC50 50	>2290[1]		>2290[1]	
IC50 100	>2085			
IC50 250	>2290[1]			

[1] ICxx value is greater than maximum dose use in the experiment.

[2] Not computable as isotonic line is completely flat.

Table NA-DA-1. Summary of model fits[1] for the NA Daphnia mortality trials				
	Single Curve for all hardness levels		Separate curve for each hardness	
	Est	**SE**	**Est**	**SE**
$AICc$	172.3		158.6	
$\Delta AICc$	13.6		0.0	
Model weight[2]	0.0		1.0	
β_0 **40**	-10.78	1.01	-9.85	1.74
β_0 **80**			-8.69	1.69
β_0 **160**			-21.00	5.56
β_0 **320**			-13.69	2.58
β_1 **40**	1.51	0.14	1.49	0.26
β_1 **80**			1.22	0.24
β_1 **160**			2.88	0.77
β_1 **320**			1.87	0.36
LC10 40	533	50	318	63
LC10 80			432	102
LC10 160			941	134
LC10 320			767	113
LC25 40	797	59	478	76
LC25 80			712	130
LC25 160			1163	133
LC25 320			1062	129
LC50 40	1245	91	753	113

Table NA-DA-1. Summary of model fits[1] for the NA Daphnia mortality trials				
	Single Curve for all hardness levels		Separate curve for each hardness	
	Est	**SE**	**Est**	**SE**
LC50 80			1237	222
LC50 160			1469	165
LC50 320			1523	194

[1] All models fit using maximum likelihood.

[2] Three additional models are not shown. Models with constrained LC10, LC25, and LC50 all had model weights of 0.0.

Table NA-DA-Repro-1. Summary of model fits for the NA Daphnia reproduction experiment.				
	Separate curve for each hardness		Single curve for all hardness levels	
	Est	**SE**	**Est**	**SE**
$AICc$	3027.2		3135.2	
$\Delta AICc$	0		108.0	
Model weight	1.0		0.0	
IC10 40	86	24	351	100
IC10 80	441	169		
IC10 160	801	105		
IC10 320	357	27		
IC25 40	125	51	791	176
IC25 80	752	125		
IC25 160	898	32		
IC25 320	419	63		
IC50 40	467	108	927	116
IC50 80	1056	137		
IC50 160	1084	40		
IC50 320	875	138		

Table NA-RO-Repro-1. Summary of model fits for the NA Rotifer reproduction experiment.				
	Separate curve for each hardness		Single curve for all hardness levels	
	Est	**SE**	**Est**	**SE**
$AICc$	304.2		310.0	
$\Delta AICc$	0.0		5.8	
Model weight	0.95		0.05	
IC10 40	697	290	888	256
IC10 80	391	175		
IC10 160	697	183		
IC10 320	843	101		
IC25 40	1051	195	1376	238
IC25 80	1820	409		
IC25 160	1286	149		
IC25 320	1026	203		
IC50 40	1641	356	2197	192
IC50 80	2196	213		
IC50 160	>1800[1]			
IC50 320	>1800[1]			

[1] ICxx value is greater than maximum dose use in the experiment.

Table NA-FM-1. Summary of model fits[1,2] for the NA Fathead Minnow mortality trials				
	Single Curve for all hardness levels		Separate curve for each hardness	
	Est	**SE**	**Est**	**SE**
$AICc$	404.5		306.8	
$\Delta AICc$	97.7		0.0	
Model weight[2]	0.0		1.0	
β_0 **40**	-5.17	0.34	-6.18	0.67
β_0 **80**			-4.60	0.57
β_0 **160**			-6.93	0.97
β_0 **320**			-4.51	0.99
β_1 **40**	0.62	0.05	0.85	0.10
β_1 **80**			0.60	0.08
β_1 **160**			0.81	0.13
β_1 **320**			0.41	0.13
LC10 40	535	58	327	60
LC10 80			420	85
LC10 160			1102	201
LC10 320			2490	923
LC25 40	1425	124	670	94
LC25 80			989	152
LC25 160			2341	381
LC25 320			10859	7814
LC50 40	4234	531	1489	206

Table NA-FM-1. Summary of model fits[1,2] for the NA Fathead Minnow mortality trials				
	Single Curve for all hardness levels		Separate curve for each hardness	
	Est	**SE**	**Est**	**SE**
LC50 80			2551	464
LC50 160			5402	1230
LC50 320			55771	67658

[1] All models fit using maximum likelihood.

[2] Three additional models are not shown. Models with constrained LC10, LC25, and LC50 all had model weights of 0.0.

Table NA-FM-Weight-1. Summary of model fits for the NA Fathead Minnow final biomass experiment.				
	Separate curve for each hardness		Single curve for all hardness levels	
	Est	**SE**	**Est**	**SE**
$AICc$	-26.2		14.2	
$\Delta AICc$	0.0		40.4	
Model weight	1.00		0.00	
IC10 40	514	167	844	360
IC10 80	1056	498		
IC10 160	296	200		
IC10 320	637	193		
IC25 40	759	128	1355	189
IC25 80	1550	187		
IC25 160	2659	700		
IC25 320	1766	599		
IC50 40	1172	185	2614	670
IC50 80	2365	287		
IC50 160	3713	421		
IC50 320	>5259			

[1] ICxx value is greater than maximum dose use in the experiment.

Table NA-TA-1. Summary of model fits[1,2] for the NA Tadpole mortality trials				
	Single Curve for all hardness levels		Separate curve for each hardness	
	Est	**SE**	**Est**	**SE**
$AICc$	72.1		69.3	
$\Delta AICc$	2.8		0.0	
Model weight[2]	0.34		0.66	
β_0 **15**	-5.15	1.05	-9.21	2.89
β_0 **80**			-4.44	1.18
β_1 **15**	0.67	0.16	1.21	0.41
β_1 **80**			0.61	0.18
LC10 15	309	89	719	174
LC10 80			183	86
LC25 15	760	144	1189	221
LC25 80			498	142
LC50 15	2066	600	2082	574
LC50 80			1518	545

[1] All models fit using maximum likelihood.

Table NA-TA-Weight-1. Summary of model fits for the NA Tadpole final biomass experiment.				
	Separate curve for each hardness		Single curve for all hardness levels	
	Est	**SE**	**Est**	**SE**
$AICc$	300.7		308.7	
$\Delta AICc$	0.0		7.9	
Model weight	0.98		0.02	
IC10 15	1097	146	1020	576
IC10 90	1148	199		
IC25 15	1326	139	1239	176
IC25 80	1266	109		
IC50 15	1819	198	1853	170
IC50 80	1491	118		

Table NA-AL-CellIncrease-1. Summary of model fits for the NA Algae increase in cell numbers experiment.				
	Separate curve for each hardness		Single curve for all hardness levels	
	Est	**SE**	**Est**	**SE**
$AICc$	732.9		730.6	
$\Delta AICc$	2.3		0.0	
Model weight	0.24		0.76	
IC10 10	700	739	1470	322
IC10 80	1345	747		
IC10 320	1377	442		
IC25 10	1112	299	1767	168
IC25 80	1763	266		
IC25 320	1727	135		
IC50 10	1430	71	2402	160
IC50 80	2742	229		
IC50 320	2518	165		

Table-Summary. Summary of estimated ECxx/Icxx values from the species tested in the two reports. The Ecxx/Icxx value si reported along with a 95% confidence interval for the parameter. Values more than 2x the maximum dose used in the experiment are reported ax >xxxx.

Rep	Species	**Endpoint**	**model**	**Hardness**	**EC10 or IC10**	**EC25 or IC25**	**EC50 or IC50**
EC	CH	mortality	Common	All	> 4340 (1633,> 4340)	> 4340 (> 4340,> 4340)	> 4340 (> 4340,> 4340)
			Separate	50	> 4340 (0,> 4340)	> 4340 (0,> 4340)	> 4340 (0,> 4340)
				100	> 4340 (> 4340,> 4340)	> 4340 (> 4340,> 4340)	> 4340 (> 4340,> 4340)
				250	1248 (979, 1591)	2678 (1817, 3948)	> 4340 (3262,> 4340)
	FH	mortality	Common	All	624 (520, 748)	1224 (1063, 1410)	2590 (2132, 3147)
			Separate	50	293 (221, 388)	498 (403, 615)	898 (740, 1089)
				100	972 (805, 1174)	1312 (1132, 1521)	1830 (1582, 2118)
				250	3130 (1,> 5758)	3373 (0,> 5758)	3666 (0,> 5758)
		weight	Common	All	1924 (1113,> 5758)	2468 (1541,> 5758)	> 5758 (2235,> 5758)
			Separate	50	805 (378, 863)	893 (822, 1057)	1062 (1009, 1482)
				100	1092 (945,> 5758)	1480 (1190,> 5758)	1859 (1667,> 5758)
				250	> 5758 (> 5758,> 5758)	> 5758 (> 5758,> 5758)	> 5758 (> 5758,> 5758)
	HY	mortality	Common	All	2235 (1647, 3034)	3824 (2758,> 4300)	> 4300 (> 4300,> 4300)
			Probit,NR,Probit	50	1401 (1007, 1947)	2176 (1649, 2871)	3550 (2149,> 4300)

				100	> 4300 (> 4300,> 4300)	> 4300 (> 4300,> 4300)	> 4300 (> 4300,> 4300)
				250	> 4300 (36,> 4300)	> 4300 (1,> 4300)	> 4300 (0,> 4300)
		weight	Common	All	653 (287, 1094)	1282 (680, 1539)	> 4300 (2120,> 4300)
			Separate	50	1326 (469, 1439)	1880 (1208,> 4300)	> 4300 (> 4300,> 4300)
				100	647 (135, 770)	912 (552, 1269)	> 4300 (1647,> 4300)
				250	533 (285, 905)	1130 (490, 1334)	1808 (1512,> 4300)
	LM	frond	Common	All	2200 (1213,> 4580)	> 4580 (2239,> 4580)	> 4580 (> 4580,> 4580)
			Separate	50	1107 (442,> 4580)	> 4580 (1710,> 4580)	> 4580 (> 4580,> 4580)
				100	1760 (812,> 4580)	> 4580 (1850,> 4580)	> 4580 (> 4580,> 4580)
				250	> 4580 (1638,> 4580)	> 4580 (> 4580,> 4580)	> 4580 (> 4580,> 4580)
		weight	Common	All	> 4580 (2262,> 4580)	> 4580 (> 4580,> 4580)	> 4580 (> 4580,> 4580)
			Separate	50	> 4580 (2047,> 4580)	> 4580 (> 4580,> 4580)	> 4580 (> 4580,> 4580)
				100	> 4580 (949,> 4580)	> 4580 (> 4580,> 4580)	> 4580 (> 4580,> 4580)
				250	> 4580 (> 4580,> 4580)	> 4580 (> 4580,> 4580)	> 4580 (> 4580,> 4580)
	MY	mortality	Common	All	223 (64, 772)	2619 (487,> 4230)	> 4230 (823,> 4230)
			Common LC10	50	132 (36, 483)	494 (211, 1158)	2143 (564,> 4230)
				100	132 (36, 484)	0 (0, 0)	0 (0, 0)

				250	132 (36, 484)	> 4230 (> 4230,> 4230)	> 4230 (> 4230,> 4230)
			Separate	50	131 (36, 475)	492 (212, 1138)	2143 (571,> 4230)
				100	> 4230 (0,> 4230)	> 4230 (0,> 4230)	> 4230 (0,> 4230)
				250	584 (28,> 4230)	> 4230 (0,> 4230)	> 4230 (0,> 4230)
	RT	mortality	Common	All	153 (66, 356)	399 (236, 673)	1159 (688, 1951)
			Common LC10	50	149 (65, 342)	317 (170, 588)	730 (399, 1333)
				100	149 (65, 342)	414 (222, 771)	1285 (492, 3358)
				250	149 (65, 342)	503 (246, 1031)	1942 (498,> 4230)
			Common LC25	50	186 (76, 453)	381 (200, 726)	847 (463, 1550)
				100	129 (35, 471)	381 (200, 726)	1269 (485, 3323)
				250	107 (18, 643)	381 (200, 727)	1565 (438,> 4230)
			Common LC50	50	148 (25, 865)	402 (160, 1009)	1214 (723, 2038)
				100	142 (31, 644)	393 (170, 905)	1215 (724, 2039)
				250	244 (74, 804)	522 (246, 1110)	1216 (725, 2039)
			Separate	50	128 (40, 410)	285 (127, 639)	696 (363, 1334)
				100	139 (29, 662)	400 (162, 988)	1295 (471, 3560)
				250	275 (73, 1041)	662 (282, 1554)	1752 (661,> 4230)
			Separate, Cmn NR	50	111 (37, 332)	256 (124, 526)	647 (363, 1154)
				100	148 (35, 623)	419 (184, 953)	1329 (500, 3533)
				250	246 (78, 774)	613 (309, 1215)	1694 (661,> 4230)

NA	AL	cell.incre	Common	All	1470 (960, 1503)	1767 (1491, 1869)	2402 (2167, 2730)
			Separate	10	700 (26, 1144)	1112 (366, 1292)	1429 (1253, 1605)
				80	1345 (209, 1477)	1763 (795, 2018)	2741 (2063, 3047)
				320	1377 (535, 1506)	1727 (1410, 1891)	2517 (2184, 2884)
	DA	mortality	Common	All	534 (443, 643)	797 (688, 924)	1245 (1077, 1439)
			Separate	40	318 (216, 470)	479 (350, 656)	753 (560, 1013)
				80	433 (272, 688)	712 (496, 1021)	1238 (868, 1764)
				160	942 (711, 1246)	1163 (927, 1457)	1469 (1177, 1835)
				320	768 (573, 1027)	1062 (835, 1350)	1524 (1185, 1960)
		repro	Common	All	351 (179, 599)	791 (472, 884)	927 (886, 1255)
			Separate	40	86 (82, 212)	126 (110, 391)	467 (308, 858)
				80	441 (189, 693)	752 (435, 850)	1056 (810, 1235)
				160	801 (396, 831)	898 (848, 925)	1085 (1027, 1151)
				320	357 (344, 484)	419 (383, 685)	875 (616, 1053)
	FH	mortality	Common	All	623 (509, 764)	1543 (1307, 1821)	4223 (3349, 5325)
			Separate	40	327 (227, 473)	671 (507, 888)	1489 (1131, 1961)
				80	420 (282, 625)	987 (727, 1339)	2551 (1777, 3660)

				180	1102 (766, 1585)	2340 (1693, 3235)	5403 (3438, 8491)
				360	2490 (1193, 5200)	10860 (2603,>11000)	>11000 (5017,>11000)
		weight	Common	All	844 (259, 956)	1356 (1038, 1666)	2612 (2084, 5261)
			Separate	40	514 (283, 670)	759 (483, 977)	1173 (919, 1657)
				80	1056 (512, 1428)	1550 (1190,>11000)	2366 (1863,>11000)
				160	296 (178, 1897)	2660 (1274,>11000)	3715 (2657,>11000)
				320	637 (369, 1128)	1767 (1120, 3210)	>11000 (>11000,>11000)
	RO	growth	Common	All	888 (457, 1143)	1376 (968, 1839)	2197 (1810,> 8000)
			Separate	40	697 (228,> 8000)	1052 (681,> 8000)	1641 (1121,> 8000)
				80	391 (170, 968)	1819 (531,> 8000)	2197 (2071,> 8000)
				160	696 (351, 970)	1286 (1074, 1569)	> 8000 (1786,> 8000)
				320	843 (790, 1131)	1026 (916,> 8000)	> 8000 (1133,> 8000)
	TA	mortality	Common	All	309 (171, 559)	761 (517, 1120)	2066 (1142, 3740)
			Common LC10	15	263 (131, 527)	962 (426, 2175)	> 3850 (571,> 3850)
				80	263 (131, 527)	606 (364, 1009)	1533 (813, 2889)
			Common LC25	15	545 (245, 1212)	947 (582, 1542)	1751 (1103, 2779)
				80	269 (95, 763)	947 (582, 1542)	3839 (662,> 3850)
			Common LC50	15	714 (440, 1159)	1143 (809, 1615)	1927 (1239, 2999)

				80	170 (54, 541)	538 (295, 979)	1927 (1238, 2999)
			Separate	15	719 (438, 1182)	1190 (813, 1741)	2082 (1186, 3657)
				80	183 (70, 480)	499 (278, 894)	1519 (729, 3164)
		weight	Common	All	1020 (216, 1139)	1239 (1046, 1691)	1852 (1470, 1880)
			Separate	15	1097 (601, 1190)	1326 (1114, 1703)	1819 (1536,> 3850)
				80	1147 (683, 1173)	1266 (831, 1337)	1491 (1151, 1664)

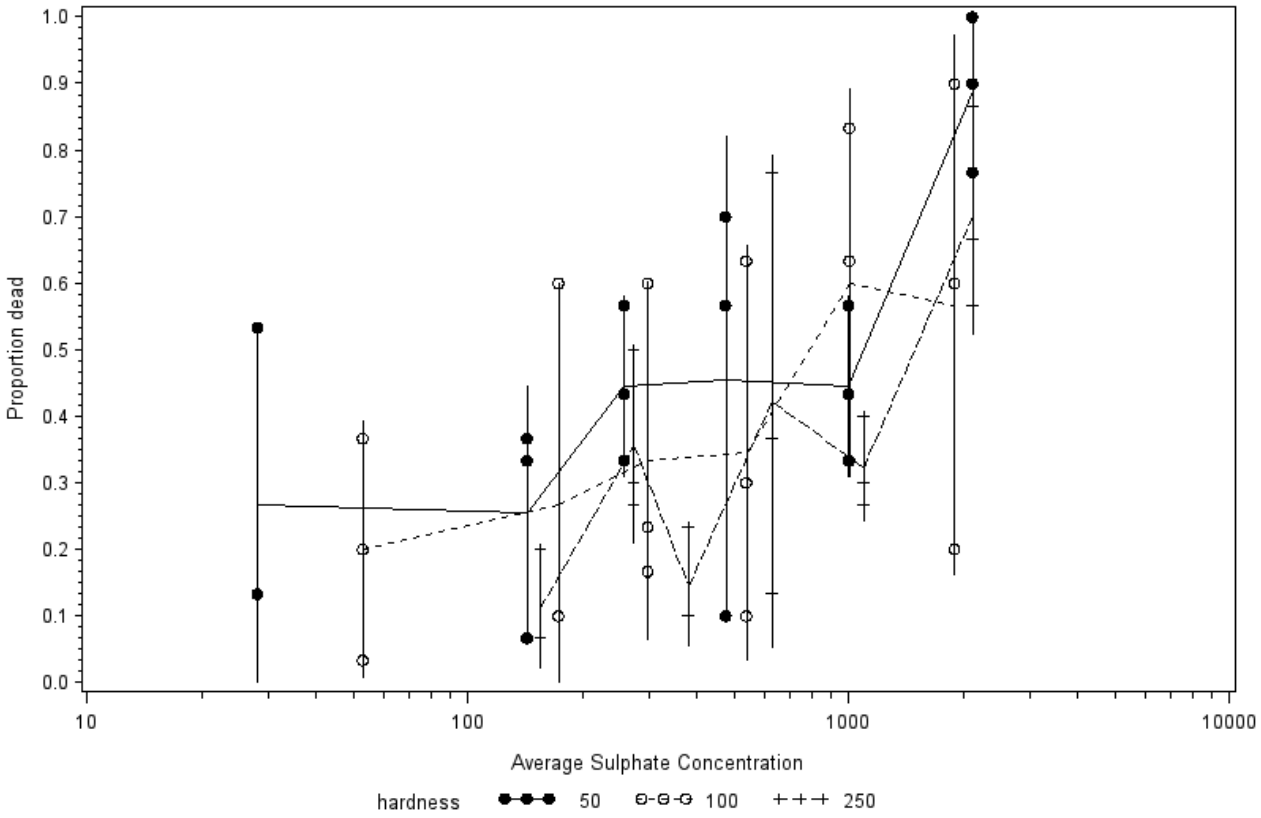

Figure EC-RT-1. Empirical mortality observed in the EC rainbow trout toxicity trials.

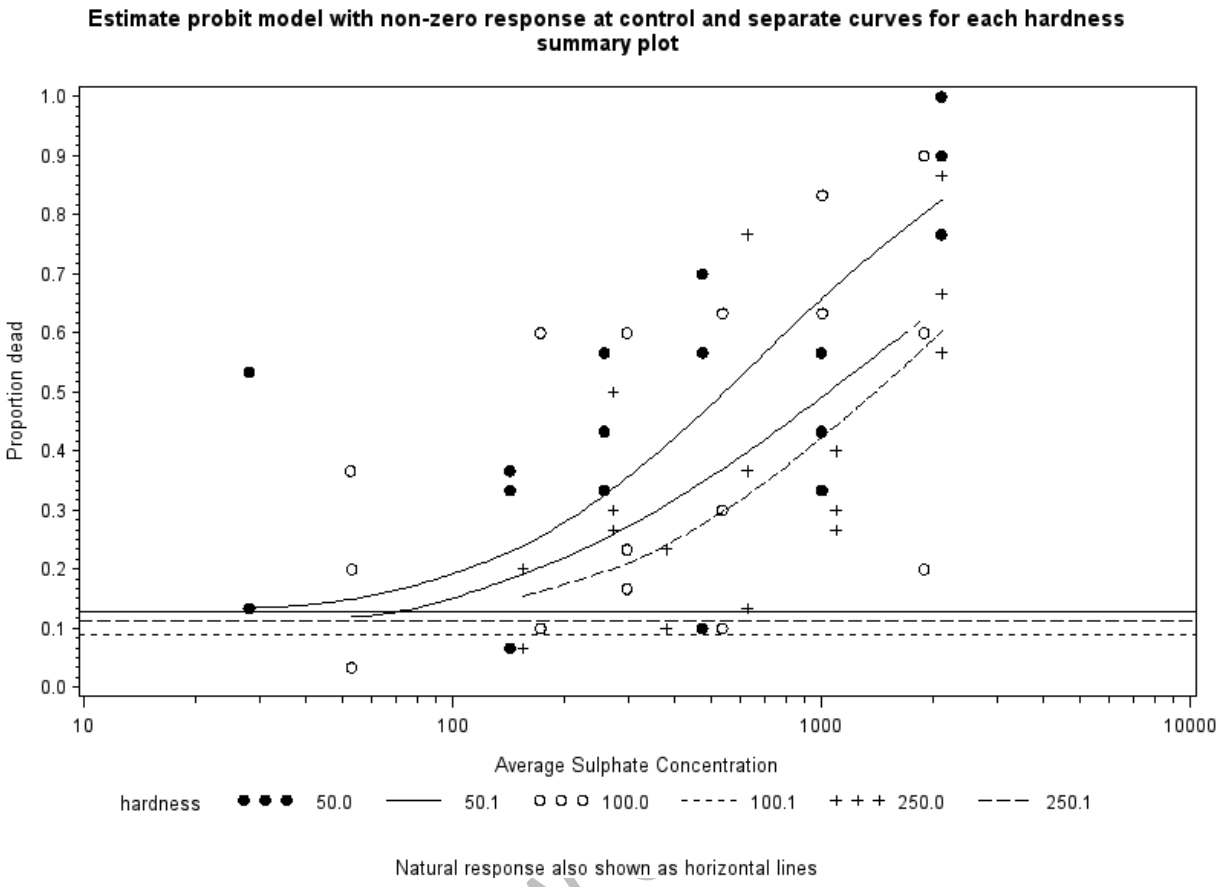

Figure EC-RT-2 Fitted probit curves for model with separate curve for each hardness for the EC rainbow trout egg tests.

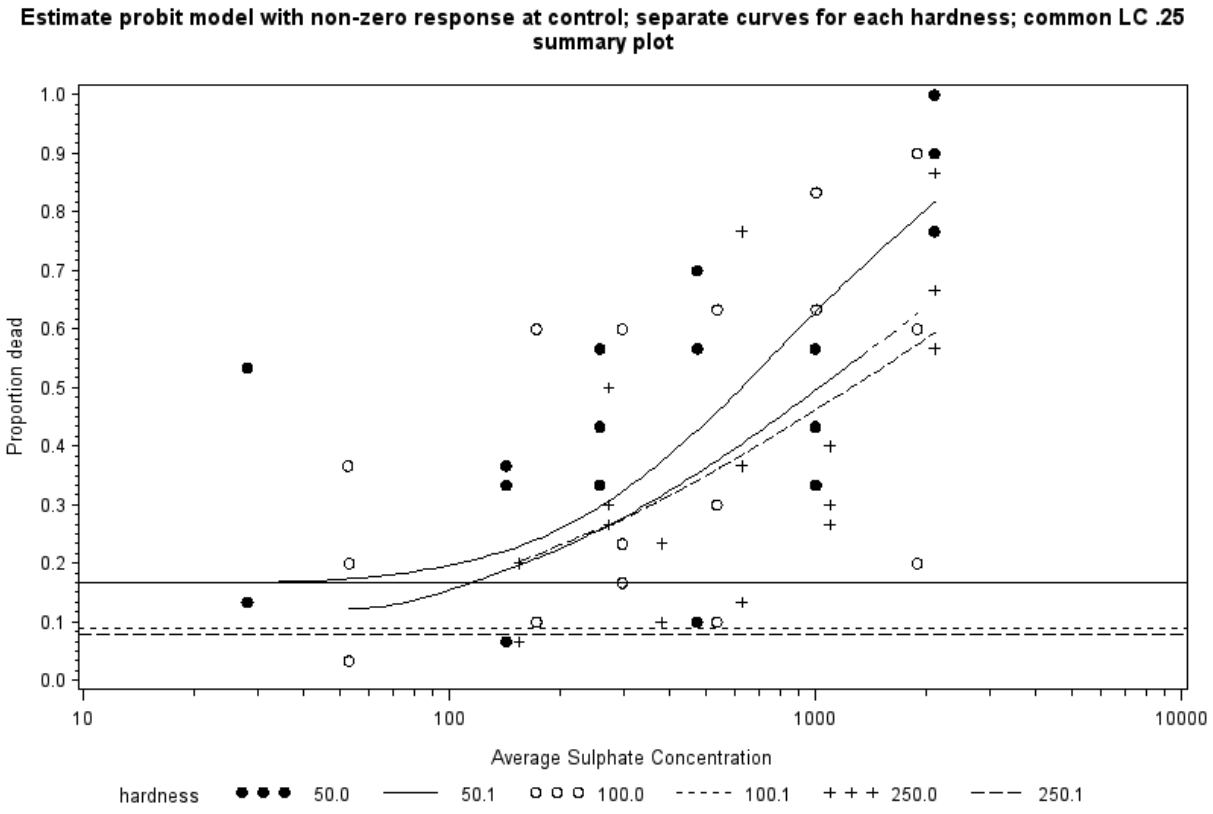

Figure EC-RT-3. Fitted probit curves for model where LC25 are common among models in the EC rainbow trout study. The common LC25 value is 381 (SE 122). Notice that the curves do not all meet at this point because of differing natural response values.

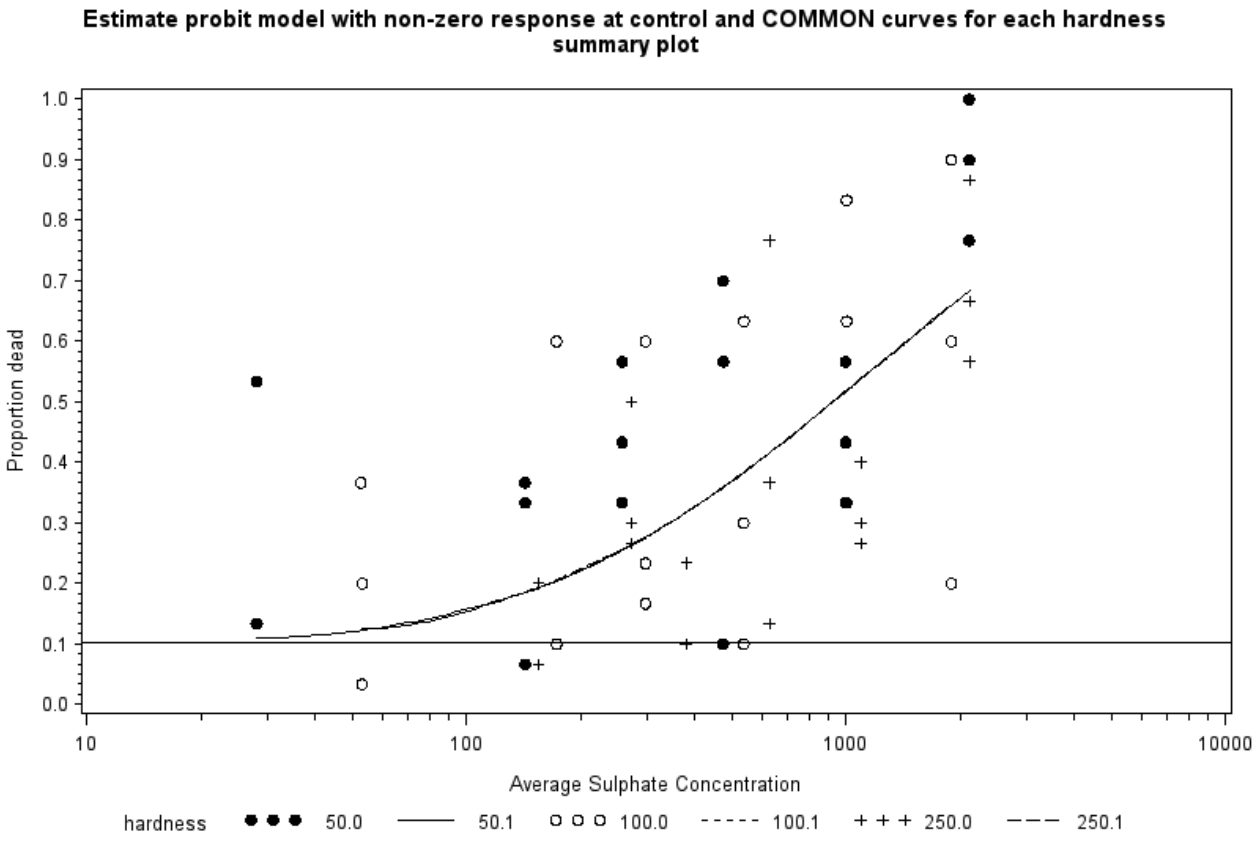

Figure EC-RT-4. Fitted probit curves for model where a single dose-response curve was fit to all hardness levels for the EC rainbow trout egg tests.

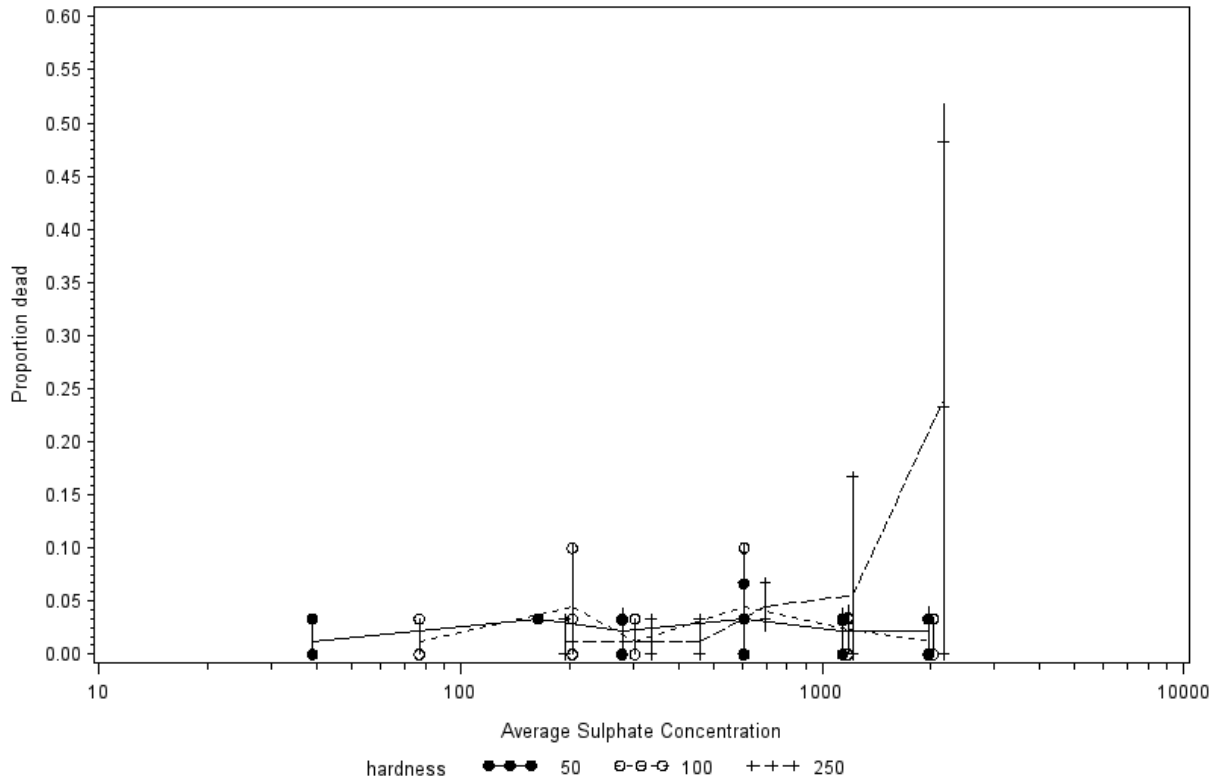

Figure EC-CH-1. Empirical mortality observed in the EC Chinook egg mortality trials.

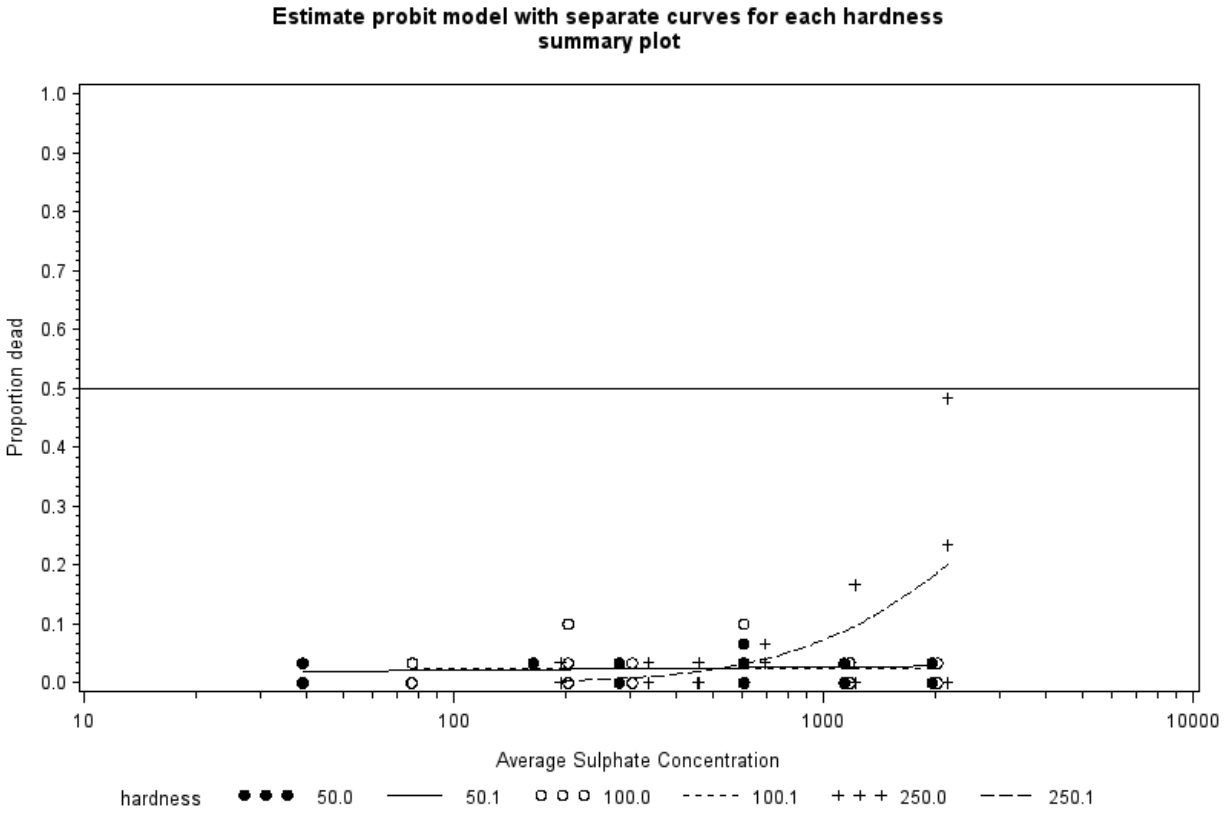

Figure EC-CH-2 Fitted probit curves for model with separate curve for each hardness for the EC Chinook egg tests.

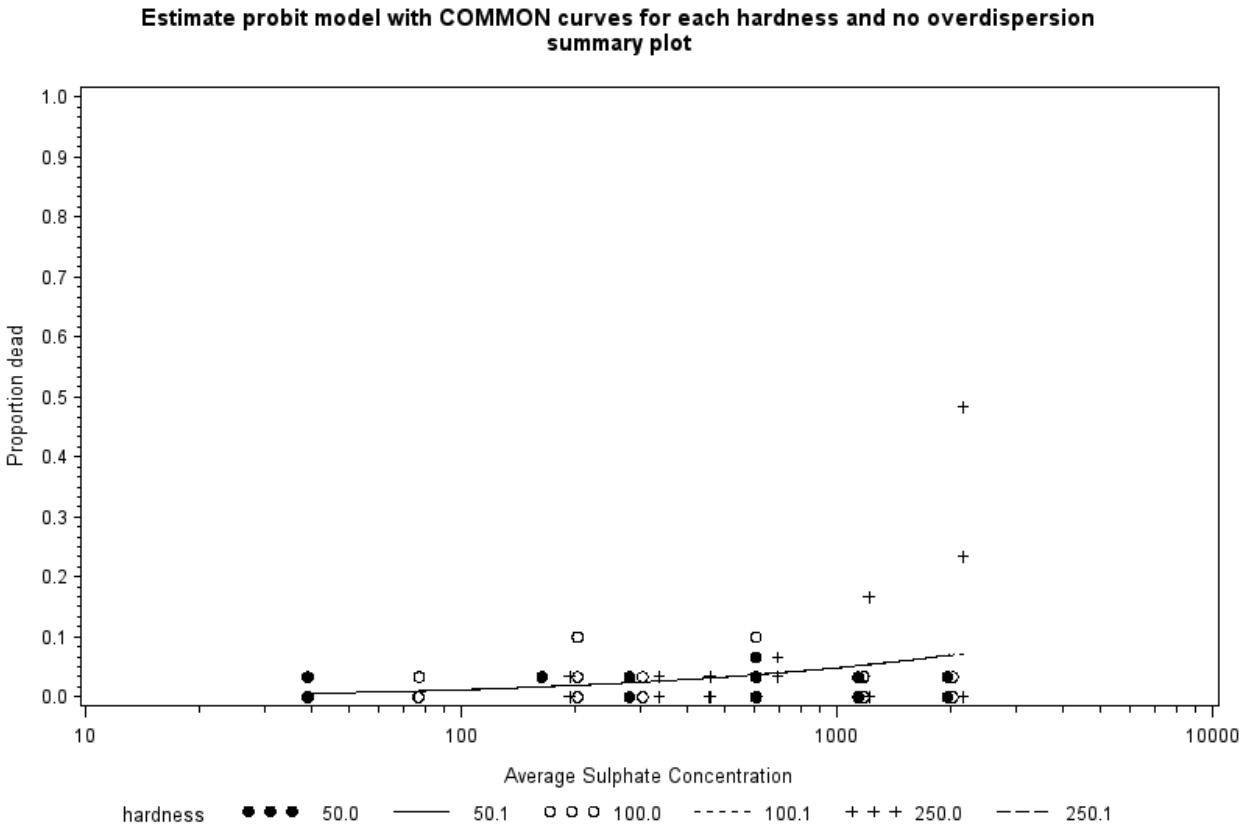

Figure EC-CH-3. Fitted probit curves for model where a single dose-response curve was fit to all hardness levels for the EC Chinook egg study.

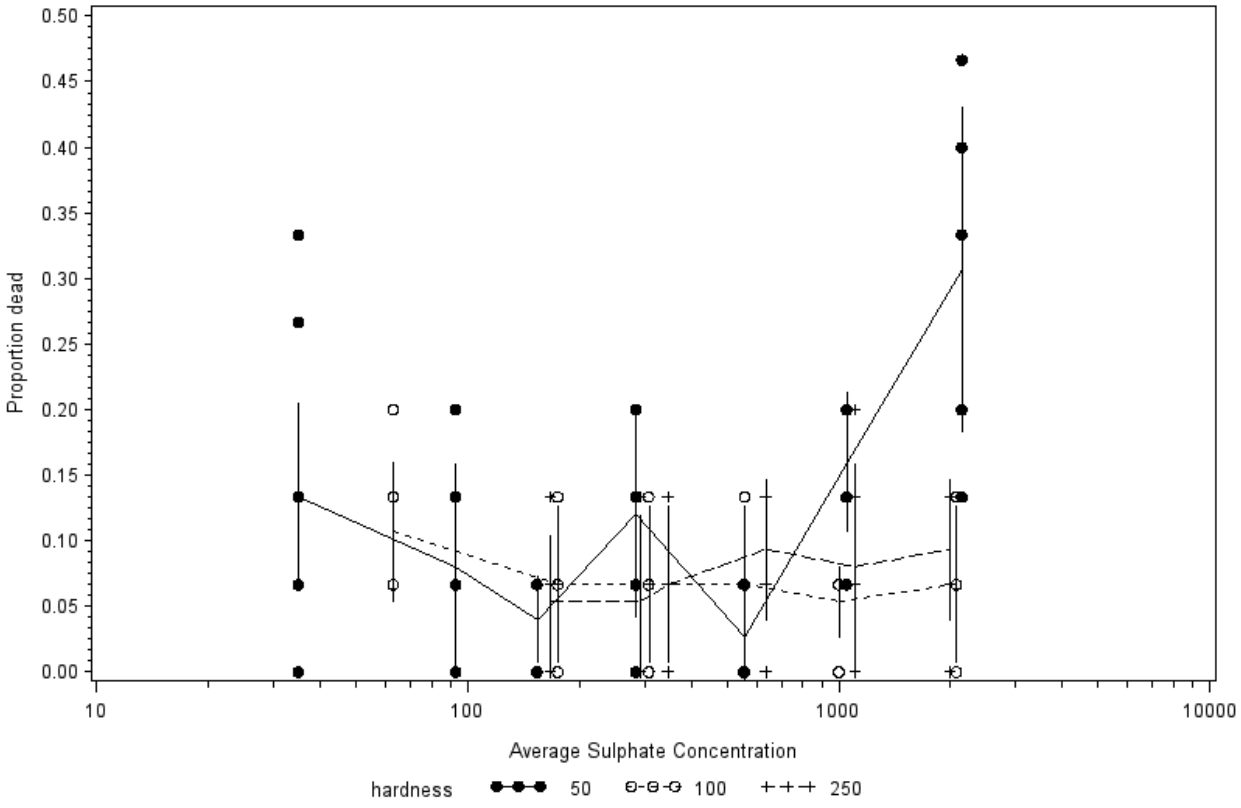

Figure EC-HY-Mort-1. Empirical mortality observed in the EC Hyalella mortality trials

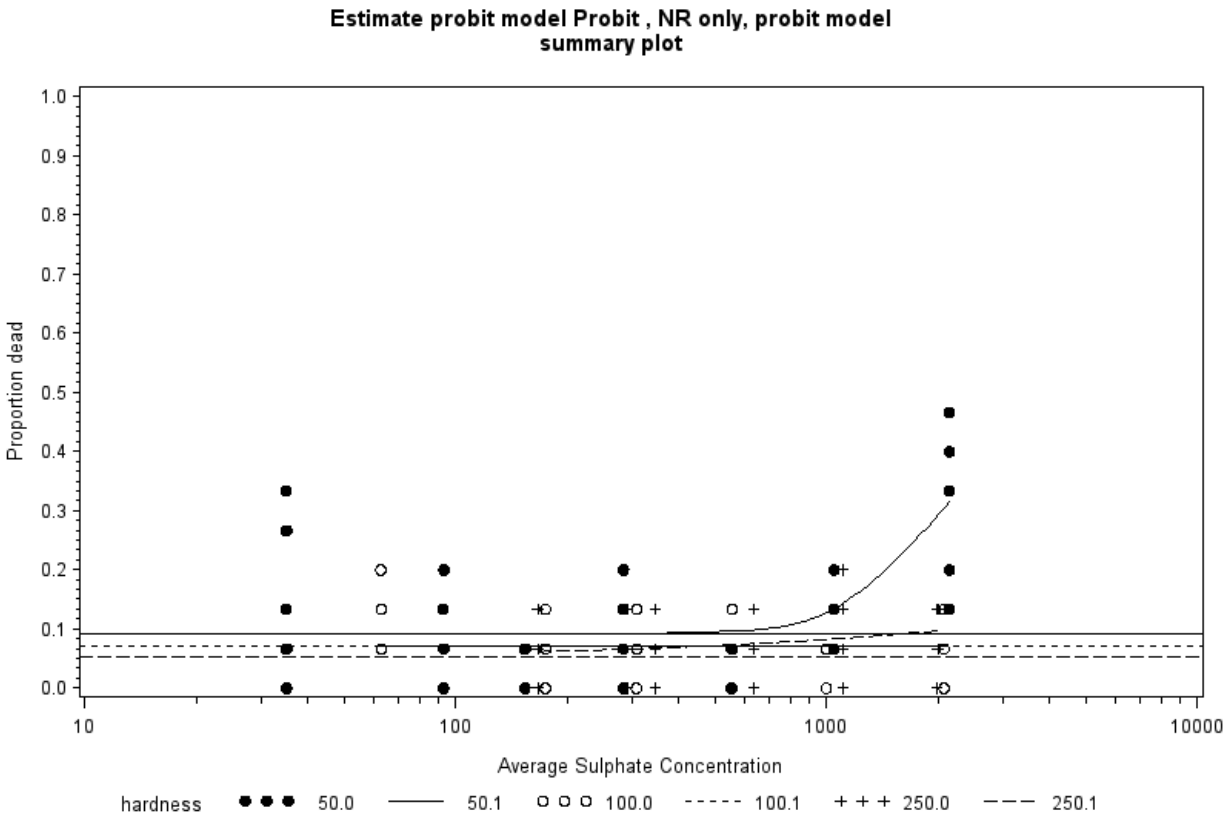

Figure EC-HY-Mort-2 Fitted probit curves for model with separate curve for each hardness for the EC Hyalella mortality tests. Note that only a constant natural mortality could be fit for the medium hardness.

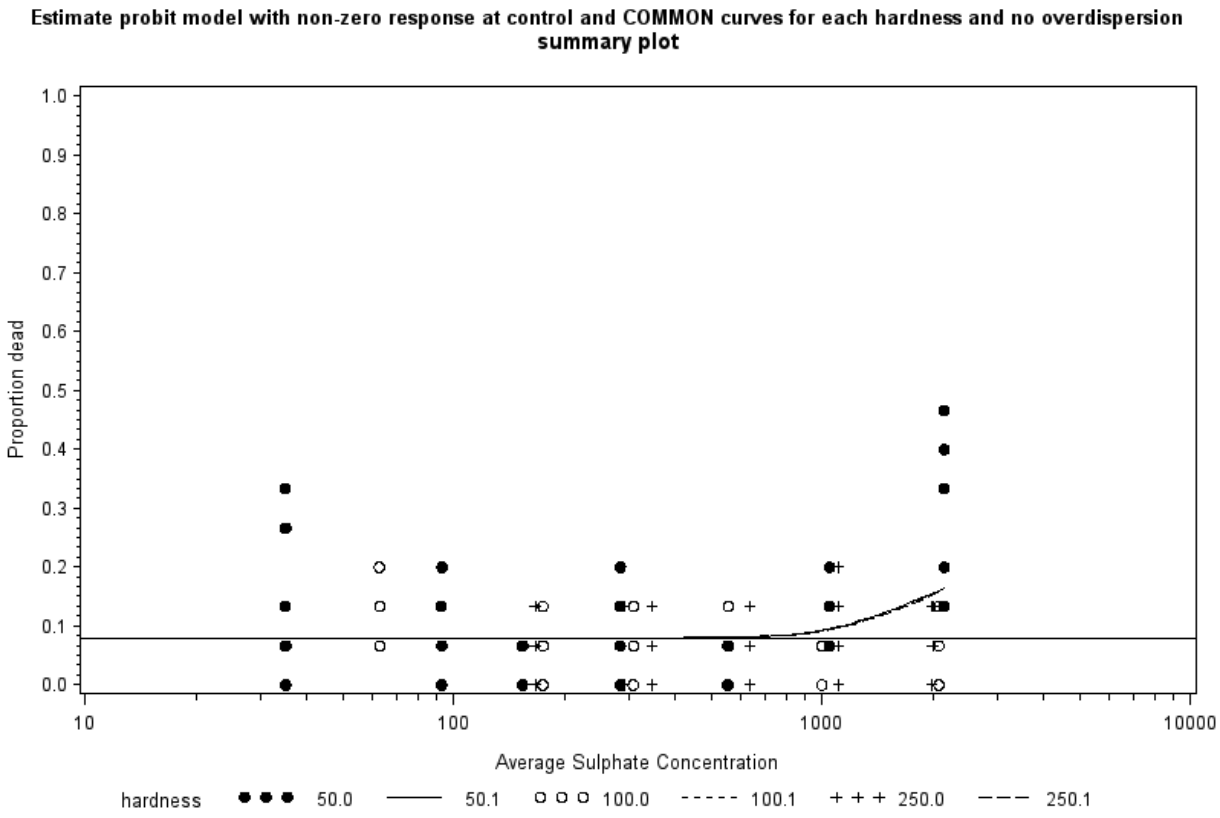

Figure EC-HY-Mort-3. Fitted probit curves for model where a single dose-response curve was fit to all hardness levels for the EC Hyalella mortality tests.

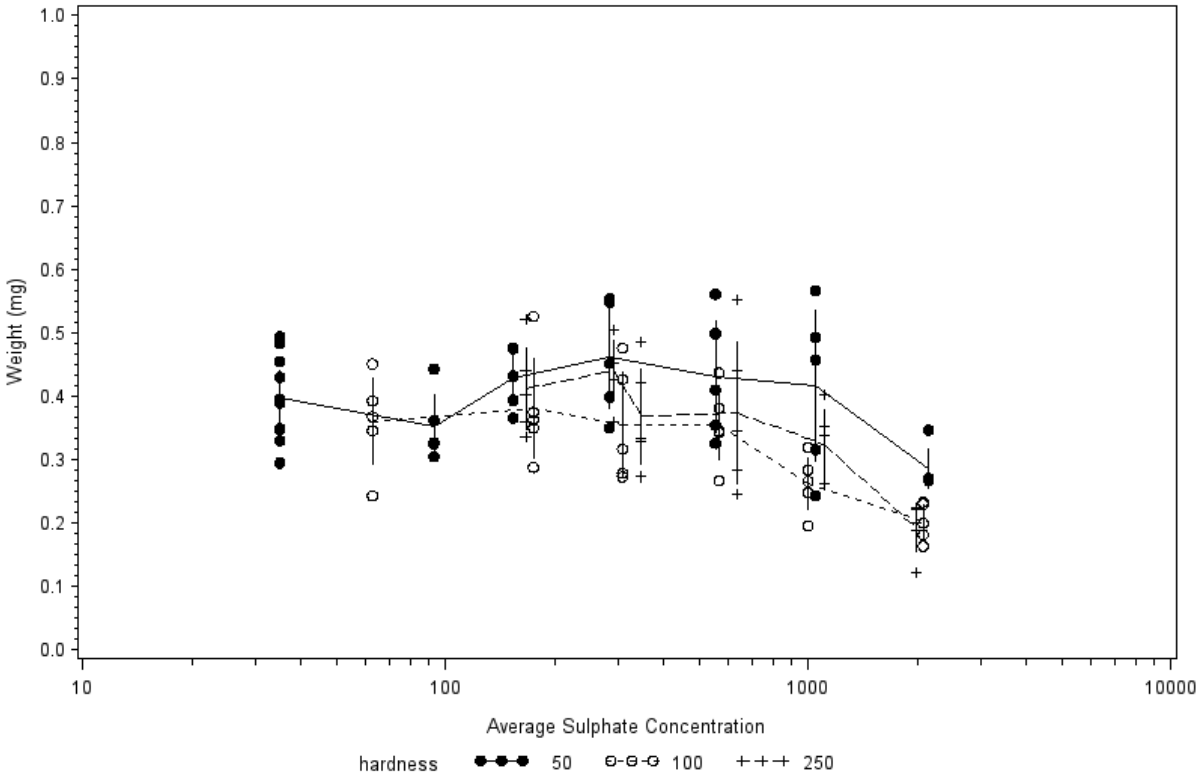

Figure EC-HY-Weight-1. Empirical mean weight at the end of the experiment observed in the EC Hyalella trials

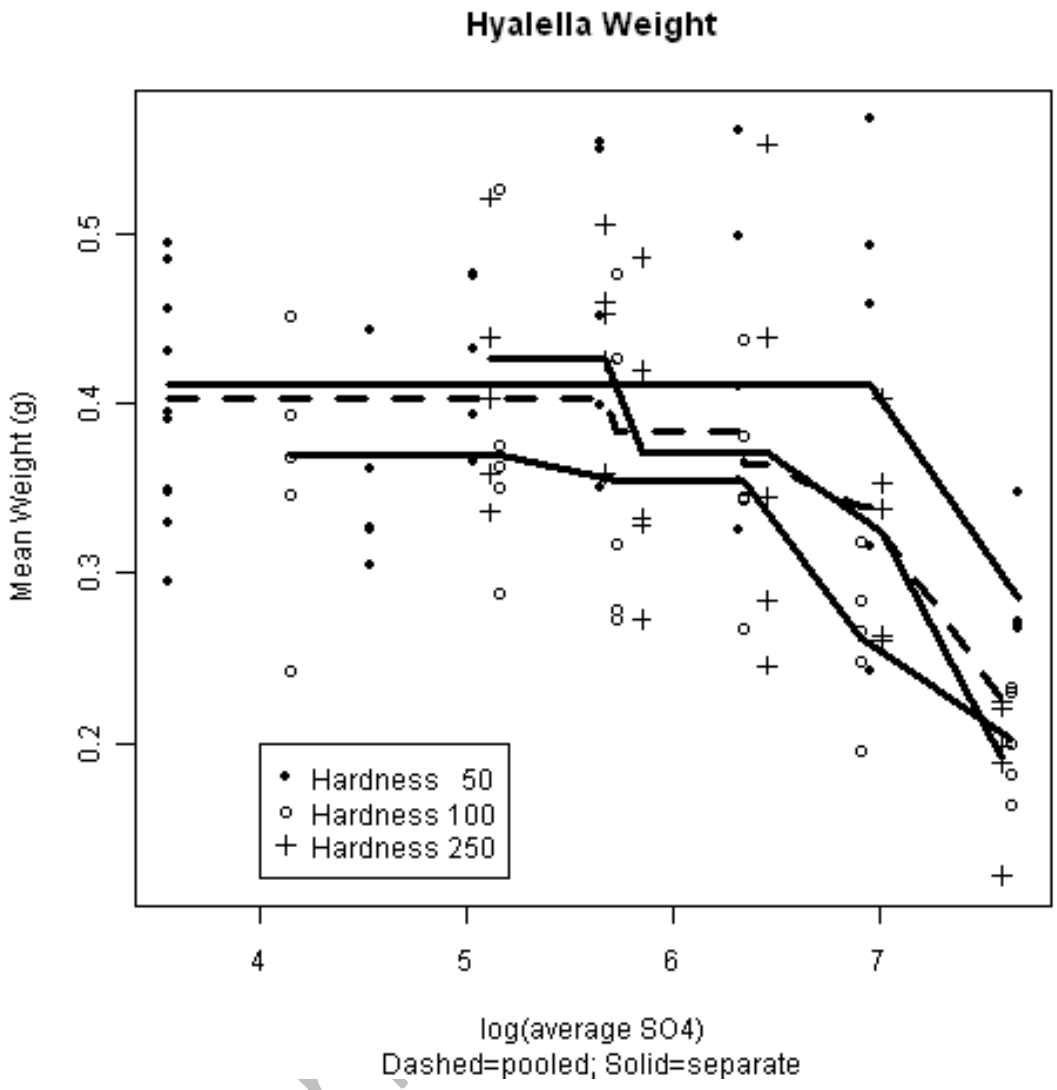

Figure EC-HY-Weight-2. Fitted isotonic curves for the mean weight at the end of the experiment in the EC Hyalella trials

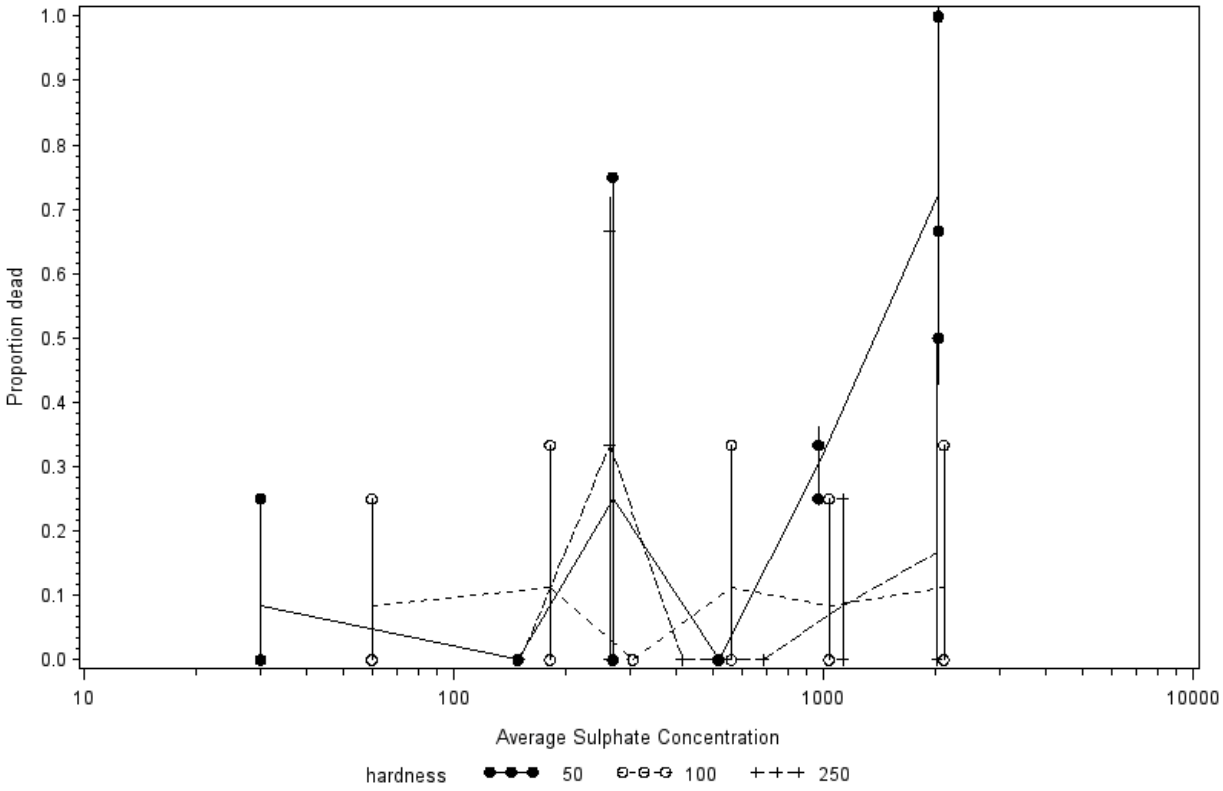

Figure EC-MY-1. Empirical mortality observed in the EC mussel mortality trials.

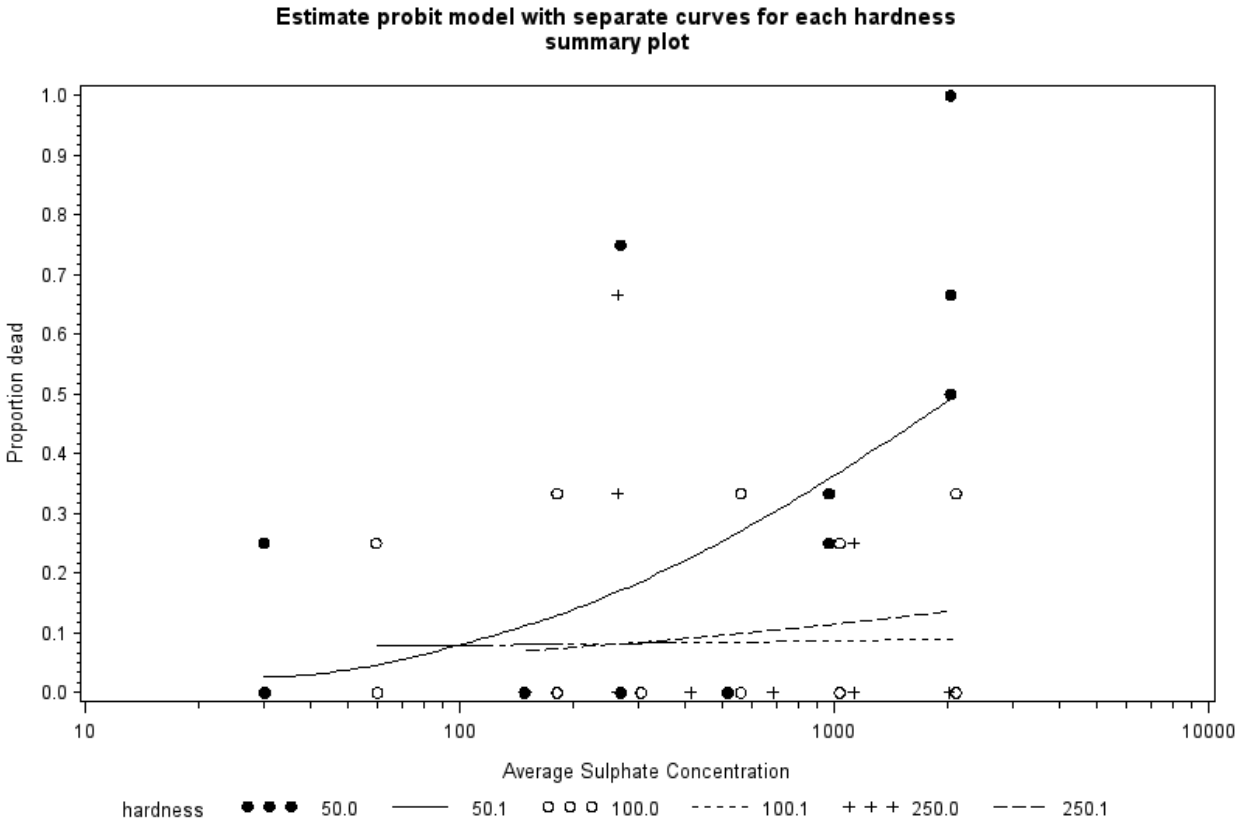

Figure EC-MY-2 Fitted probit curves for model with separate curve for each hardness for EC mussel mortality trials.

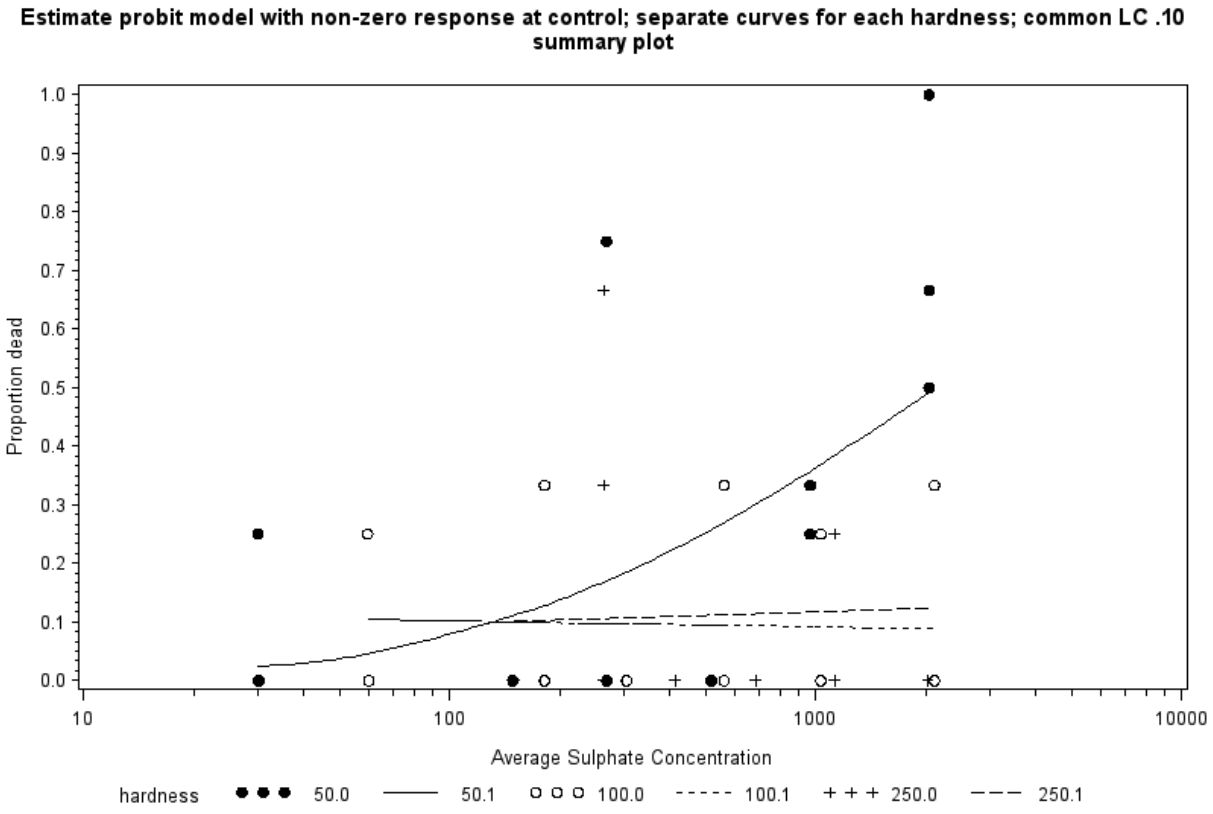

Figure EC-MY-3. Fitted probit curves for model with separate curve for each hardness but the LC10 are constrained to be equal for EC mussel mortality trials.

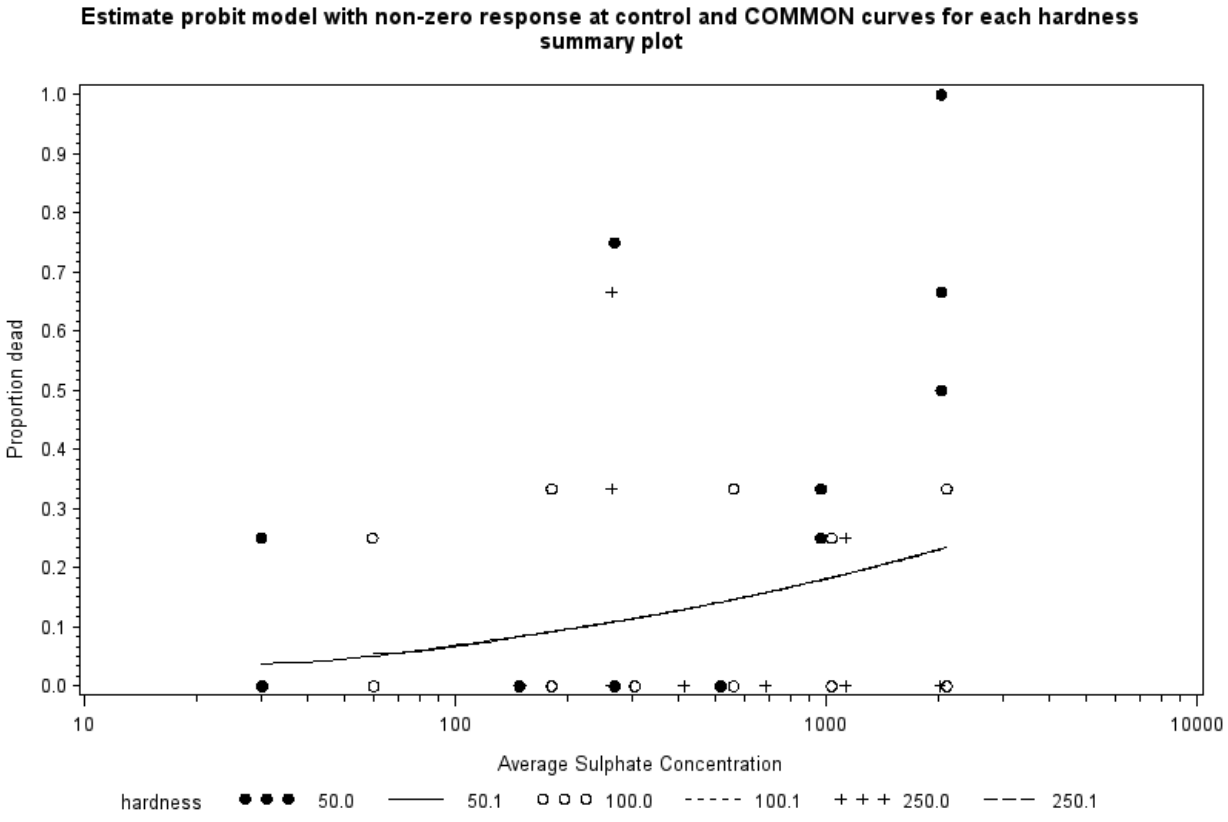

Figure EC-MY-4. Fitted probit curves for model with common probit model for all hardness levels in EC mussel study.

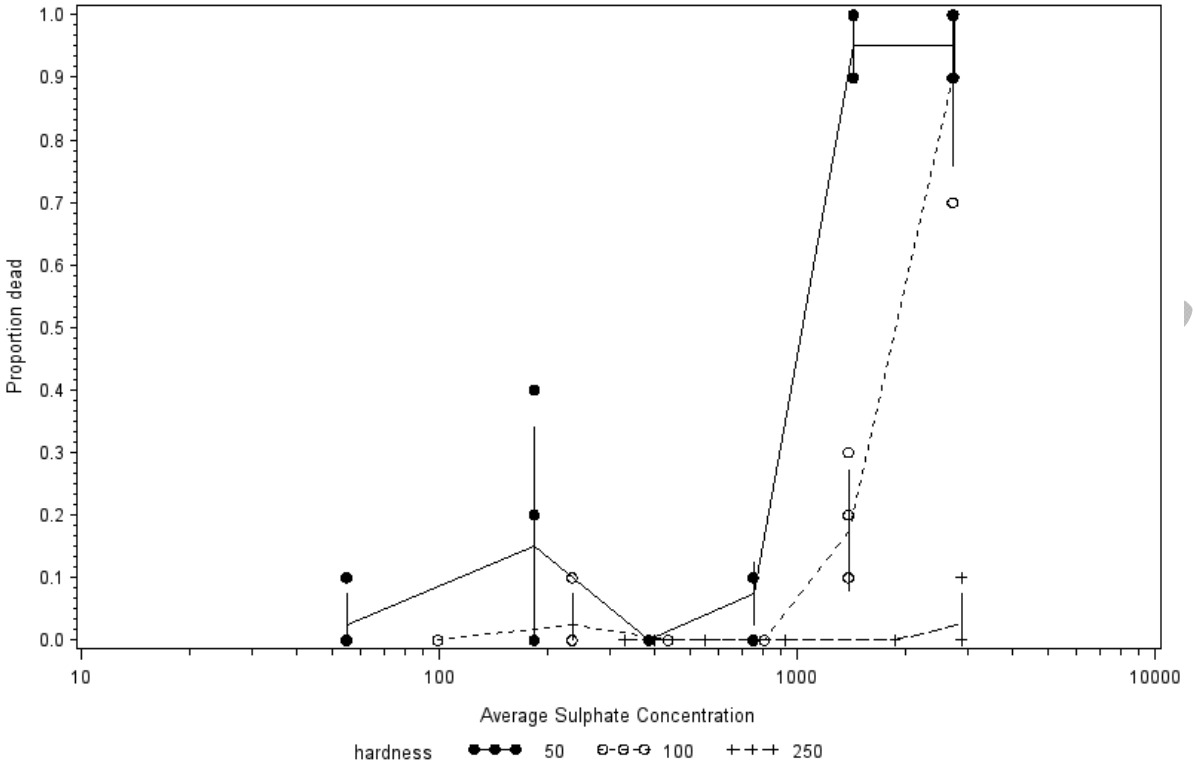

Figure EC-FM-Mort-1. Empirical mortality observed in the EC fat head minnow mortality trials.

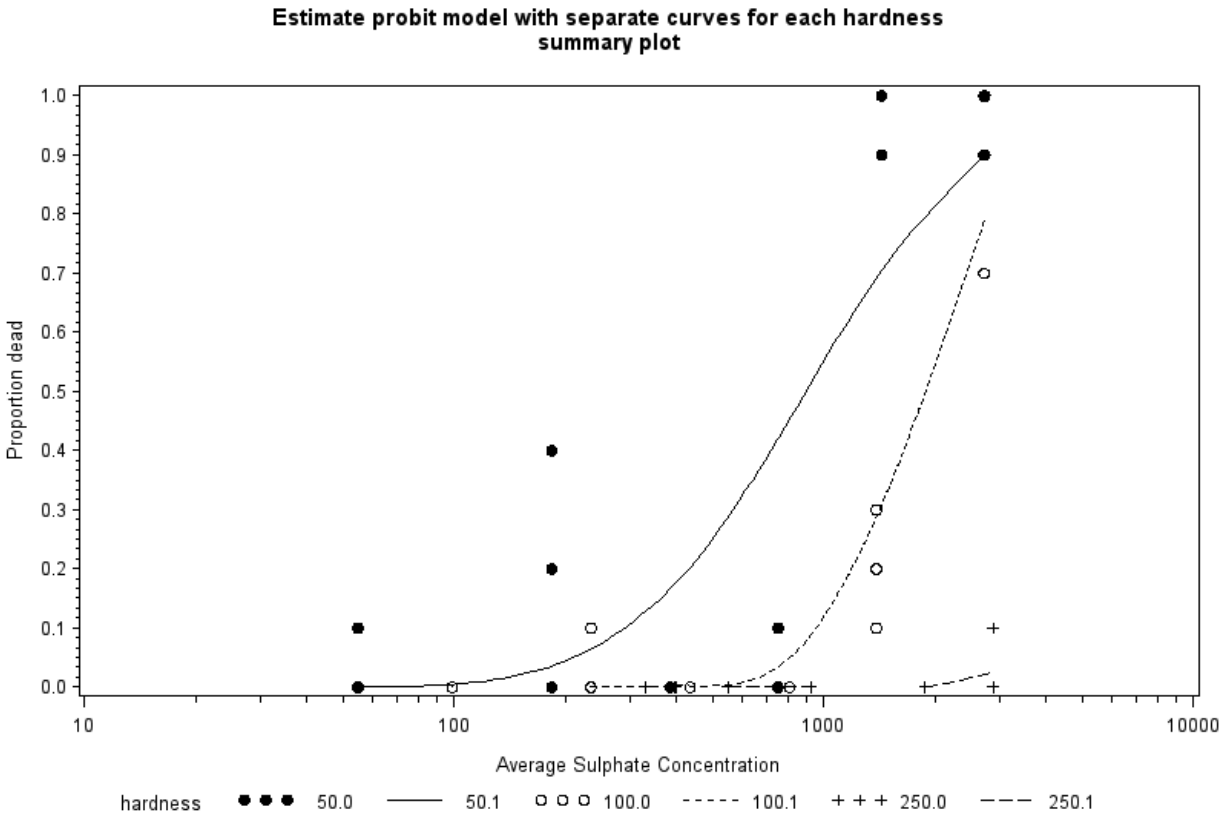

Figure EC-FM-Mort-2 Fitted probit curves for model with separate curve for each hardness for EC fat head minnow mortality trials.

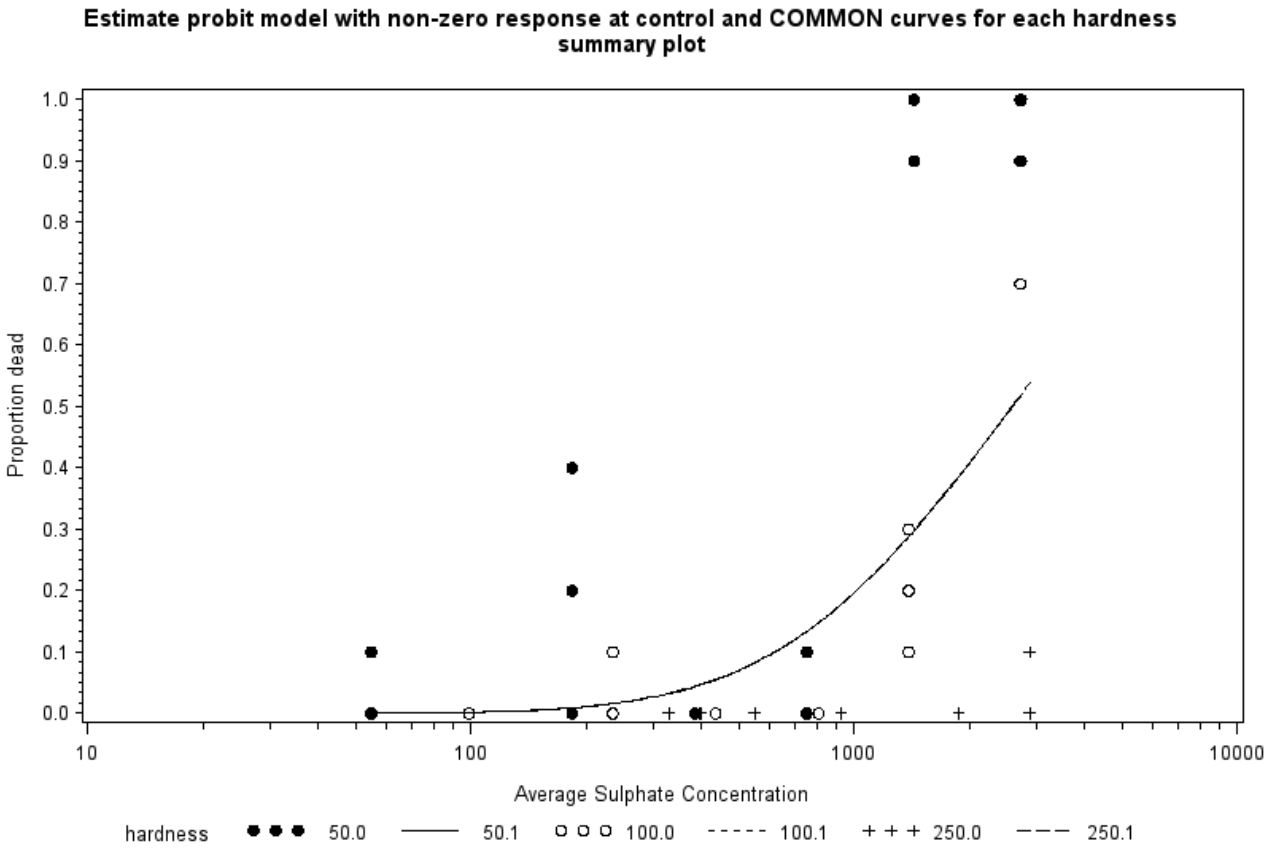

Figure EC-FM-Mort-3. Fitted probit curves for model with common curve for all hardness levels for EC fat head minnow mortality trials.

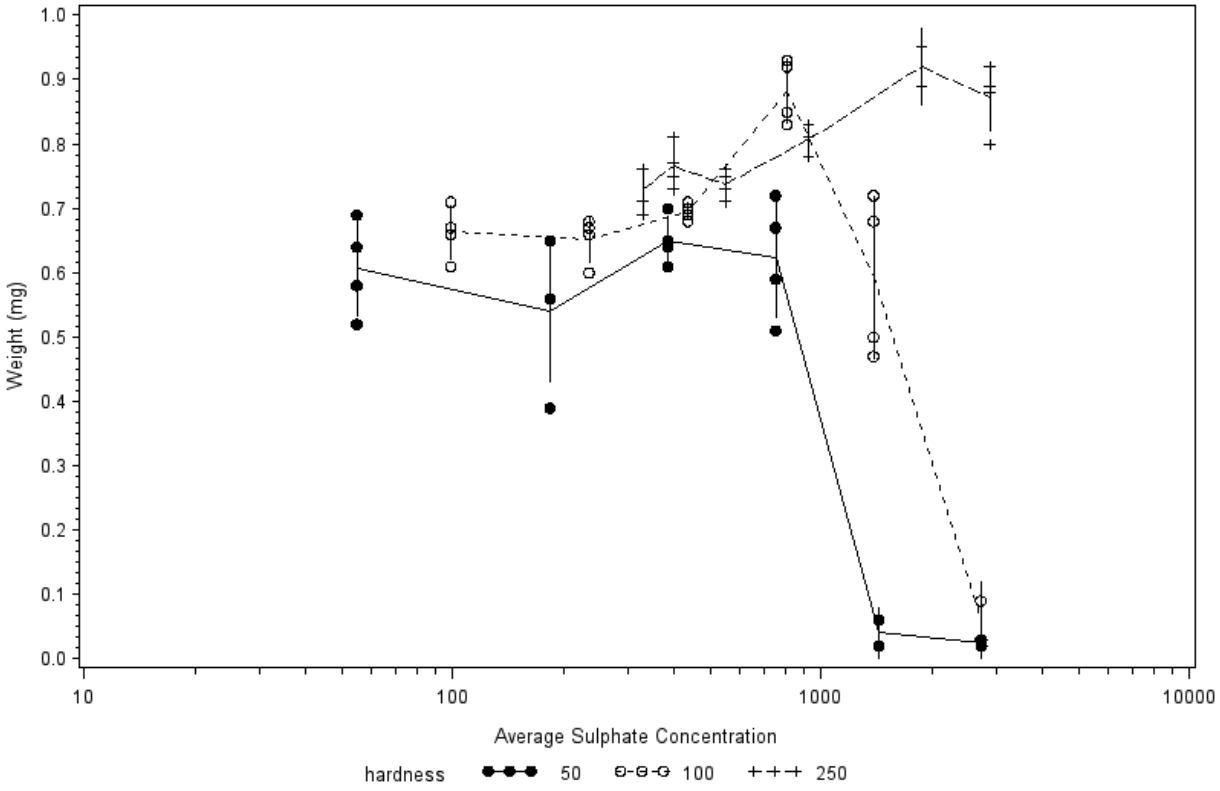

Figure EC-FM-Weight-1. Empirical mean final weight observed in the EC fat head minnow mortality trials.

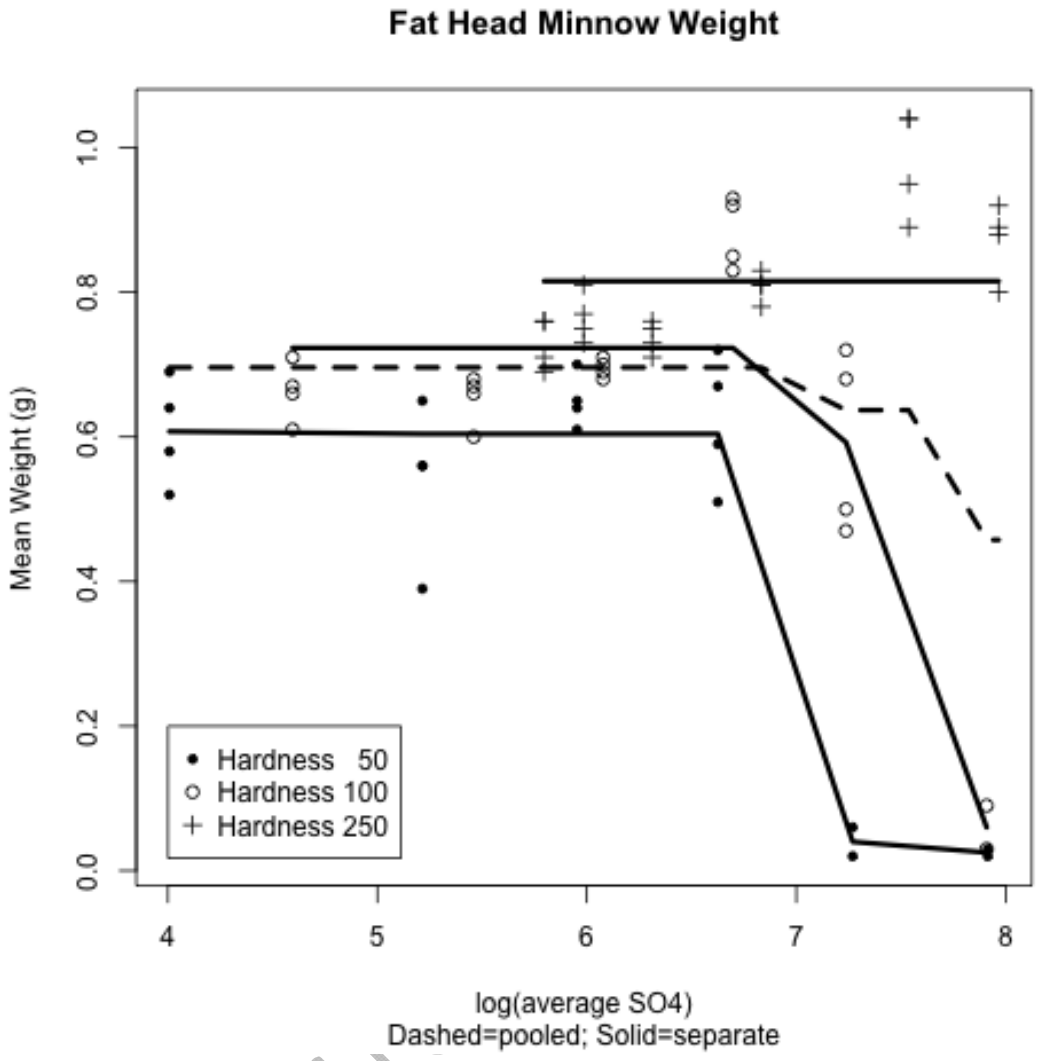

Figure EC-FM-Weight-2. Fitted isotonic curves for the mean weight at the end of the experiment in the EC Fat Head Minnow weight trials

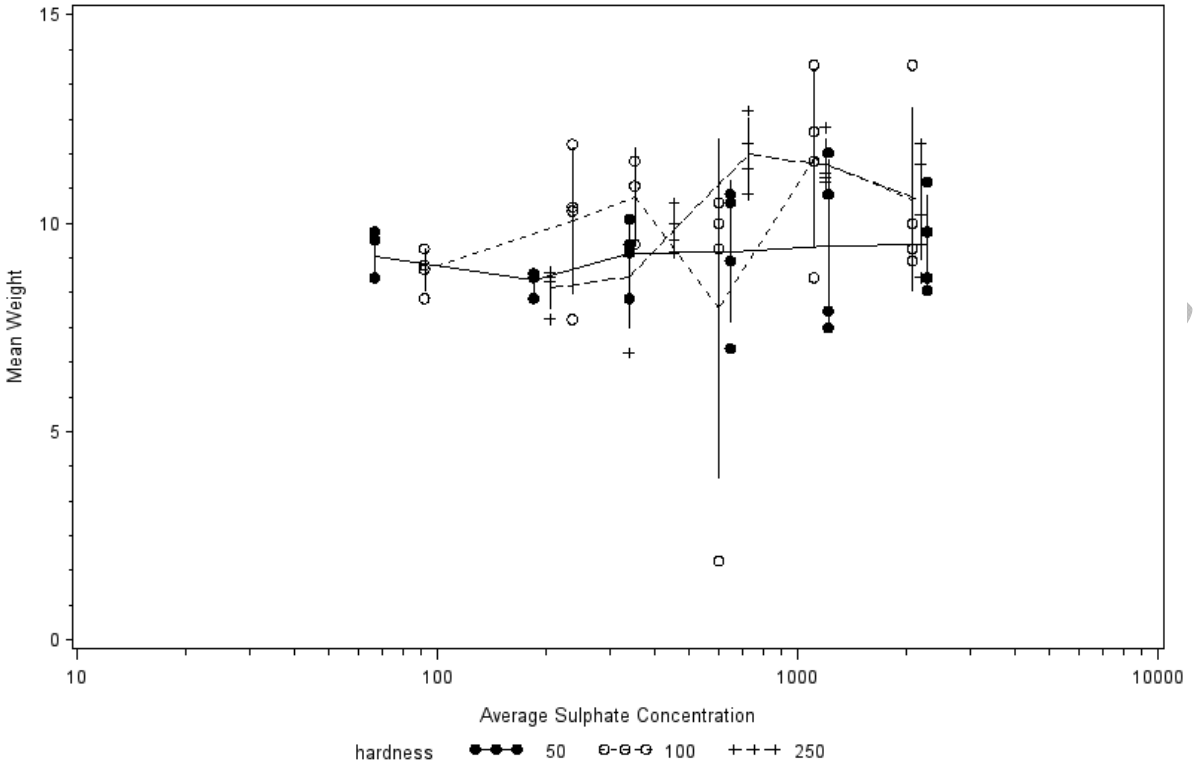

Figure EC-LM-Weight-1. Empirical mean final weight observed in the EC Lemna growth trials.

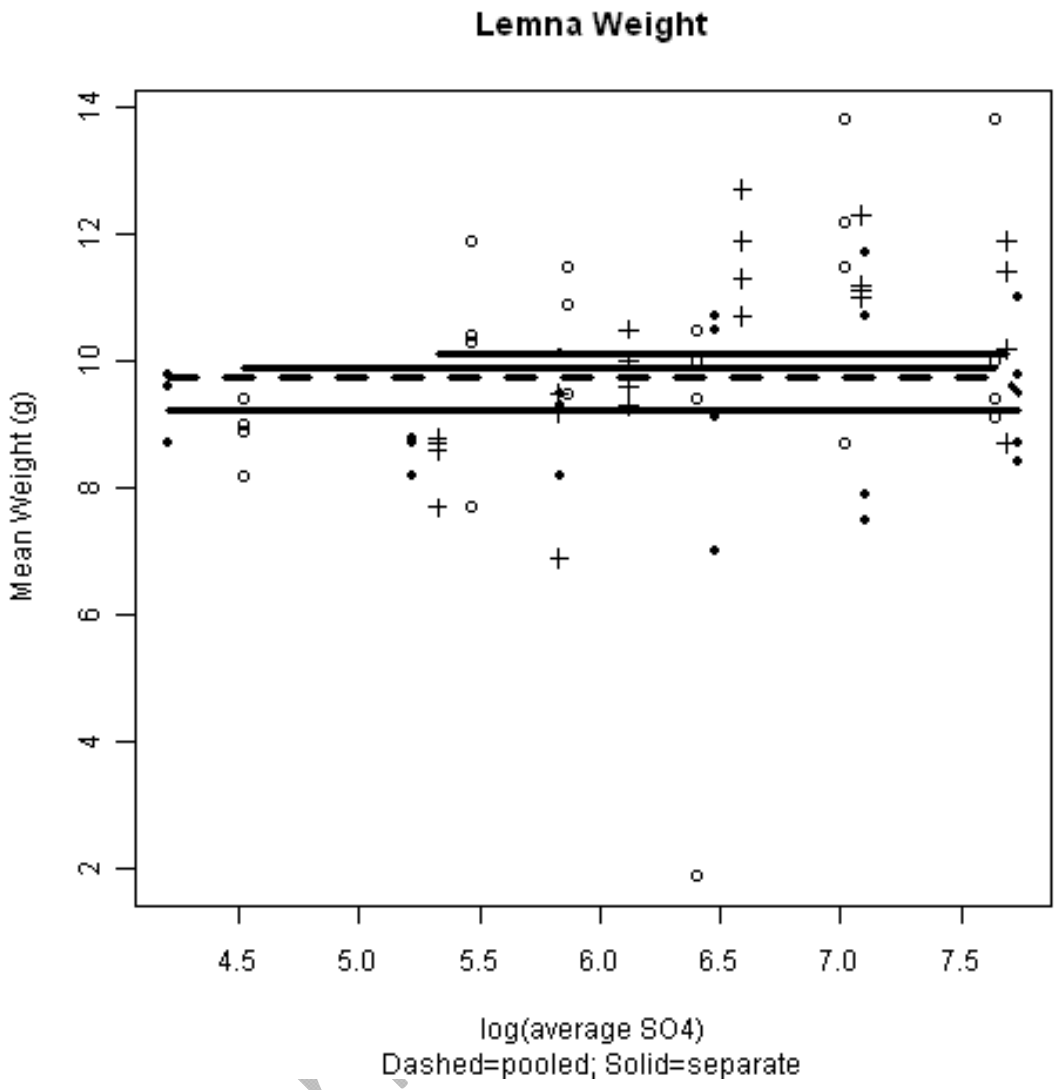

Figure EC-LM-Weight-2.. Fitted isotonic curves for the mean weight at the end of the experiment in the EC Lemna trials.

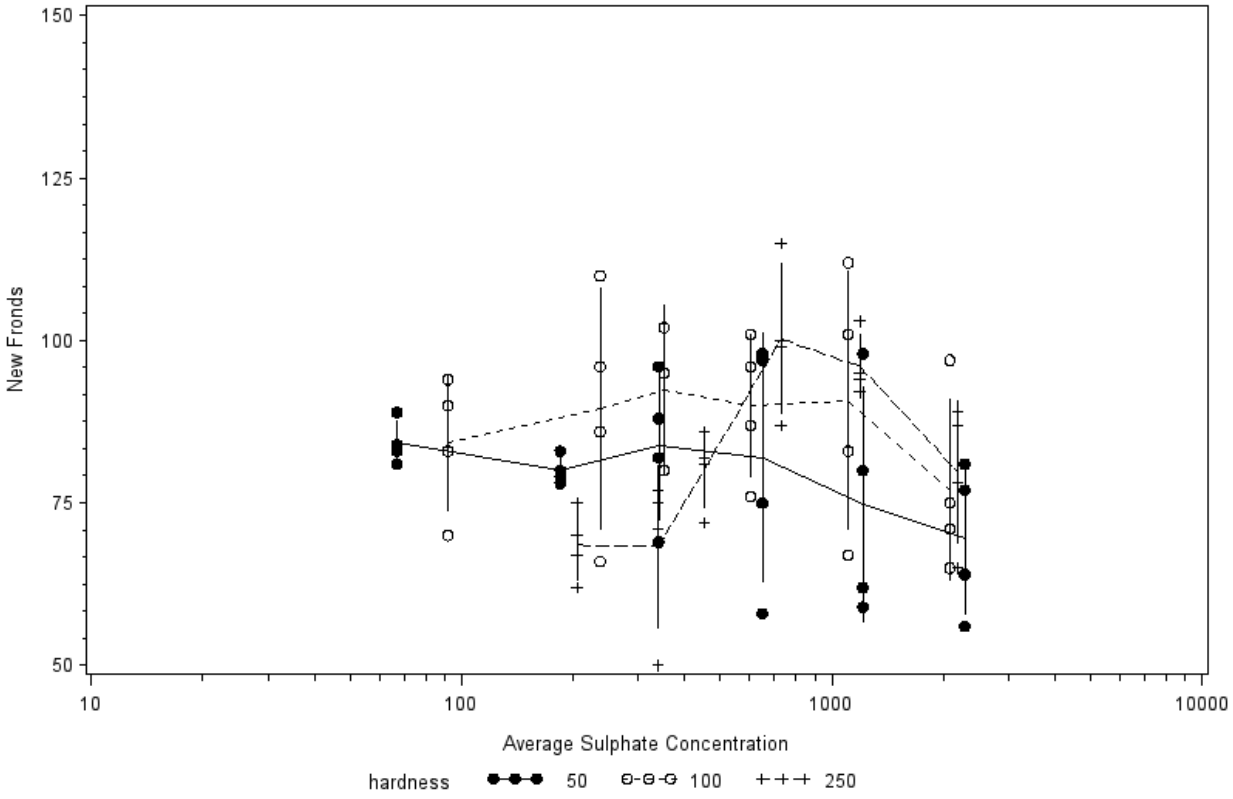

Figure EC-LM-Frond-1. Empirical increase in frond numbers observed in theEC Lemna growth trials.

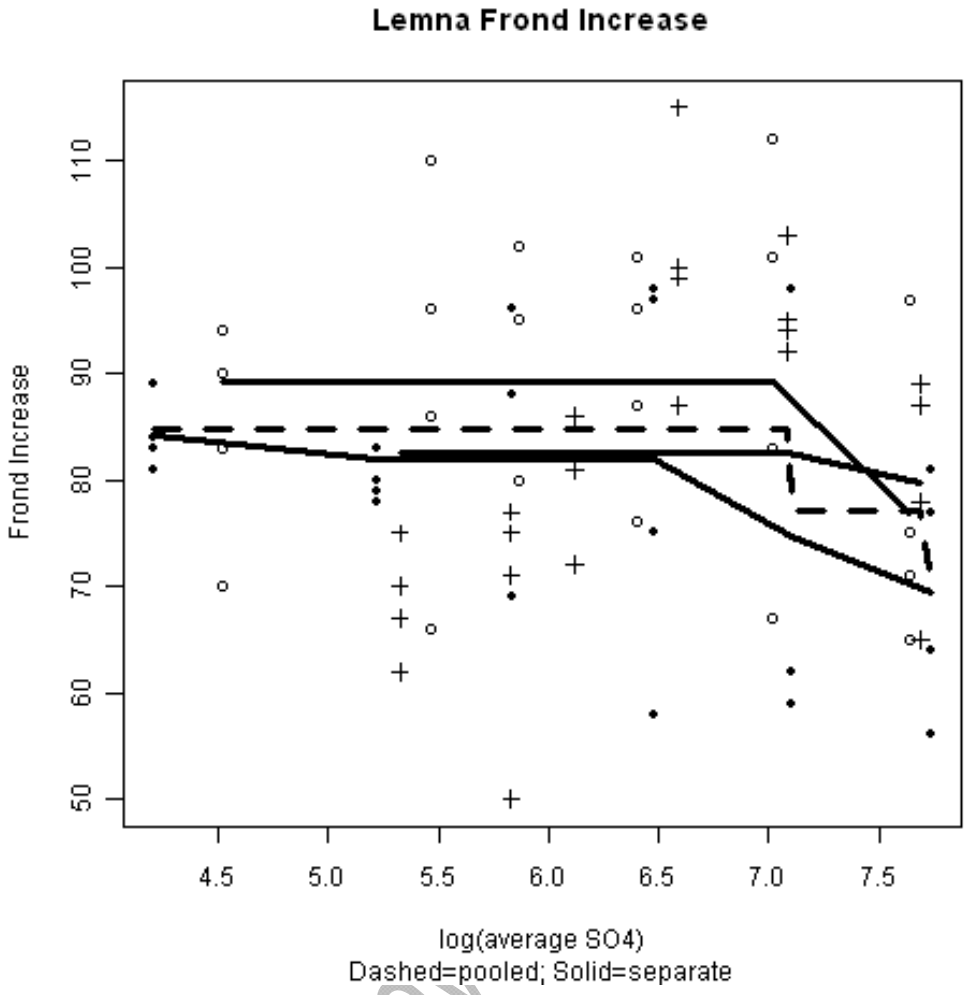

Figure EC-LM-Frond-2. Fitted isotonic curves for the increase in the number of fronds at the end of the experiment in the EC Lemna trials.

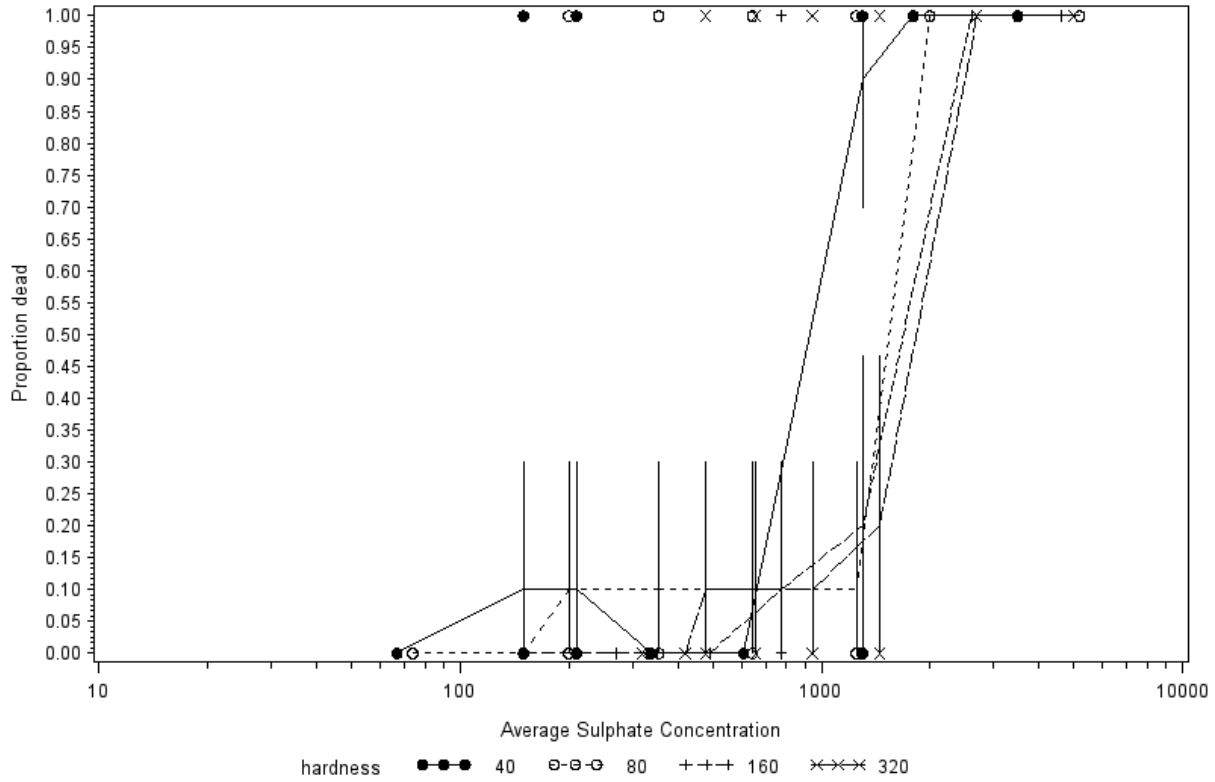

Figure NA-DA-Mort-1. Empirical mortality observed in the NA Daphnia mortality study.

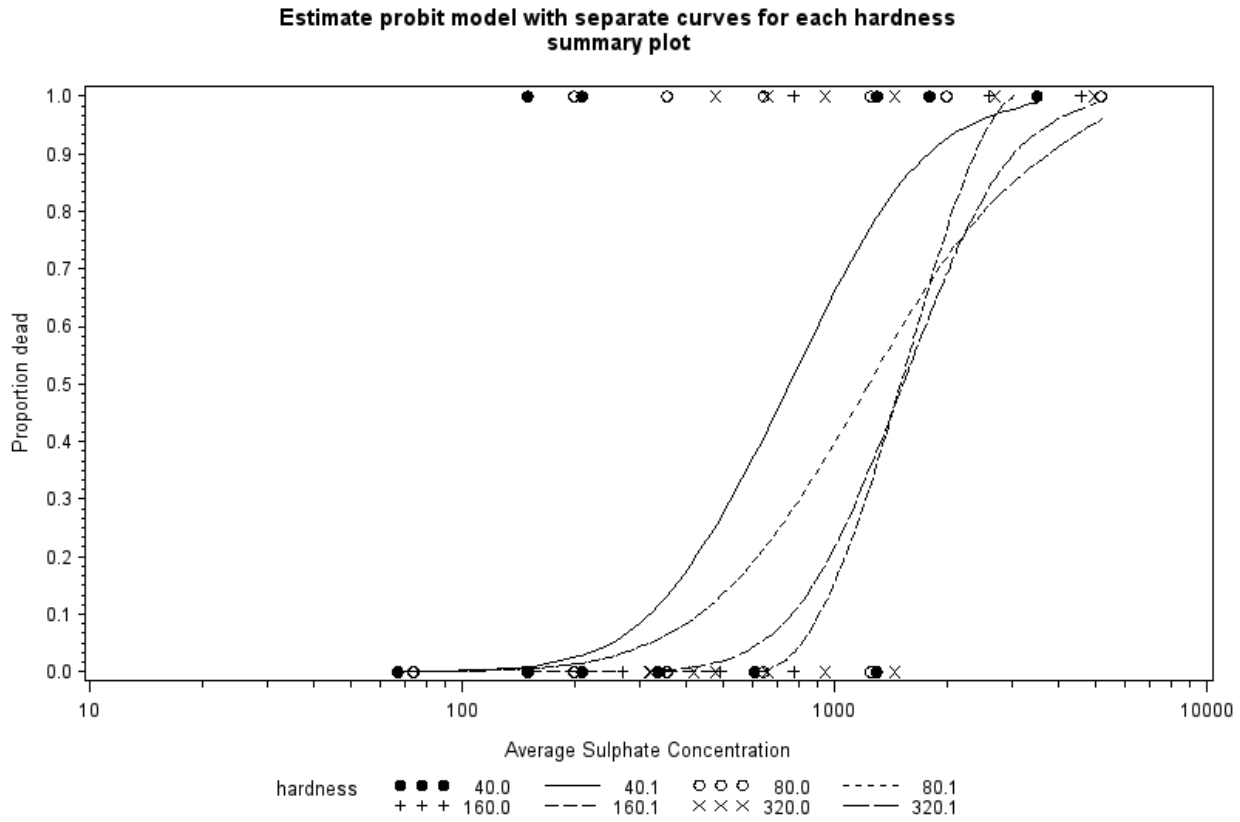

Figure NA-DA-Mort-2. Fitted probit curves for model with separate curve for each hardness for NA Daphnia mortality study.

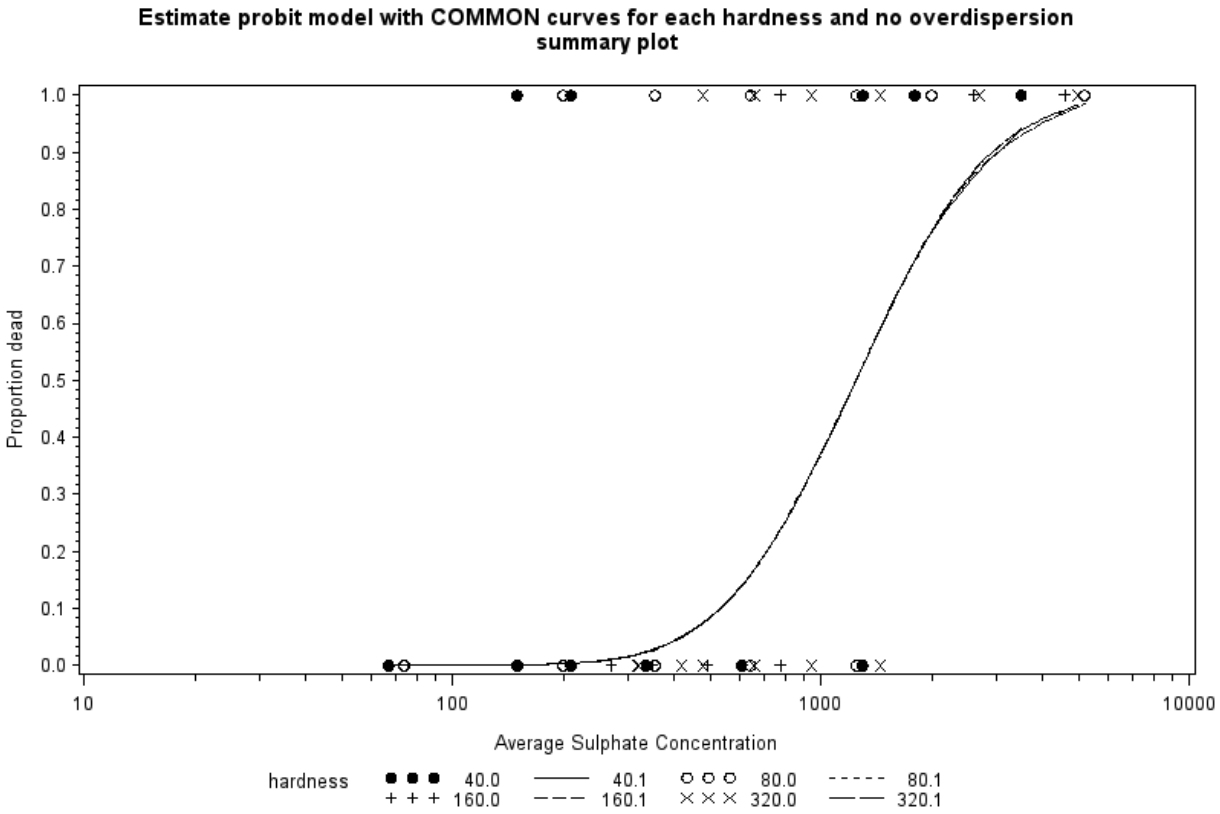

Figure NA-DA-Mort-3. Fitted probit curves for model with common curve for all hardness levels for NA Daphnia mortality study.

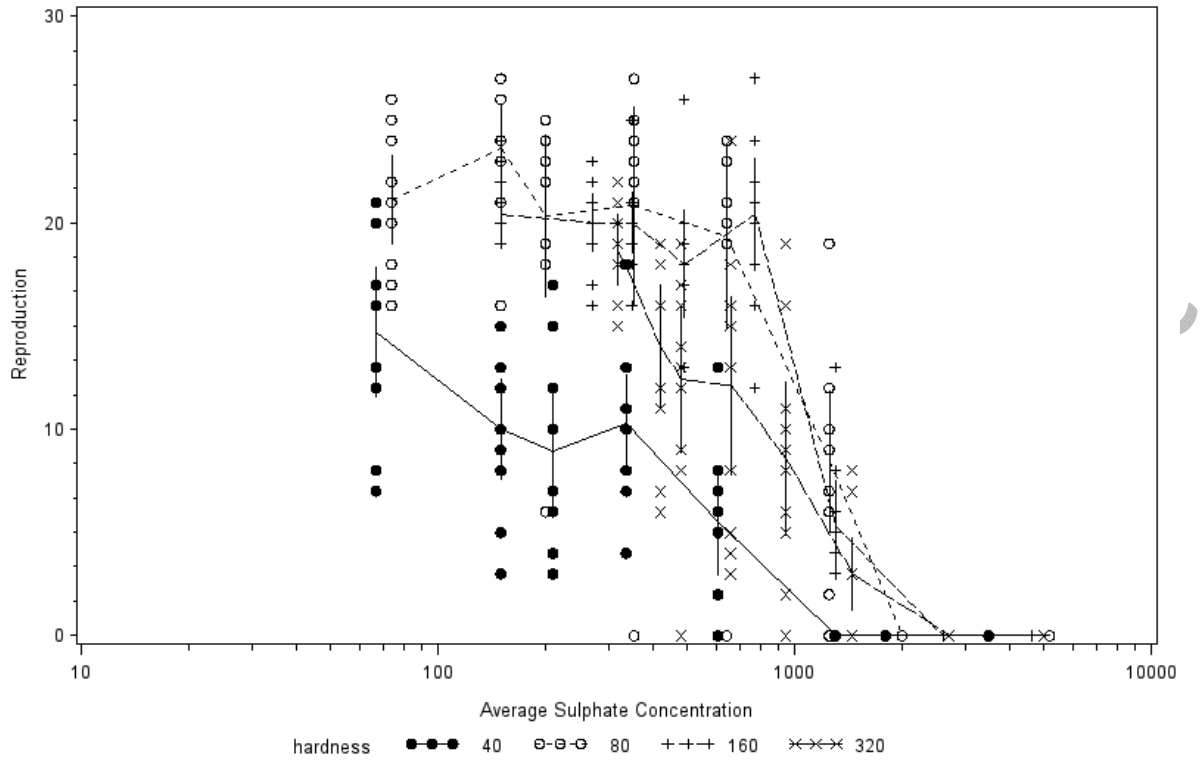

Figure NA-DA-Repro-1. Empirical mean reproductions at the end of the experiment observed in the NA Daphnia trials

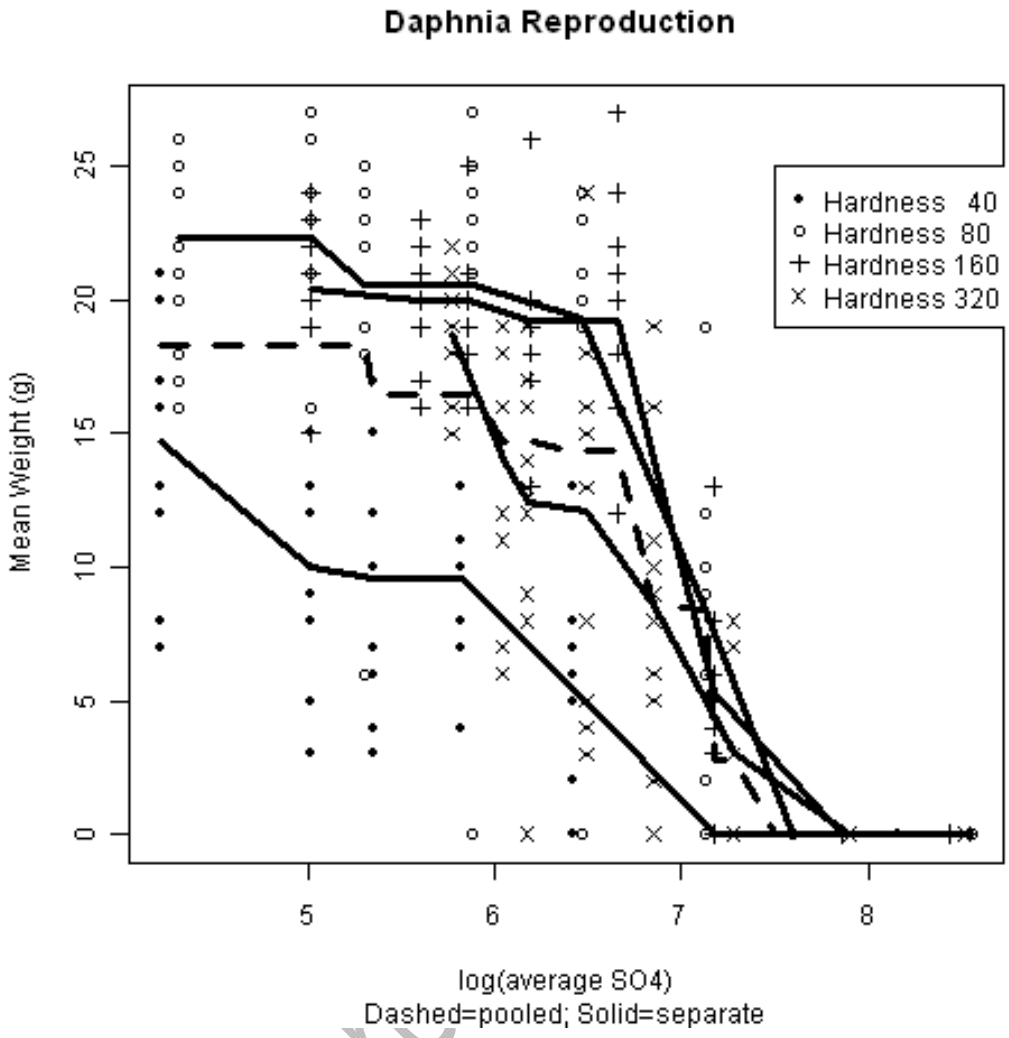

Figure NA-DA-Repro-2. Fitted isotonic curves for the mean reproduction at the end of the experiment in the NA Daphnia trials.

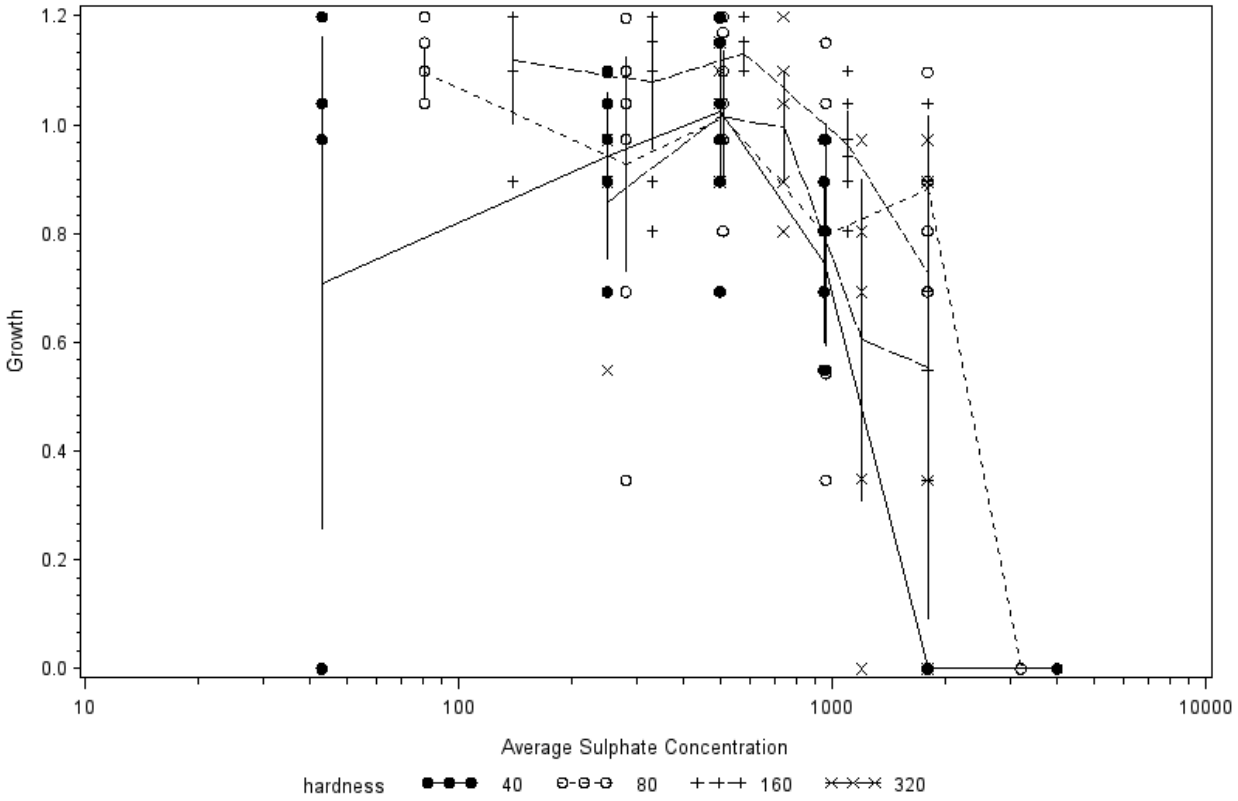

Figure NA-RO-Repro-1. Empirical mean reproductions at the end of the experiment observed in the NA Rotifer trials.

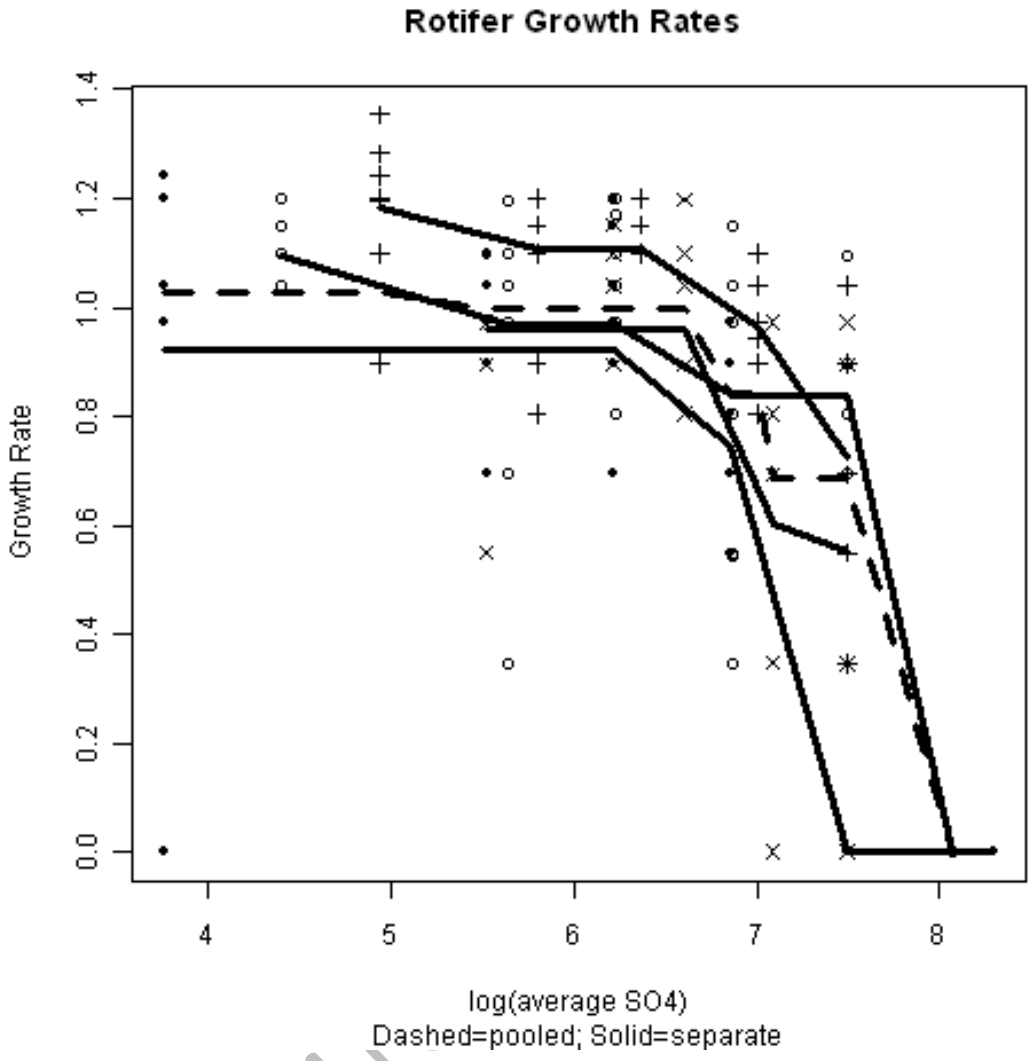

Figure NA-DA-Repro-2. Fitted isotonic curves for the mean reproduction at the end of the experiment in the NA Rotifer trials.

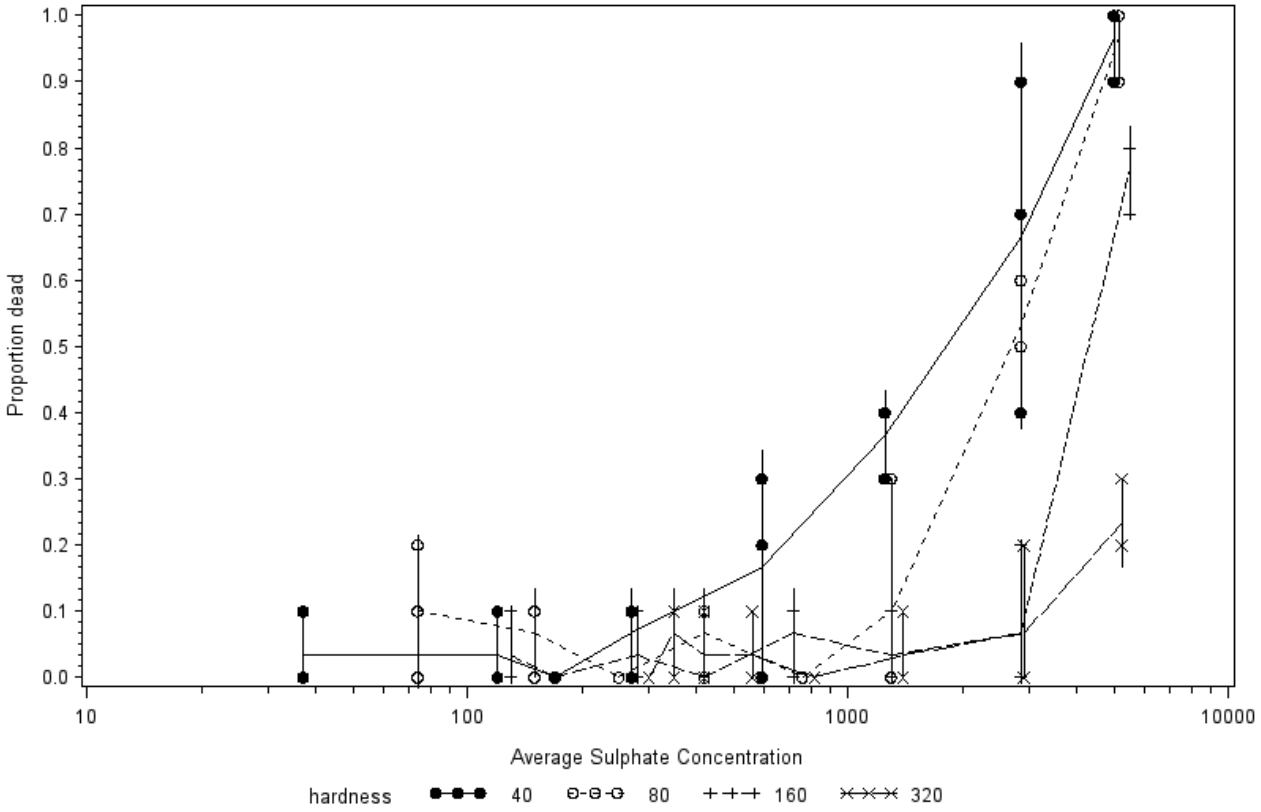

Figure NA-FM-Mort-1. Empirical mortality observed in the NA fathead minnow mortality study.

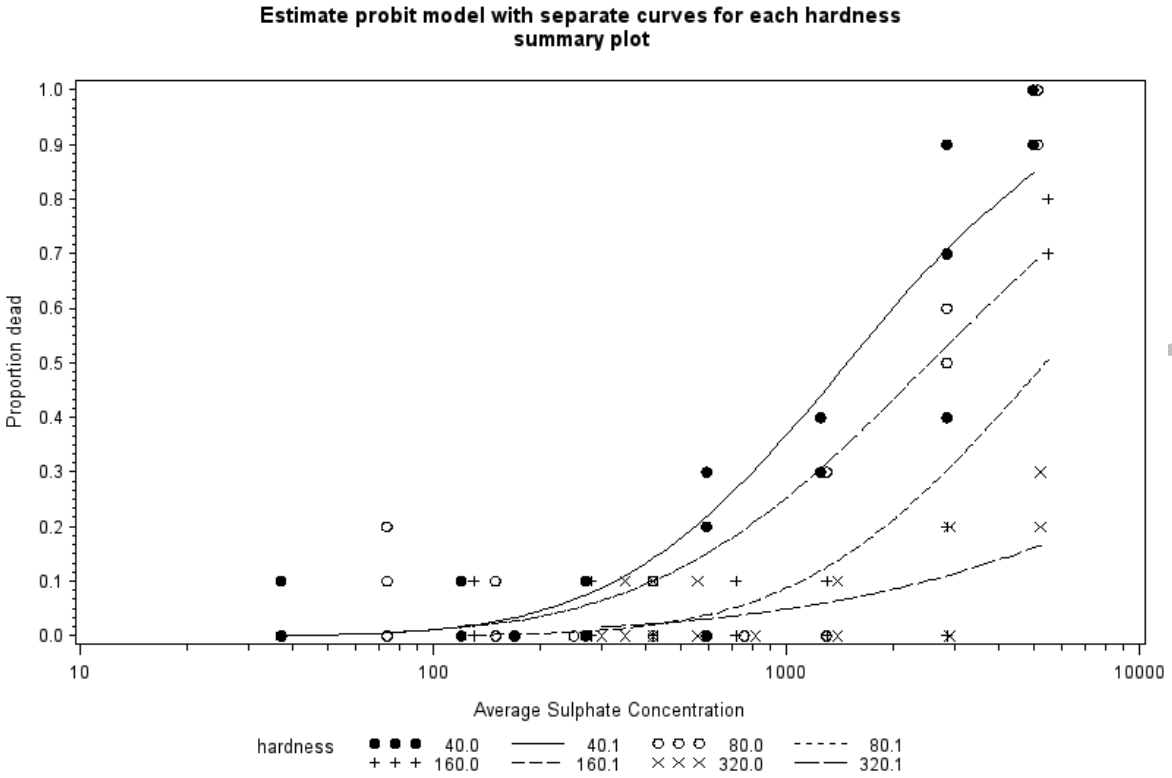

Figure NA-FM-Mort-2. Fitted probit curves for model with separate curve for each hardness for NA Fathead Minnow mortality study.

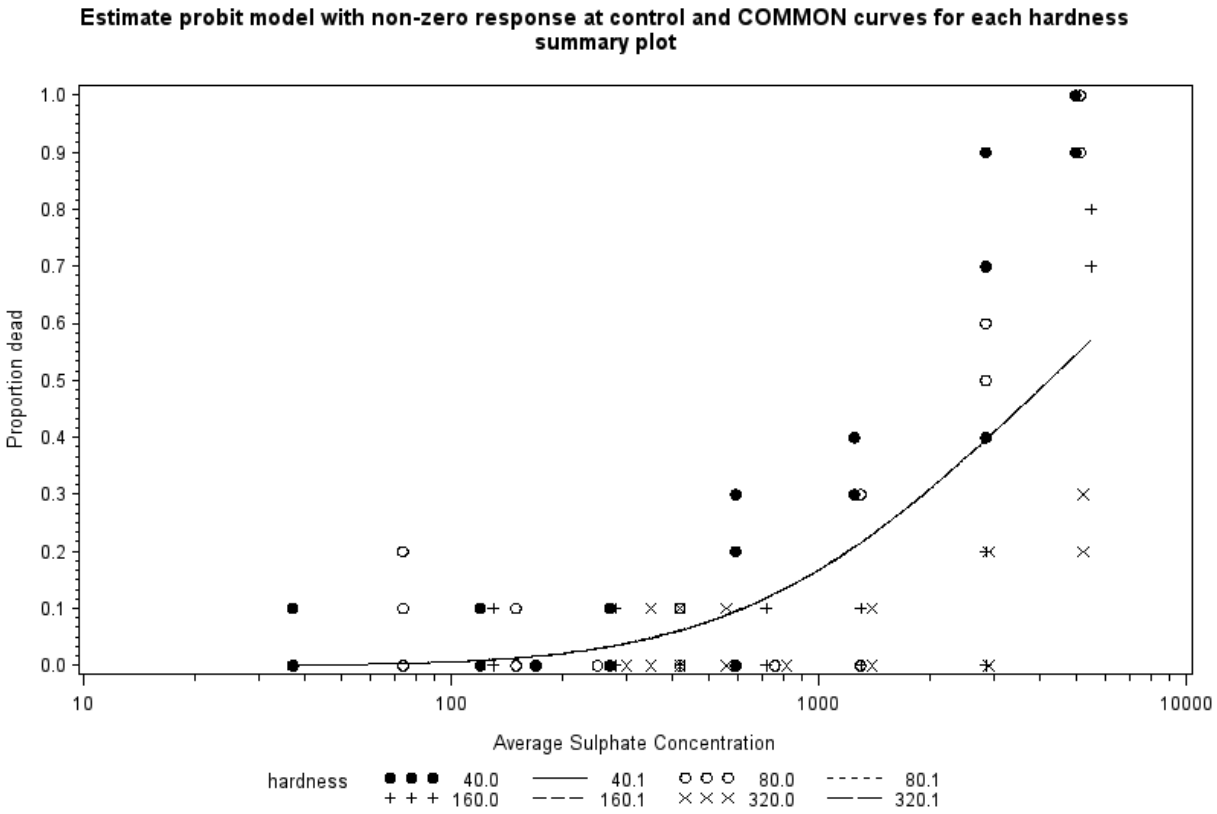

Figure NA-FM-Mort-3. Fitted probit curves for model with common curve for all hardness levels for NA Fathead Minnow mortality study.

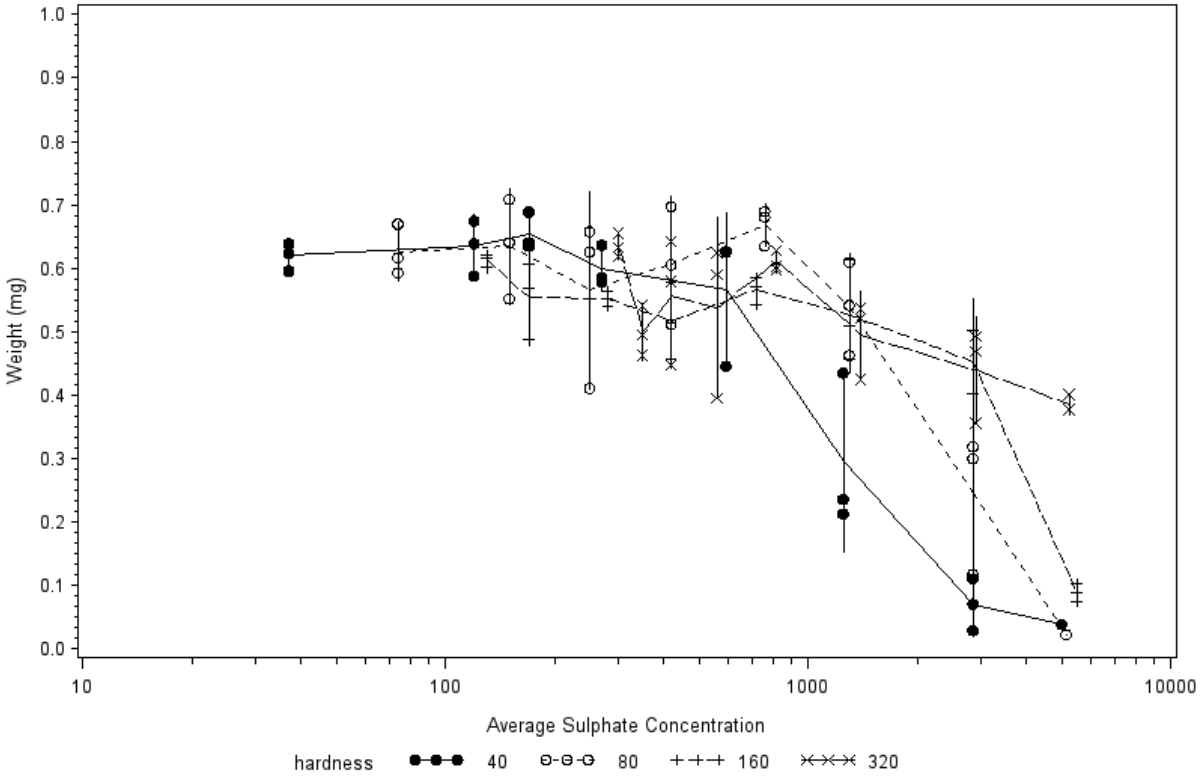

Figure NA-FM-Weight-1. Empirical mean weight at the end of the experiment observed in the NA Fathead Minnow trials.

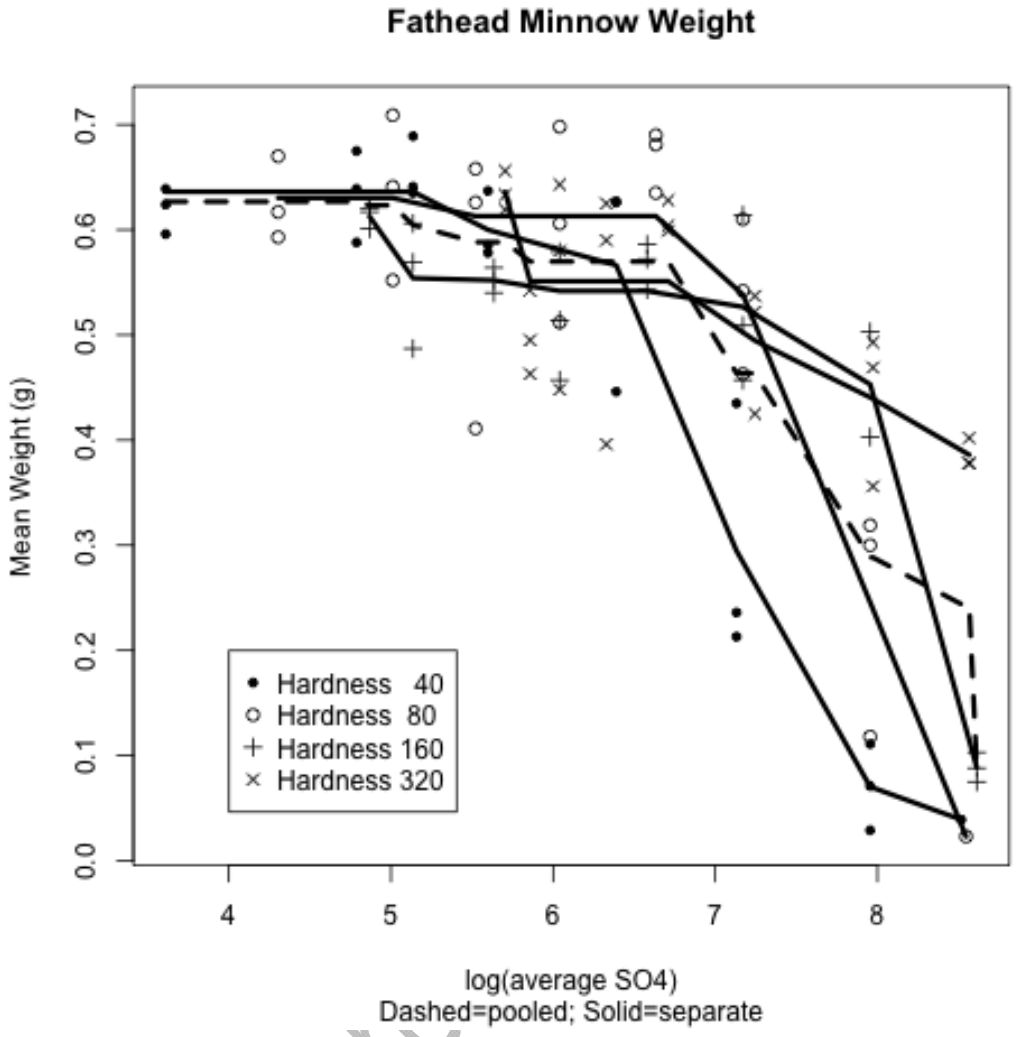

Figure NA-RM-Weight-2. Fitted isotonic curves for the mean weight at the end of the experiment in the NA Fathead Minnow trials.

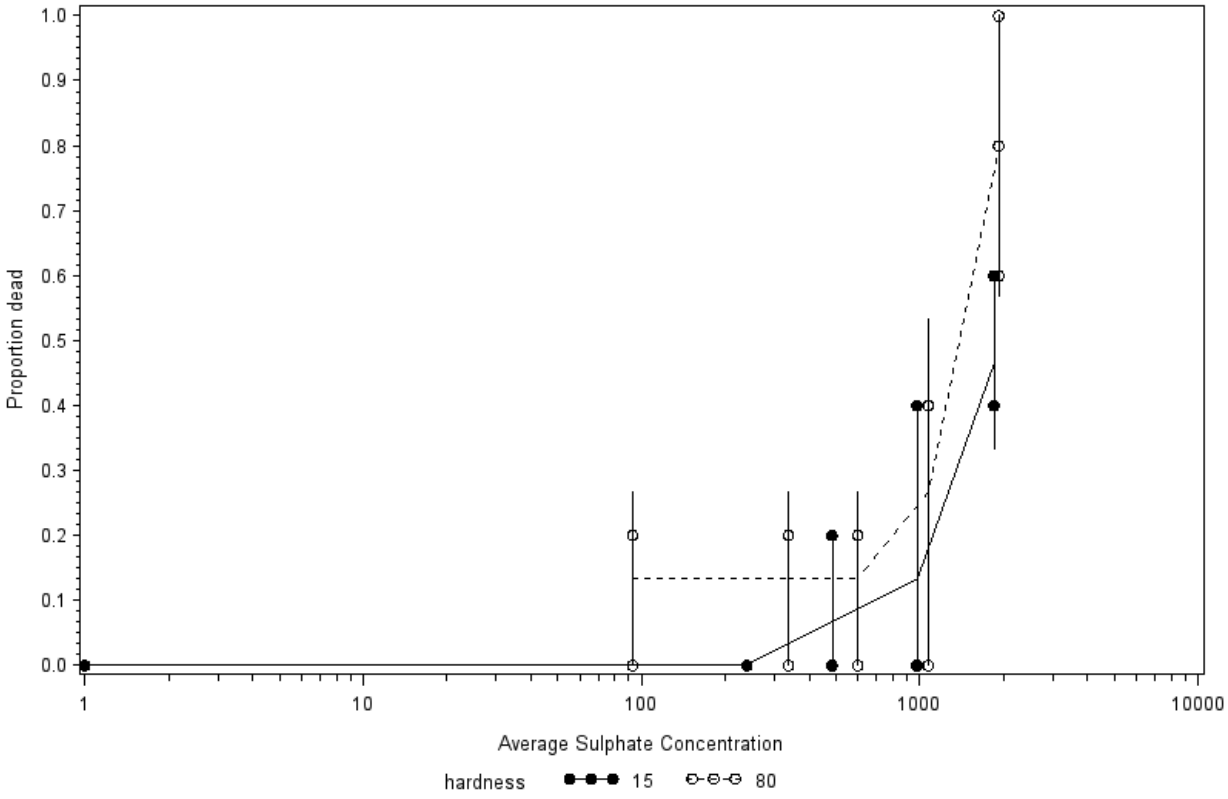

Figure NA-TA-Mort-1. Empirical mortality observed in the NA Tadpole mortality study.

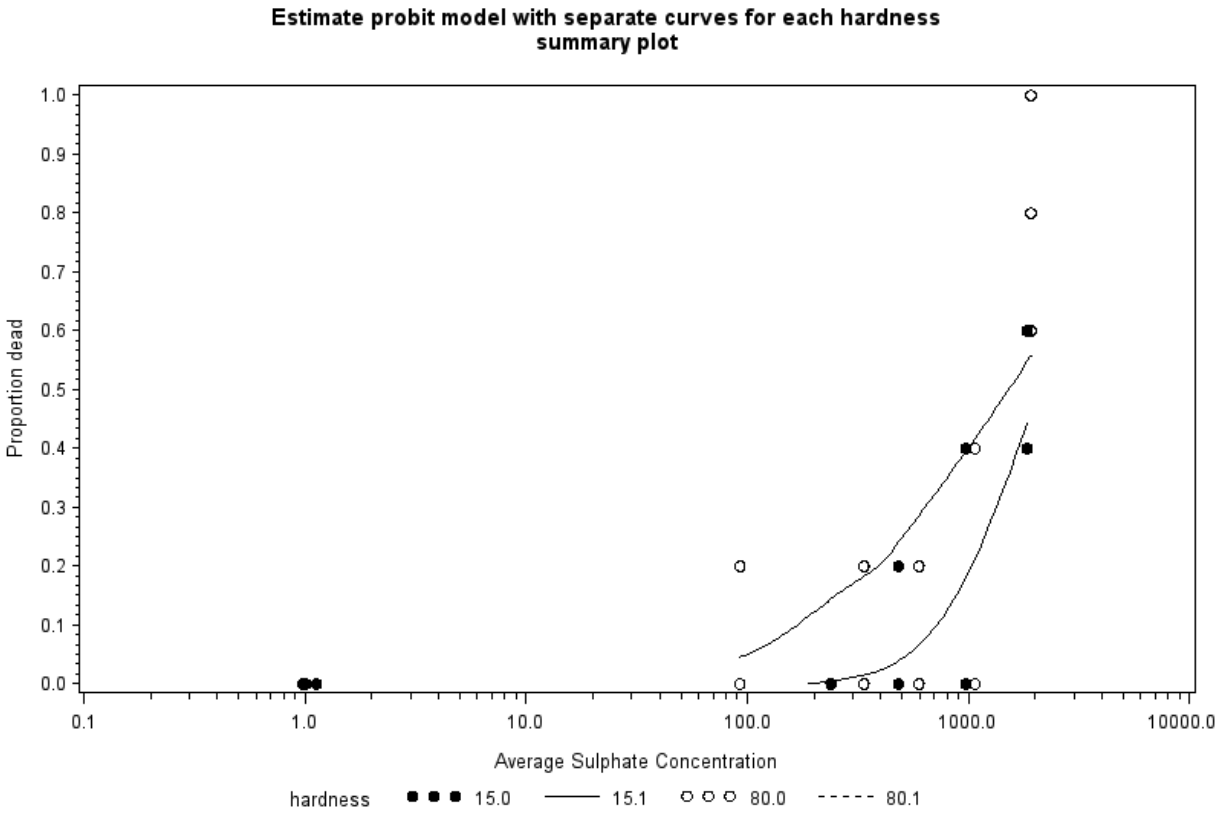

Figure NA-TA-Mort-2. Fitted probit curves for model with separate curve for each hardness for NA Tadpole mortality study.

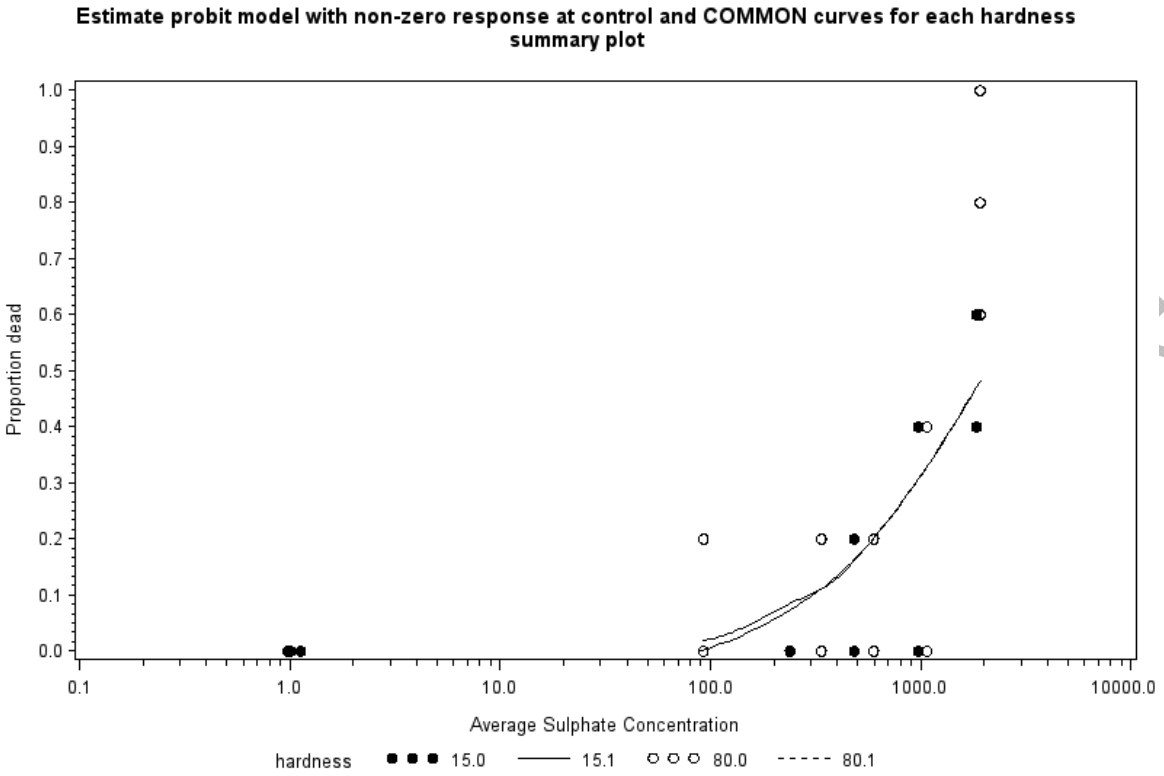

Figure NA-TA-Mort-3. Fitted probit curves for model with common curve for all hardness levels for NA Tadpole mortality study.

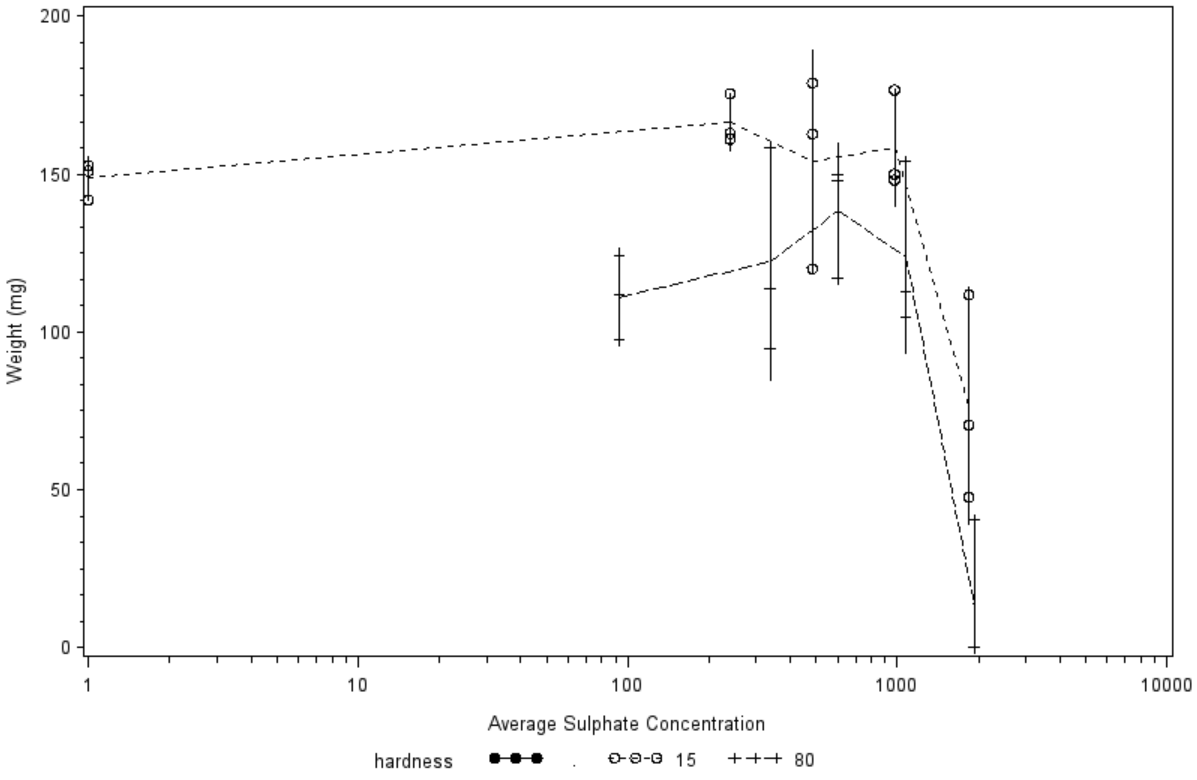

Figure NA-TA-Weight-1. Empirical mean weight at the end of the experiment in the NA Tadpole trials.

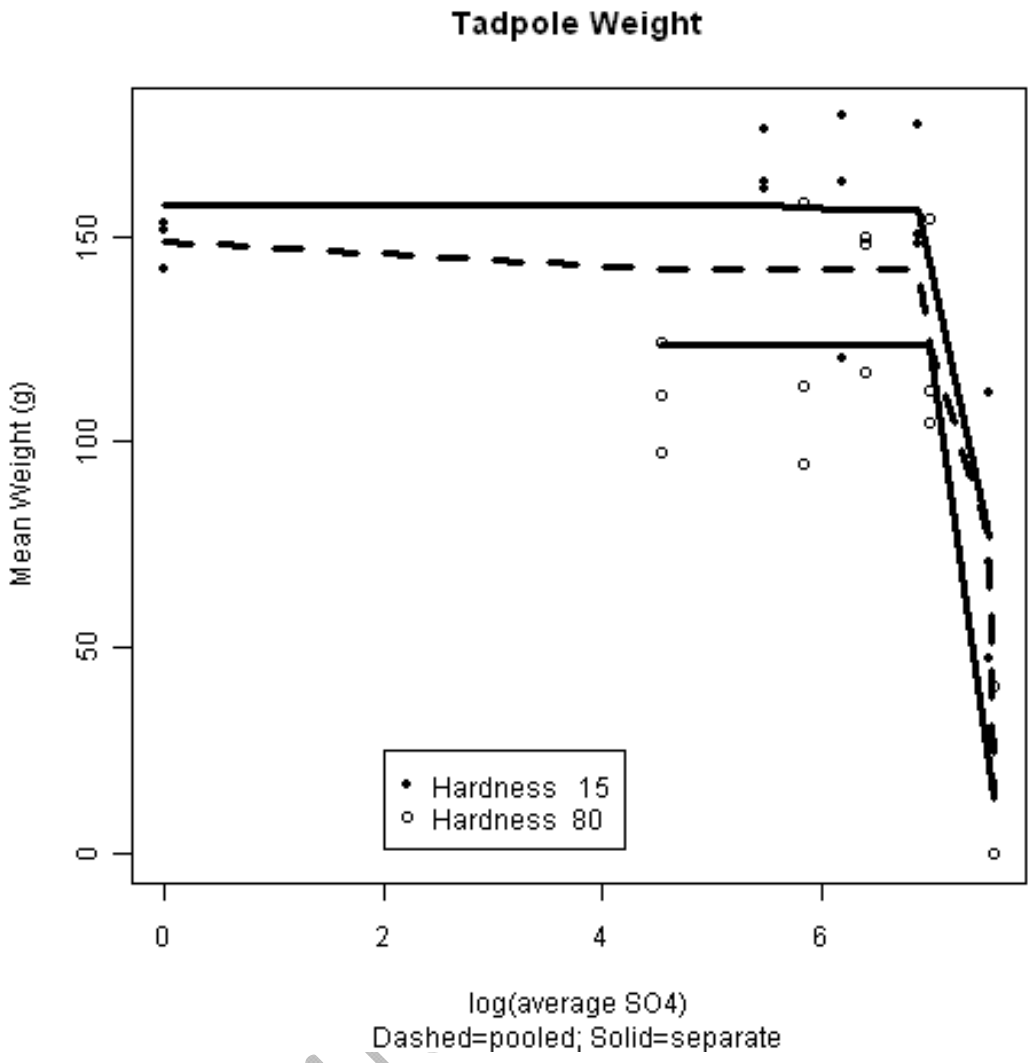

Figure NA-TA-weight-2. Fitted isotonic curves for the mean weight at the end of the experiment in the NA Tadpole trials.

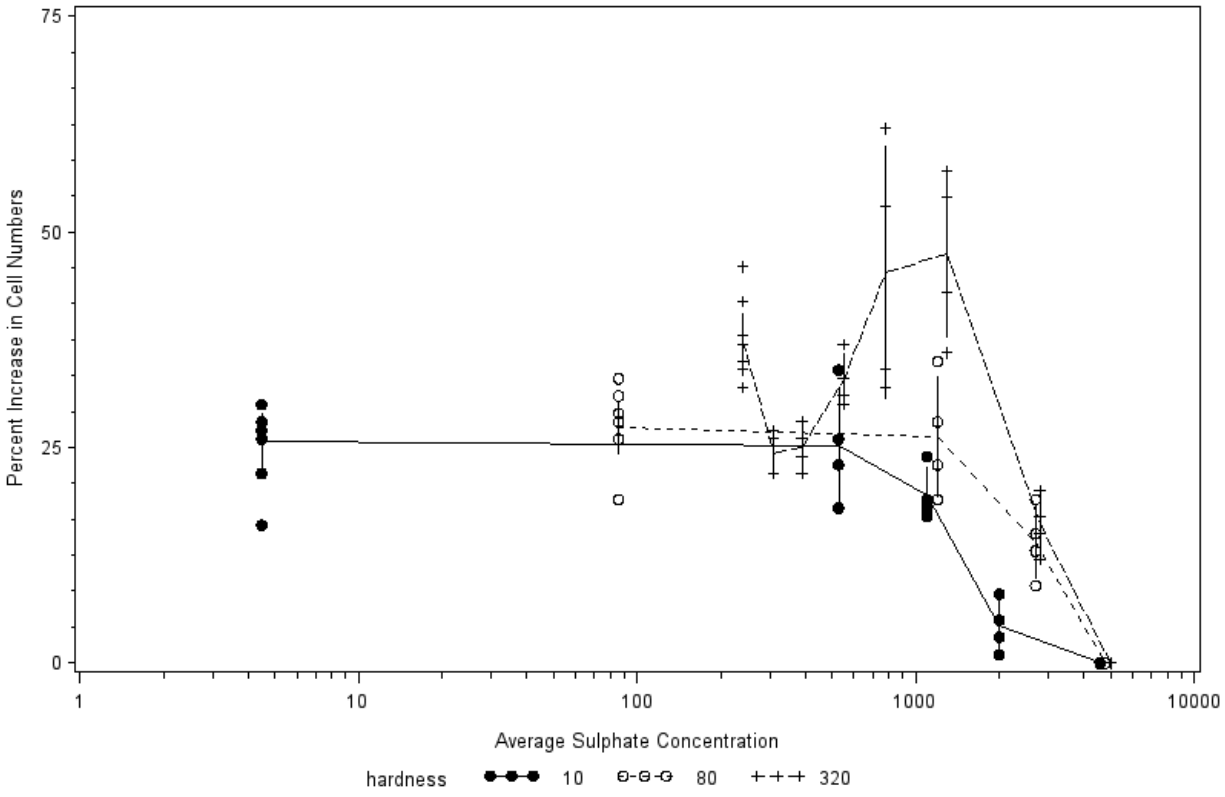

Figure NA-AL-Repro-1. Empirical mean percent increase in cell numbers at the end of the experiment observed in the NA Algae trials.

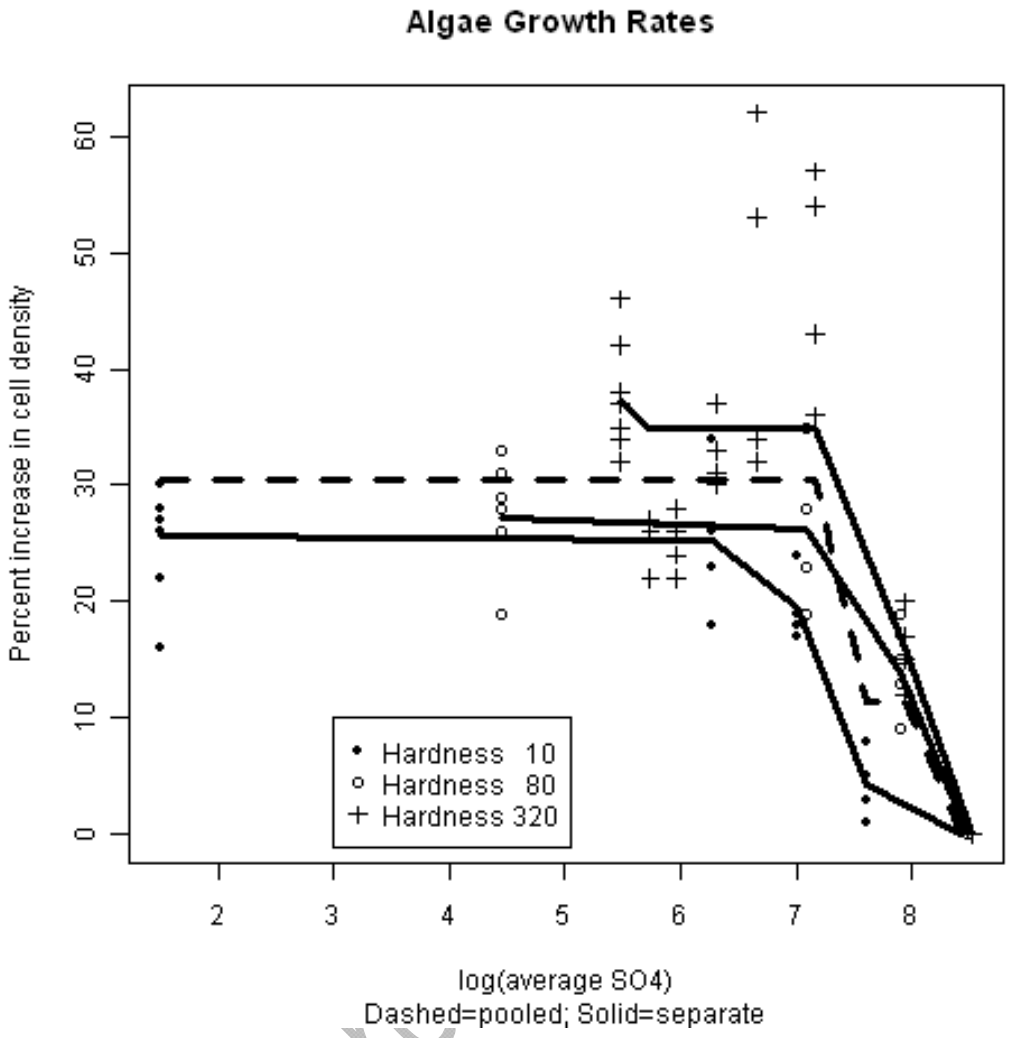

Figure NA-AL-Repro-2. Fitted isotonic curves for the mean percent increase in cell numbers at the end of the experiment in the NA Algae trials.