THE DESIGN AND ANALYSIS OF SALMONID TAGGING STUDIES IN THE COLUMBIA BASIN

VOLUME IX

A Comparison of Statistical Methods of Estimating Treatment-Control Ratios (Transportation Benefit Ratios) Based on Spring Chinook Salmon on the Columbia River, 1986-1988

Prepared by:

Richard L. Townsend
John R. Skalski
Center for Quantitative Science
School of Fisheries
University of Washington
Seattle, Washington

Prepared for:

U.S. Department of Energy
Bonneville Power Administration
Division of Fish and Wildlife
P.O. Box 3621
Portland, OR 97283-3621

Project Number 91-051
Task Number DE-AT79-91BP16570
Contract Number DE-AI79-87BP35885

August 1997

Executive Summary

Objective

The strength of a salmon run is often measured as the adult return rate from some previous brood year (i.e. the percent of a smolt population returning to spawn or captured in fisheries). The U.S. Army Corp of Engineers (ACOE) program of barge transportation of smolts from collector dams is one mitigation measure used to improve smolt survival, and the adult return rate of transported smolt has been tracked, along with the smolt captured, tagged and released back into the river. A ratio of the recovered percentages of adult salmon, those transported in the smolt stage over the salmon not transported (controls), is often used to summarize the program effectiveness. There are a number of ways to estimate this transportation/control (T/C) ratio[1], and this paper explores alternative statistical models to improve accuracy and precision of the estimate.

Results

Assuming the proportion of adult recoveries are binomially distributed, the data were analyzed using linear regression of arc-sine square-root and logit transformations; general linear model regression (GLM) with logit- and log-links; and a maximum-likelihood estimation (MLE) of the T/C ratio. Profile likelihood intervals were calculated to generate 95% confidence interval estimates of the T/C ratio. Depending on the analytical method, T/C ratios varied greatly. Arc-sine square-root and logit transformations gave individual release T/C ratios which ranged from 1.0934 to 4.0076 and -1.2193 to 1.9057, respectively. The negative T/C ratio is due to the back-transformation properties of the logit transformation. The GLM and MLE approaches produced mean T/C ratios (after adjusting for the individual release batch effects) ranging from 1.4964 to 1.4974. The recommended method from this analysis, a binomial maximum likelihood estimate adjusted for over-dispersion, produced a T/C ratio of 1.4965 with a 95% confidence interval of (1.0618, 1.9312). This means that yearling chinook salmon smolt transported from McNary Dam to a release site below Bonneville Dam survived at rate approximately 1.5 times greater than the control smolt permitted to out-migrate in-river.

Management Implications

Point estimates for transportation effect on adult survival were initially estimated from these site recoveries, but are likely inappropriate, as the assumption that an equal proportion of fish from these recovery sites were released as treatment and control smolt can not be confirmed. In this paper, the total adult recoveries across all of the recovery sites for each group of releases were analyzed, yielding a transportation effect estimate for the site of capture, McNary Dam, only. The transportation effect was assumed to be constant throughout the season and across years, yielding one number (the T/C ratio) to summarize the three years of the program Until better tracking techniques or the ability to identify the originating source of smolt are available, this is the limit of determing the effectiveness of programs of this type.

1. This is also known as the Transportation Benefit Ratio (TBR).

Recommendations

The recommended method from this analysis is a binomial maximum likelihood estimate, adjusted for over-dispersion, because of the ability to isolate the treatment effect of transportation better than the alternative statistical models. A close second in recommendation is the binomial log-link general linear regression model, which can be implemented by many statistical packages using existing software.

.

Table of Contents

List of Figures

List of Tables

Acknowledgments

We wish to express thanks to the National Marine Fisheries Service that has expended considerable resources in the generation, assembly, analysis and sharing of the Columbia River tranportation information. Deserving particular thanks is Ben Sanford of NMFS for answering our many questions about the transportation program and providing updated adult recovery data. I would also like to thank the reviewers of this paper, Drs. Lyle D. Calvin, Al Giorgi, Ken Newman and Steve Smith, for their input in presenting a more accurate and easier-to-understand paper.

Funding support for this work came from the Pacific Northwest region's electrical rate-payers through the Columbia River Fish and Wildlife Program administered by the Bonneville Power Administration.

Introduction

Transporting salmon smolt down the Columbia River bypasses a number of hazards and is believed to improve the survival rate. To evaluate this assumption, from 1986 through 1988, the National Marine Fisheries Service (NMFS) marked spring/summer chinook salmon smolt (Oncorhynchus tshawyscha) (Achord et al. 1992; Harmon et al. 1989, 1993; Matthews et al. 1987, 88, 90, 92) to evaluate the effectiveness of the U.S. Army Corp of Engineers (ACOE) program of barge transportation of smolts from collector dams. Outmigrating salmon smolt were captured, marked and then either released at the site of capture (McNary Dam) or transported by barge down the Columbia River to an area below Bonneville Dam. Freeze-brands were used to readily identify the fish visually and the adipose fin was removed to indicate a coded-wire tag (CWT) had been inserted into the nasal cavity of the fish. Adult recoveries from the returning salmon involved in the experiment were recorded for a number of sites: river system traps at the Bonneville, Lower Granite and Priest River dams; ocean and Indian fisheries; Rapid River, Tucannon and Leavenworth hatcheries; and stream surveys.

A ratio of the percentage of treated (transported) returning adult salmon to untreated (control) returning adult salmon (T/C ratio) is often used to determine the success of experiments such as this. The purpose of this paper is to show how alternative statistical analyses may influence the estimation of the T/C ratio and to recommend the most appropriate approach to obtain reliable results. The methods included are linear regression of arc-sine square-root and logit transformations of the proportion of adult recoveries from each batch; general linear model regressions using the logit- and log-links; and maximum-likelihood estimation of the T/C ratio.

This paper is a cursory exploration into alternative methods of calculating the T/C ratio and to improve estimation of the confidence interval, which may aid in future experimental design. Six statistical methods are presented and a comparison of the results is provided in the Results section. The analyses used blocking by release batches to reduce random variation introduced through time by unaccounted factors (weather, fish condition, time of year, etc.), and assumed that the transportation effect was constant throughout the experiment. The analysis weighted each regression by release size, as release numbers differed considerably throughout the program.

Description of Data

The evaluation of the transportation program occurred in 1986, 1987 and 1988. Ten batches[1] of fish were marked and released in 1986 and 1988, and 8 batches were released in 1987 (Table A1). The date of each batch release was recorded, along with the numbers released (approximately 5000 for each group, 10,000 total per batch). Salmon recovered as adults were tracked through 1992. Recovery sites were grouped into several categories: traps in fish ladders at Bonneville, Lower Granite and Priest Rapids Dams; ocean, commercial and Indian fisheries; hatcheries; and sport fishing and stream surveys. Though the site where an adult salmon was recovered is available, the effect of transporting smolt *originating* from that site can not be deter-

1. A batch consists of one experimental group and its control group.

mined. The problem lies in that the origin of the smolt tagged at McNary is unknown. Example 1 demonstrates how the estimate of the transportation-control (T/C) ratio can be miscalculated, should the assumption of origin of the salmon recovered at each site be wrong. Because of this, only the total returns summed over all recovery sites will be used. Although this level of pooling will not give the T/C ratio for a specific recovery site, it will give the over-all effectiveness of the transportation of smolt at the point of capture, McNary Dam.

Example 1: How the point estimate of the T/C ratio depends on the smolt-origin assumption.

A single release of smolt tagged at McNary Dam consists of 2200 smolt, divided evenly between transport and control. Eight-hundred of those originated from a hatchery, 1400 more from other sources. Using simulated adult return data from Table 1, the T/C ratio for a hatchery would be estimated by taking the proportion of adult recaptures of the transported smolt over the proportion of adult recaptures of the control group which returned to the hatchery. The estimate for this release would be 1.28 $\left(\frac{32/1100}{25/1100}\right)$. If origin were known for every adult recapture, the true T/C would be calculated by using the total adult numbers from only the hatchery across all recovery sites, a value of 0.89 $\left(\frac{70/500}{47/300}\right)$! It is because of the possible disparity between estimated and actual T/C ratios that the total adult returns are used to calculate a T/C ratio ($\frac{125/1100}{107/1100} = 1.17$ in this example), estimating the effect of transportation on smolt that are captured and marked at McNary Dam.

Table 1: Example data to show how transport-control ratios can be miscalculated.

	Smolt Origin Sites					
	hatchery		other		total	
	transport	control	transport	control	transport	control
Released	500	300	600	800	1100	1100
Adult Recovery Sites						
hatchery	25	15	7	10	32	25
Indian fishery	10	8	25	22	35	30
spawning grounds	5	4	10	12	15	16
dam traps	30	20	13	16	43	36
total	70	47	55	60	125	107

Statistical Methods

Transportation-Control Ratio

The transportation-control (T/C) ratio is the returning percentage of adult salmon transported as smolt to the returning percentage of adult salmon released as controls. This is a convenient summarization of whether transportation helped or hindered smolt-to-adult survival. In the 1986-88 ACOE transportation studies (Achord et al. 1992; Harmon et al. 1989, 1993; Matthews et al. 1987, 88, 90, 92), two methods of calculating the T/C ratio and a 95% confidence interval were used. A composite T/C ratio was calculated using a log transformation of the ratio and a theoretical variance of:

$$\ln\left(\frac{n_t / N_t}{n_c / N_c}\right) \pm 1.96\sqrt{\frac{1}{n_t} + \frac{1}{n_c} - \frac{1}{N_t} - \frac{1}{N_c}} \qquad (1)$$

where n_t = number of recovered adult salmon transported as smolt,
n_c = number of recovered adult salmon used as controls,
N_t = total number of smolt transported, and
N_c = total number of smolt used as controls.

This formula (1) was derived from the variance estimation of the untransformed T/C ratio put forth by Burnam et al. (1987). Another T/C ratio was estimated by averaging the log of the T/C ratio across all releases, and an empirical variance of:

$$\ln\left(\frac{T}{C}\right) \pm t_{0.05}^{n-1}\, SE\left(\ln\left(\frac{T}{C}\right)\right) \qquad (2)$$

where n = the number of release batches, and
t = Student's t-distribution with n-1 degrees of freedom and $\alpha = 0.05$.

Since the transportation studies reports were published, a few more adults have returned. Using more up-to-date (December 1995) recapture data, supplied by the National Marine Fisheries Service (personal communication: Ben Sanford), this method gives varying T/C ratios, depending on how the recoveries are grouped, i.e. by year or across all years (Table 2). Confidence interval widths are very large.

Table 2: Updated T/C ratios for Army Corp of Engineers transportation study in 1986 to 1988 using lognormal distribution.

	Combined T/C ratio	Theoretical 95% c.i.	Averaged T/C ratio	Empirical 95% c.i.
T/C for all 3 years	1.3635	(1.0946, 1.6986)	1.5365[a]	(0.1754, 13.4556)
Annual T/C ratio				
1986	0.7288	(0.3238, 1.6407)	0.5630[a]	(0.0051, 62.192)
1987	1.6528	(1.1987, 2.2789)	1.7931	(0.0992, 32.4209)
1988	1.5207	(1.0984, 2.1054)	1.8351	(0.3872, 8.6977)

a. 1986 had 7 release batches excluded from the T/C empirical calculations: 3 release batches with no control returns were discarded due to the division by zero, and 4 with no treatment returns were discarded as the log of zero is negative infinity.

Normal Error, Arc-sine Square-root Transformation

Proportional data are commonly analyzed with the arc-sine square-root transformation in conjunction with regression techniques that assume a normally distributed error. The assumption is that the transformation of the response variable makes the error distribution (the difference between the fitted coefficients and the actual, unknown coefficients) normal in distribution, justifying least-squares linear-regression in determining the release batch effect and the transportation effect on the adult recapture rate. The regression model performed on the transformation of the proportion of adult recaptures from each release would be:

$$arcsin\left(\sqrt{\frac{n_{ij}}{N_{ij}}}\right) = \alpha + \beta_i + \tau_j \qquad (2)$$

where: i = 1 to 28 for the 28 release batch groups,
j = 1 for salmon released as a control smolt, 2 for salmon transported as smolt,
n_{ij} = number of adults recaptured from release batch i, treatment j,
N_{ij} = number of smolt released in batch i, treatment j,
α = the intercept, or base-line arc-sine square-root of percentage of adults expected to be recovered,
β_i = release batch effect, and
τ_j = transportation effect.

An analysis of deviance of this model (Table 3) indicates that the batch effect is significant (p = 0.0002), and that the effect of transportation is barely nonsignificant (p = 0.1086). So according to this regression model, there is not an over-all appreciable difference in adult salmon recovery proportions of transported fish to control fish. Table 4 has the treatment coefficient and standard error, but what exactly *is* the effect of transportation? This characterization of the coefficient does not easily explain the relationship of transportation to adult salmon recovery.

Table 3: Analyses of deviance table based on the arc-sine square-root transformation of data and normal error distribution.

Source	Degrees Freedom	Deviance	Mean Dev.	F	p
Total_{corr}	55	86.3250			
batch	27	68.3433	2.5312	4.1883	0.0002
treatment	1	1.6641	1.6641	2.7535	0.1086
error	27	16.3176	0.6044		

Table 4: Estimation of treatment coefficient and standard error obtained from arc-sine transformation of data, normal error distribution.

	Coefficients	Standard error	t value
treatment (τ)	0.00462	0.00278	1.659

Now, use the fact that, by rearranging model (2):

$$\frac{n_{ij}}{N_{ij}} = (\sin(\alpha + \beta_i + \tau_j))^2 \qquad (3)$$

the back-transformation can be used to estimate the recovery proportion of adults. Substituting the back-transformation into the transportation-control (T/C) ratio formula:

$$\text{T/C}_i = \left[\frac{\sin(\alpha + \beta_i + \tau_2)}{\sin(\alpha + \beta_i + \tau_1)}\right]^2 \qquad (4)$$

where: i = 1 to 28 for the 28 release batch groups,
α = the intercept,
β_i = batch i release group effect,
τ_1 = 0, for the control releases, and
τ_2 = 0.00462, for the transportation effect from Table (4).

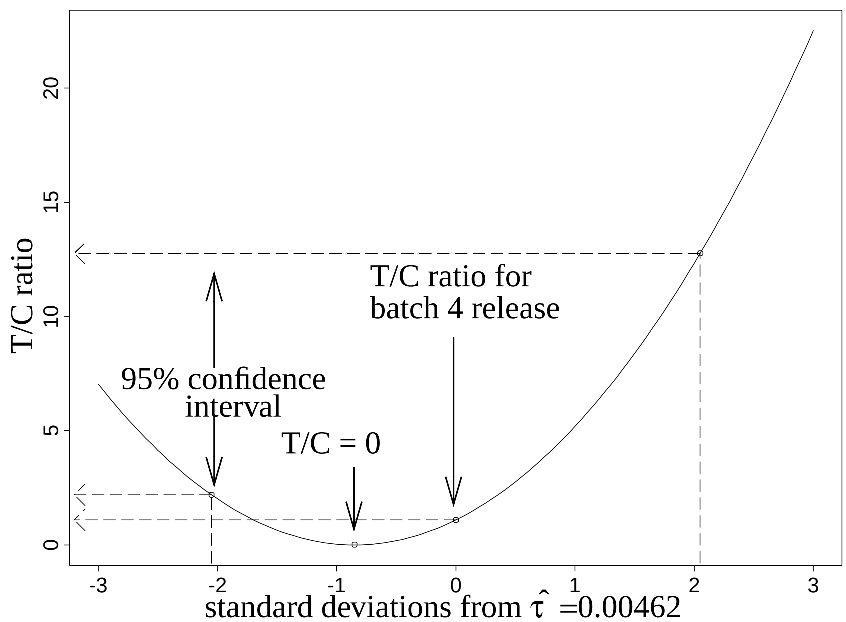
The 95% confidence interval is estimated by replacing t2 with $\tau_2 \pm t_{27df}^{0.95} \bullet se(\tau_2)$ (i.e. $2.0518 \bullet 0.00278$). The regression model is additive, resulting in a complex back-transformed T/C ratio. As batch effect cannot be isolated from the treatment effect, (to set β_i to zero would be the same as calculating the T/C ratio for the first release batch only), the estimated T/C ratio differs from batch to batch. The minimum and maximum batch T/C ratios shows a wide difference (Table 5). In addition, the cyclical nature of sine functions and the back-transformation create a situation (Figure 1) where the back-transformed T/C ratio in batch 4 is not included in the back-transformed confidence interval of the estimated mean!

Table 5: Comparison of the minimum and maximum batch estimates of the T/C and 95% confidence interval widths[a], based on the arc-sine-square-root transformation of data.

batch #	T/C	95% confidence interval
4	1.0934	2.2018-12.7816
3	4.0076	0.5822-10.5022

a. *t* distribution with 27 degrees of freedom

Figure 1: Plot of the back-transformed T/C ratios versus the estimated τ (treatment effect) 3 standard deviations for release batch 4. The back-transformation for the different values of τ shows a concave pattern, and the estimated T/C ratio for the batch 4 release to be outside of its 95% confidence interval.



Normal Error, Logit Transformation

The logit is another commonly used transformation of proportional data. The transformation is the log of the odds ratio[2] in the form:

2. Odds ratio is the ratio of percent of successes (recovered adults) over the percent of failures (non-recovered adult salmon).

$$\log\left(\frac{n_{ij}/N_{ij}}{1-(n_{ij}/N_{ij})}\right) \tag{5}$$

where: i = 1 to 28 for the 28 release batch groups,
j = 1 for salmon released as a control smolt, 2 for salmon transported as smolt,
n_{ij} = number of adults recaptured from release batch i, treatment j, and
N_{ij} = number of smolt released in batch i, treatment j.

Unfortunately, this transformation is undefined whenever there are no adults recaptured or 100 percent of the adults are recovered. In this data set, 8 of the 56 releases would be discarded because of no adult recoveries. The result would be to overestimate the expected adult recapture rate, as the zero data would not be included in the regression. To correct this, a small constant is often added to the numerator and subtracted from the denominator so that the log of zero or one will never occur. The transformation is now rewritten as:

$$\log\left(\frac{\frac{n_{ij}+c}{N_{ij}}}{1-\frac{n_{ij}-c}{N_{ij}}}\right), \text{ which reduces to: } \quad \log\left(\frac{n_{ij}+c}{N_{ij}-n_{ij}+c}\right) \tag{6}$$

where i = 1 to 28 for the 28 release batch groups,
j = 1 for salmon released as a control smolt, 2 for salmon transported as smolt,
n_{ij} = number of adults recaptured from release batch i, treatment j,
N_{ij} = number of smolt released in batch i, treatment j, and
c = some arbitrary constant value.

The response model used in the analysis of the T/C data can then be written as:

$$\log\left(\frac{n_{ij}+c}{N_{ij}-n_{ij}+c}\right) = \alpha + \beta_i + \tau_j \tag{7}$$

where i = 1 to 28 for the 28 release batch groups,
j = 1 for salmon released as a control smolt, 2 for salmon transported as smolt,
n_{ij} = number of adults recaptured from release batch i, treatment j,
N_{ij} = number of smolt released in batch i, treatment j,
α = the intercept, or base-line log-odds of adult proportion recaptured,
β_i = batch i release group effect,
τ_j = transportation effect.

The addition of the small constant “c” can affect the value of the transformation, and subsequently, the outcome of the analysis. The effect of the added constant on the regression results using model (7) is demonstrated in Figure 2. From these plots, it can be seen that, as c is increased from 0 to approximately 0.3, there is a rapid change in the estimated parameters from the regression. Constants greater than 0.3 appear to give approximately the same parameter estimates, though these are still different from those found by adding a smaller constant than 0.3 (Fig. 2a). An analysis of deviance can differ tremendously, as the estimated p-value of the treatment coefficient drops quickly to become highly significant (Fig. 2b). As the constant *c* increases, the logit transformation tends to bring the transformed response closer to 1 for all responses, and the residual deviances from the regression model (7) and the null model converge to zero. The recommended constant, $c = 0.5$ (the vertical dashed line in Fig. 2), is called the *empirical logistic transformation* (Cox 1970), and has the property of having an asymptotic bias of order $O(N^{-2})$[3]. Any other constant has a bias of $O(N^{-1})$(McCullagh and Nelder 1991, pp:106-7).

Table 6 is the analysis of deviance table for the logit response model (7). Again, there is a significant batch effect (p=0.0002) and a nonsignificant difference in the proportions of transported smolt to control smolt recovered as adults (p = 0.1658).

Table 6: Analyses of deviance table for the empirical logit transformation (7) of data and normal error distribution.

Source	Degrees Freedom	Deviance	Mean Dev.	F	p
$Total_{corr}$	55	371054.5542			
batch	27	294064.0314	10891.2604	4.1065	0.0002
treatment	1	5380.3071	5380.3071	2.0286	0.1658
error	27	71610.2157	2652.2302		

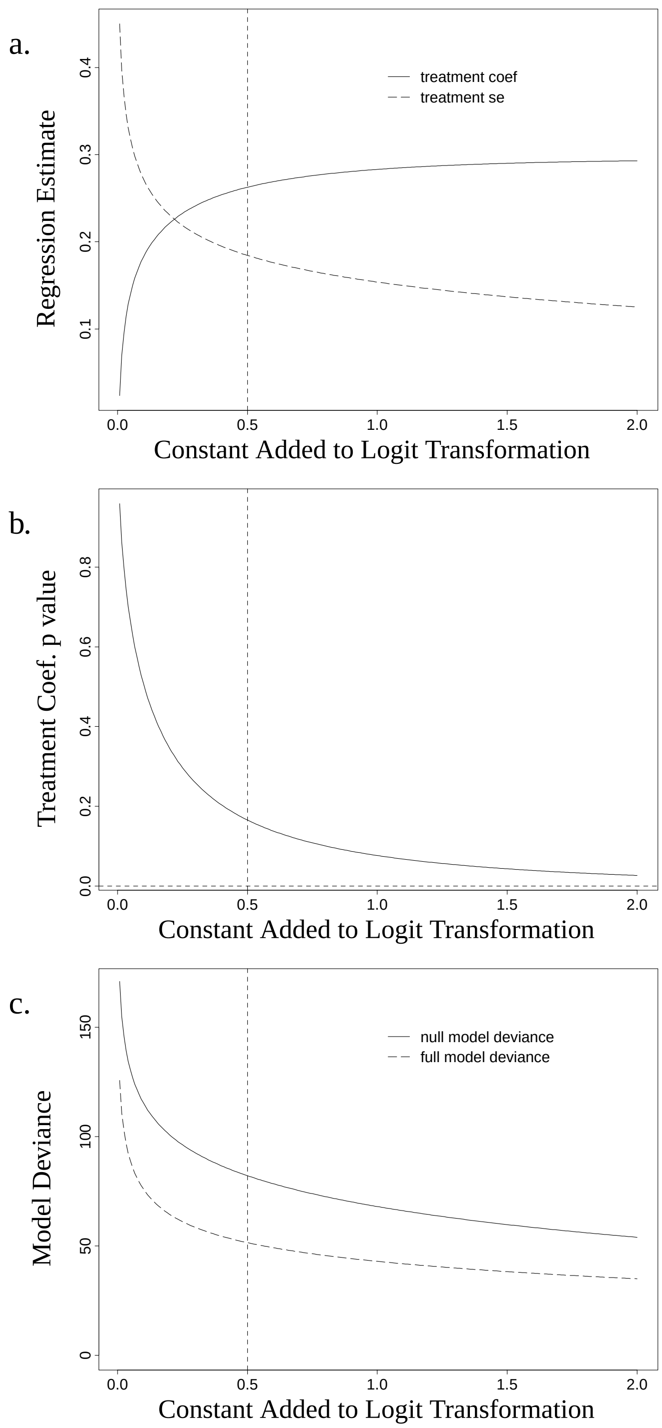
The treatment coefficient (τ) (Table 7) is the estimated treatment effect on the log odds-ratio of adult recovery. Transporting smolt increased the log-odds by 0.2626. By taking the exponential of the coefficient to get the multiplicative effect, transportation increases the odds-ratio of adult recovery by 1.3 times. Not wanting to give Vegas-styled odds on the results of an experiment, the T/C ratio can be calculated by the back-transformation of the odds-ratio.

Table 7: Estimation of treatment coefficient and standard error from logit transformation of data, normal error distribution.

	Coefficients	Standard error	t value
treatment (τ)	0.2626	0.1844	1.4243

3. $O(N^{-2})$ means that the bias of the transformation goes to zero at the rate of $\frac{1}{N^2}$, as $N \to \infty$. In this paper, N is the number of released salmon smolt.

Figure 2: Demonstration of effect of a constant added to the logit transformation of data on the linear regression results of model (7). (a) Treatment coefficient (t) rapidly rises from a very small value to approximately 0.3 and then asymptotes. The treatment standard error drops rapidly, becoming less than the treatment coefficient (t) as the constant increases, explaining (b) the increase in significance of the transportation effect. (c) Model residual deviance and null model (intercept only) deviance quickly converge as the value of the arbitrary constant increases. The vertical dashed line represents the recommended added constant of 0.05.



Solving model (7) for the probability of adult recovery from a release:

$$E\left(\frac{n_{ij}}{N_{ij}}\right) = \frac{\left(1 + \frac{c}{N_{ij}}\right)e^{\alpha + \beta_i + \tau_j} - \left(\frac{c}{N_{ij}}\right)}{1 + e^{\alpha + \beta_i + \tau_j}} \qquad (8)$$

where: i = 1 to 28 for the 28 release batch groups,

j = 1 for salmon released as a control smolt, 2 for salmon transported as smolt,

n_{ij} = number of adults recaptured from release batch i, treatment j,

N_{ij} = number of smolt released in batch i, treatment j,

c = the constant added in the transformation,

α = the intercept,

β_i = batch i release group effect, and

τ_j = transportation effect.

Substituting Eq. 8 into the T/C ratio formula:

$$E(T/C)_i = \frac{\dfrac{\left(1 + \frac{c}{N_{i2}}\right)e^{\alpha + \beta_i + \tau} - \frac{c}{N_{i2}}}{1 + e^{\alpha + \beta_i + \tau}}}{\dfrac{\left(1 + \frac{c}{N_{i1}}\right)e^{\alpha + \beta_i} - \frac{c}{N_{i1}}}{1 + e^{\alpha + \beta_i}}} \quad . \qquad (9)$$

It is tempting to discount the correction factor (c/N_{ij}), since for large releases, it is very close to zero and eliminating it would simplify the T/C equation greatly. The percentage of adult recaptures in this data set are very small though, and are greatly influenced by the correction factor c. Also notice that the T/C ratio estimated in this manner is not independent of the batch effect (a similar result occurred with the arc-sine square-root transformation). Calculating the T/C ratio for the minimum and maximum batches reveals how much the batch effect influences the results (Table 8).

Table 8: Comparison of the minimum and maximum batch estimates of the T/C ratio and 95% confidence interval widths[a] based on the empirical logit transformation of data.

Batch #	T/C	95% Confidence Interval
4	-1.2193	(-6.5372) - 2.4238
3	1.9057	0.6915 - 3.6780

a. *t* distribution with 27 degrees of freedom

Batch #4 has a *negative* T/C ratio due to the correction creating a negative probability for recovering control smolt (i.e. the *corrected* probability of recovering adult control fish = $-(9.64 \cdot 10^{-06})$ and the corrected probability of transported smolt returning as adults = $11.75 \cdot 10^{-06}$). The correction constant caused the T/C ratio to become negative. The lower 95% confidence bound has a negative probability for both the treated and control smolt, resulting in a positive lower bound for the batch 4 T/C ratio. These problems are a consequence of the use of data transformations in calculating a confidence interval.

All this contributes to a very messy determination of the actual T/C ratio and it's confidence interval. In this vein, the General Linear Model is introduced, using a logit "link" rather than a data transformation of the response variable.

Binomial Error, Logit Link

A "link" differs from a transformation in that it is not the response variable that is transformed, rather it is the fitted value from a model that is back-transformed and compared to the response variable. In this case, the back-transformation is taken of model (7) (without the added constant) to create the regression model (10).

$$E\left(\frac{n_{ij}}{N_{ij}}\right) = \frac{e^{\alpha + \beta_i + \tau_j}}{1 + e^{\alpha + \beta_i + \tau_j}} \quad , \tag{10}$$

where: i = 1 to 28 for the 28 release batch groups,
j = 1 for salmon released as a control smolt, 2 for salmon transported as smolt,
n_{ij} = number of adults recaptured from release batch i, treatment j,
N_{ij} = number of smolt released in batch i, treatment j,
α = the intercept,
β_i = batch i release group effect, and
τ_j = transportation effect.

This statistical approach has the advantage that no arbitrary constant need be added and normality need not be assumed. Using Iterative Reweighted Least Squares (IRLS) to fit the model, the log of the response is never actually calculated. Instead, difference between the exponential of the right-hand side of the model and the response variable is minimized. This allows for regression with the inclusion of those batches that had either no adult recoveries or 100 percent of the adults recovered from a smolt batch release.

The results of this approach show that both batch and treatment effects are significant (p = 0.0002 and 0.0179, respectively) (Table 9), a different finding from that of both of the response-transformation approaches, models (2) and (7).

Table 9: Analyses of deviance table for a general linear model with logit-link and binomially distributed error assumed for model (10).

Source	Degrees Freedom	Deviance	Mean Dev.	F	p
$Total_{corr}$	55	291.4511			
batch	27	224.0609	8.2986	4.1070	0.0002
treatment	1	12.8352	12.8352	6.3522	0.0179
error	27	54.5550	2.0206		

Substituting model (10) into the T/C ratio equation, a much simpler formula than equation (9) appears:

$$T/C_i = \quad e^{\tau} \bullet \frac{1 + e^{\alpha + \beta_i}}{1 + e^{\alpha + \beta_i + \tau}} \tag{11}$$

where: i = 1 to 28 for the 28 release batch groups,
j = 1 for salmon released as a control smolt, 2 for salmon transported as smolt,
n_{ij} = number of adults recaptured from release batch *i*, treatment *j*,
N_{ij} = number of smolt released in batch *i*, treatment *j*,
α = the intercept,
β_i = batch *i* release group effect, and
τ = transportation effect.

The denominator and numerator in the second half of Equation 11 are so close in value for all batches that the exponential of the treatment coefficient (e^{τ}) is effectively multiplied by one (Table 10). This has the bonus of giving the same T/C ratio (1.4974) to all batches. In data sets with a greater treatment effect, this would not be true.

Table 10: Estimation of treatment coefficient and standard error, general linear model regression with the logit-link and binomially distributed error, and the resulting T/C ratio and 95% confidence interval width[a].

	Coefficients	Standard error	t value	T/C	95% Confidence Interval
treatment (τ)	0.4037	0.1524	2.6487	1.4974	1.0953 - 2.0473

a. *t* distribution with 27 degrees of freedom

Compared to the results from the logit transformation of the response variable, the treatment coefficient (and thus the T/C ratio) estimate has increased and the standard error and T/C confidence interval width have decreased. The next method investigated in determining the T/C ratio is a less complicated model than the logit model.

Poisson Error, Log Link

Another way to look at the recovery data is as a discrete count of adult salmon recoveries randomly occurring throughout the experimental time frame. A Poisson distribution describes the expected distribution of those counts and the associated errors. The assumption is that the probability of recovering an adult is constant through time, and that there may be factors that affect that probability one way or the other. In this analysis, we are testing to see if transportation affects the probability of recovering adult salmon. The regression model (12) would use the count of recovered adults from each release as the response variable, vice the percent recovered as in the response transformation models (2) and (7) and the binomial-link GLM method (10). The link function in this regression approach is the natural log function. Remember that with a link function, it is the right-hand side of the equation which is transformed and not the response variable.

$$E(n_{ij}) = e^{N_{ij} + \alpha + \beta_i + \tau_j} \qquad (12)$$

where: i = 1 to 28 for the 28 release batch groups,
j = 1 for salmon released as a control smolt, 2 for salmon transported as smolt,
n_{ij} = number of adults recaptured from release batch i, treatment j,
N_{ij} = number of smolt released in batch i, treatment j,
α = the mean count of recovered adult salmon,
β_i = batch i release group effect, and
τ_j = transportation effect.

To account for the differences between the expected number of adult salmon recovered from each batch release due to the different release sizes (i.e., the larger releases would be expected to have a greater number of recoveries), the release size, N_{ij} is entered into regression as an offset value. Since N_{ij} is a known quantity, a regression coefficient is not calculated for it.

The p values in Table 11 of the blocking and treatment effects are both significant (p = 0.0002 and 0.0180, respectively), and approximately equivalent to the binomial logit-link model (10). An additional advantage to using the log-link is that the regression model is multiplicative (i.e. $e^{a+b+c}=e^a e^b e^c$). Taking the exponential of the treatment coefficient gives the T/C ratio directly (i.e. $T/C = e^{\tau}$), and the ratio will be the same across all batches. Table 12 has the treatment coefficient and the estimated effect on the probability of recovering adult salmon from each release. Under the Poisson error assumption, the T/C ratio's confidence interval is slightly larger than its counterpart using a binomial error assumption and a logit link. The next method examines using a binomial error distribution with the multiplicative regression model.

Table 11: Analyses of deviance table for a general linear model with log-link and Poisson distributed error for model (12).

Source	Degrees Freedom	Deviance	Mean Dev.	F	p
$Total_{corr}$	55	291.1585			
batch	27	223.8468	8.2906	4.1073	0.0002
treatment	1	12.8135	12.8135	6.3480	0.0180
error	27	54.4982	2.0185		

Table 12: Estimate of treatment coefficient and standard error, using a general linear model with a log-link and Poisson distributed error, and the resulting T/C ratio and 95% confidence interval width[a].

	Coefficients	Standard error	t value	T/C	95% Confidence Interval
treatment (τ)	0.4031	0.1522	2.6480	1.4964	1.0950 - 2.0449

a. *t* distribution with 27 degrees of freedom

Binomial Error, Log Link

The regression model assuming a binomial distribution with a logit-link gave a better estimate of the T/C ratio than the response transformation models (2) and (7), but was a little difficult to actually calculate, while the model assuming the Poisson distribution of the adult recoveries with a log-link made the T/C ratio calculations easier but used an offset, a function which may not exist or may take some effort to incorporate into a less sophisticated statistical software package. Theoretically, the binomial distribution is also more appropriate for this analysis than the Poisson distribution. When N, the release size, is big, and n, the number of recovered adult salmon is small, the Poisson distribution will approximate a binomial distribution quite closely. What makes one distribution more appropriate than the other is the knowledge of N. When N is unknown, and thus an upper limit to the possible number of adults that can be recovered is unknown, a Poisson distribution is used. When N is known, as in this case, a binomial distribution more accurately describes the distribution of the response variable.The regression model (13) assumes a binomial distribution of the percent of adult salmon recovered from each batch, and uses a log-link. This method is an attempt to use the best of both worlds, a multiplicative model without the offset term.

$$E\left(\frac{n_{ij}}{N_{ij}}\right) = e^{\alpha + \beta_i + \tau_j} \tag{13}$$

where: i = 1 to 28 for the 28 release batch groups,
j = 1 for salmon released as a control smolt, 2 for salmon transported as smolt,
n_{ij} = number of adults recaptured from release batch i, treatment j,
N_{ij} = number of smolt released in batch i, treatment j,
α = the mean count of recovered adult salmon,
β_i = batch i release group effect, and
τ_j = transportation effect.

Table 13 has similar results from an analysis of deviance as the binomial logit-link and the poisson log-link approaches. Examination of the p values show that blocking and treatment effects are significant (p = 0.0002 and 0.0173, respectively). The exponential of the treatment coefficient (i.e. T/C = e^{τ}) gives the T/C ratio directly and is constant across all batches (Table 12) due to the multiplicative nature of the regression model. The treatment coefficient confidence interval width is smaller then when either the Poisson distribution is assumed (Table 12), or the binomial logit-link regression (Table 10). The last method to be examined uses maximum likelihood methods to estimate confidence intervals for the T/C ratio.

Table 13: Analyses of deviance tables for a general linear model with log link and binomially distributed error for model (13).

Source	Degrees Freedom	Deviance	Mean Dev.	F	p
$Total_{corr}$	55	291.4511			
batch	27	223.907	8.2929	4.1044	0.0002
treatment	1	12.9910	12.9910	6.4296	0.0173
error	27	54.5529	2.0205		

Table 14: Estimation of treatment coefficient and standard error from a general linear model using a log link and binomially distributed error, and the resulting T/C ratio and 95% confidence interval width[a].

	Coefficients	Standard error	t value	T/C	95% Confidence Interval
treatment (τ)	0.4031	0.1522	2.6493	1.4965	1.0952 - 2.0448

a. *t* distribution with 27 degrees of freedom

Binomial Error, Maximum Likelihood Estimation

Without using linear regression, a maximum likelihood method can be used to find the T/C ratio (assumed to be constant through the program). Assume that there is a base probability p_c of a control fish returning and probability τp_c of a transported fish returning (where τ is the defined as the T/C ratio). Furthermore, assume that this p_c is different for each batch (which we have already done when we used batch as a blocking effect in the previous analyses). We can then write the likelihood of the counts of recovered adult salmon as:

$$L\langle \tau, p_c | \underset{\sim}{n}_c, \underset{\sim}{n}_t, \underset{\sim}{R}_c, \underset{\sim}{R}_t \rangle = \prod_{i=1}^{m} \binom{\underset{\sim}{R}_{ci}}{\underset{\sim}{n}_{ci}} p_{ci}^{\underset{\sim}{n}_{ci}} (1-p_{ci})^{(\underset{\sim}{R}_{ci}-\underset{\sim}{n}_{ci})} \bullet \binom{\underset{\sim}{R}_{ti}}{\underset{\sim}{n}_{ti}} (\tau p_{ci})^{\underset{\sim}{n}_{ti}} (1-\tau p_{ci})^{\underset{\sim}{R}_{ti}-\underset{\sim}{n}_{ti}} \tag{14}$$

where $\underset{\sim}{R}_{ci}$ = number of control smolt for release batch i,
$\underset{\sim}{R}_{ti}$ = number of transported smolt in release batch i,
m = the number of release batches (m = 28),
$\underset{\sim}{n}_{ci}$ = number of recovered control adult salmon from release batch i,
$\underset{\sim}{n}_{ti}$ = number of recovered transported adult salmon from release batch i,
p_{ci} = probability of fish returning for batch i, and
τ = Treatment-Control ratio.

Maximizing the likelihood (14) for both τ and p_{ci} gives an estimate for the T/C ratio. The maximum likelihood for each batch p_{ci} is found by solving the likelihood equation (15) using a fixed τ:

$$\hat{p}_c = \frac{(\tau\underset{\sim}{n}_c + \underset{\sim}{n}_t + \underset{\sim}{R}_c + \tau\underset{\sim}{R}_t) \pm \sqrt{(\tau\underset{\sim}{n}_c + \underset{\sim}{n}_t + \underset{\sim}{R}_c + \tau\underset{\sim}{R}_t)^2 - 4\tau(\underset{\sim}{R}_c + \underset{\sim}{R}_t)(\underset{\sim}{n}_c + \underset{\sim}{n}_t)}}{2\tau(\underset{\sim}{R}_c + \underset{\sim}{R}_t)} \tag{15}$$

then maximizing for τ using the new MLE p_c (the minimum of the two p_c's obtained from Eq. (15)). The maximum likelihood for τ was solved by using the function “nlmin” in Splus (Dennis, Gay and Welsh,1981, and Dennis and Mei,1979).

To calculate a profile confidence interval (McCullagh and Nelder, 1991) for τ, a new τ (τ_{new}) is selected and used to calculate a new likelihood. Theoretically, the function of twice the negative ratio of the original likelihood over the new likelihood (16) has a χ^2 distribution with one degree of freedom.

$$-2 \cdot \log\left(\frac{L_c}{L_m}\right) \tag{16}$$

where L_m = the value of the likelihood (14) using τ, and
L_c = the value of the likelihood (14) using τ_{new}.

A 95% confidence interval can be found by reiteratively selecting τ_{new}'s away from τ until the likelihood ratio test is rejected at the 5% significance level (two-sided). At first glance, the resulting T/C ratio and confidence interval (Table 15) show a great improvement over previous methods. The T/C ratio is the same as the binomial and Poisson log-link methods (1.4965), and the confidence interval for the estimate is much smaller. Further investigation into the regression

Table 15: Maximum likelihood estimate (14) of the T/C ratio and the 95% confidence interval.

T/C	95% Confidence Interval
1.4965	1.1964 - 1.7966

residuals indicate that there is some over-dispersion present. Over- and under-dispersion occurs when the model does not quite fit the data, and there is more or less (respectively) variance in the fit than would be expected if the regression model was correct. The regression model variance estimates are automatically adjusted to account for this in the GLM methods used by S-plus. The profile likelihood process described above does not adjust for the over-dispersion. Theoretically, the dispersion (i.e. error deviance) of a model is distributed χ^2 with *n-p* degrees of freedom, where *n* is the number of observations and *p* is the number of parameters in the model (McCullagh and Nelder, 1991). By calculating this dispersion and dividing by the (*n-p*) degrees of freedom, the scale parameter for a particular model is obtained.

$$ErrorDeviance(\hat{p}) = -2\log\left(\frac{L_{\hat{p}}}{L_{\tilde{p}}}\right) \quad (17)$$

where $L_{\hat{p}}$ = the value of the likelihood (14) of the fitted model, and
$L_{\tilde{p}}$ = the value of the likelihood (14) of the null model, where $\tilde{p}_i = n_i/R_i$ and $\tau = 1$.

From the final model, the deviance calculated is 54.5501. The variance scale parameter (18) is found by dividing the deviance by 26 (55-29) degrees of freedom, for a over-dispersion estimate of 2.0981. The square-root of the over-dispersion (1.4485) is the multiplier used to correct the confidence interval width (Table 16). This adjusted confidence interval width is smaller and shifted to the left of the estimates from the GLM methods.

$$sp = \frac{ErrorDeviance(\hat{p})}{df} \quad (18)$$

where: *sp* = the scale parameter, and
df = the degrees of freedom from the regression model.

Table 16: Maximum likelihood estimate (14) of the T/C ratio and the 95% confidence interval adjusted for dispersion.

T/C	95% Confidence Interval
1.4965	1.0618 - 1.9312

Comparison of Statistical Methods

All of the statistical methods explain approximately 81% of the corrected total deviance (R^2) observed (Table 17). Residual Sum of Squares (RSS) (i.e. $\sum(y_{ijk} - \hat{y}_{ijk})^2$) are the conventional means of comparing models, and the Least-Squares estimates of the regression coefficients can be computed by minimizing the RSS. A smaller RSS indicates a better fitting model, and does not depend on the error assumption used in the regression. Further comparison of the statistical methods show two distinct groupings by goodness-of-fit. The data transformations have approximately thirteen to eighteen percent more RSS (indicating a worse fit) than the glm methods and the binomial MLE, which are very close to each other.

Table 17: Proportion of corrected total deviance explained by the fitted model, and the residual sum of squares by the fitted model, where y_{ij} is the observed recovered salmon and $\hat{y}_{ij}$ is the fitted number of adult recovered salmon.

Equation #	error	model	r^2	$\sum(y_{ij} - \hat{y}_{ij})^2$
1	normal	arc-sine transformation	0.8110	358.6978
3	normal	logit transformation	0.8070	375.5955
8	binomial	logit link	0.8128	317.5060
11	binomial	log link	0.8128	317.4694
10	Poisson	log link	0.8128	317.4716
12	binomial	mle	NA	317.4694

Depending on the statistical method used to estimate the Treatment-Control Ratio, the analysis results can vary (Table 18).The normal approximations are the least sophisticated of the methods compared. With most older statistical software able to only perform linear regressions, these methods work, but are highly susceptible to batch (or blocking) effects. Blocking effects were found to be highly significant in all models, indicating that there was a large amount of variance from batch to batch. The generalized linear model regression is a better technique, providing smaller confidence interval widths than the response transformations, and a better fit to the data by allowing analysis of the extreme cases of $n = 0$ or $n = N$. The binomial logit-link and the Poisson log-link models are very similar when the probability of success (an adult salmon is recovered) is very small. The log-link with the binomial error is the most appropriate because the binomial distribution is the proper error structure and the T/C ratio is defined as a multiplicative effect. This results in a small confidence interval width combined with a more direct calculation of the T/C ratio. In all of the models, with the exception of the MLE method, the estimation of the confidence interval used the Student's t-distribution with 27 degrees of freedom (2.0518). The profile confidence interval for the T/C, though, gives the smallest confidence interval width even after accounting for the observed over-dispersion. The binomial MLE approach is the recommended method for Transportation-Control ratio estimation, followed closely by the GLM using a binomial log-link regression model.

Table 18: A comparison of estimates of the T/C ratio's and 95% confidence intervals for methods used in this report.

Equation #	error	model	release batch (where applicable)	T/C	95% Confidence Interval
2	normal	arc-sine transformation	4 (min)	1.0934	2.2018 - 12.7816
			3 (max)	4.0076	0.5822 - 10.5022
7		logit transformation	4 (min)	-1.2193	2.4238 - (-6.5372)
			3 (max)	1.9057	0.6915 - 3.6780
10	binomial	logit link		1.4974	1.0953 - 2.0473
12	Poisson	log link		1.4964	1.0950 - 2.0449
13	binomial	log link		1.4965	1.0952 - 2.0448
14		mle		1.4965	1.1964 - 1.7966
14		mle (adjusted)		1.4965	1.0618 - 1.9312

Literature Cited

Achord, S., J.R. Harmon, D. M. Marsh, B.P. Sandford, K.W. McIntyre, K.L. Thomas, N.N. Paasch and G.M. Matthews. 1992. Research Related to Transportation of Juvenile Salmonids on the Columbia and Snake Rivers, 1991. Annual report to U.S. Army Corps of Engineers, Contract DACW68-84-H0034. 57 p. plus Appendix. (Available from Northwest Fisheries Science Center, 2725 Montlake Blvd. E., Seattle, WA 98112-2097).

Burnham, K.P, D.R. Anderson, G.C. White, C. Brownie and K.H. Pollack. 1987. Design and Analysis Methods for Fish Survival Experiments Based on Release-Recapture. American Fisheries Society, Monograph 5, Bethesda, MD., p. 84.

Cox, D.R. 1970. The Analysis of Binary Data. Chapman and Hall, London.

Cox, D.R. and E.J. Snell. 1968. "A General Definition of Residuals." Journal for the Royal Statistical Society of London, B 30, pp:258-9.

Dennis, J.E., D.M. Gay and R.E. Welsh. 1981. "An Adaptive Nonlinear Least-squares Algorithm." ACM Transactions on Mathematical Software 7, pp:348-383.

Dennis, J.E. and H.H. Mei. 1979. “Two New Unconstrained Optimization Algorithms Which Use Functions and Gradient Values.” Journal of Optimization Theory and Applications 28, pp:453-483.

Harmon, J.R, G.M. Matthews, D.L. Park and T.E. Ruehle. 1989. Evaluation of Transportation of Juvenile Salmonids and Related Research on the Columbia and Snake Rivers, 1988. Annual report to U.S. Army Corps of Engineers, Contract # DACW68-84-H0034. 88 p. (Available from Northwest Fisheries Science Center, 2725 Montlake Blvd. E., Seattle, WA 98112-2097).

Harmon, J.R, B.P. Sandford, K.L. Thomas, N.N. Paasch, K.W. McIntyre and G.M. Matthews. 1993. Research Related to Transportation of Juvenile Salmonids on the Columbia and Snake Rivers, 1992. Annual report to U.S. Army Corps of Engineers, Contract # DACW68-84-H0034. 25 p. plus Appendix. (Available from Northwest Fisheries Science Center, 2725 Montlake Blvd. E., Seattle, WA 98112-2097).

Matthews, G.M., S. Achord, J.R. Harmon, O.W. Johnson, D.M. Marsh, B.P. Sandford, N.N. Paasch, K.W. McIntyre and K.L. Thomas. 1992. Evaluation of Transportation of Juvenile Salmonids and Related Research on the Columbia and Snake Rivers, 1990. Annual report to U.S. Army Corps of Engineers, Contract # DACW68-84-H0034. 51 p. plus Appendix. (Available from Northwest Fisheries Science Center, 2725 Montlake Blvd. E., Seattle, WA 98112-2097).

Matthews, G.M., J.R. Harmon, S. Achord, O.W. Johnson and L.A. Kubin. 1990. Evaluation of Transportation of Juvenile Salmonids and Related Research on the Columbia and Snake Rivers, 1989. Annual report to U.S. Army Corps of Engineers, Contract # DACW68-84-H0034. 59 p. plus Appendix. (Available from Northwest Fisheries Science Center, 2725 Montlake Blvd. E., Seattle, WA 98112-2097).

Matthews, G.M., D.L. Park, J.R. Harmon C.S. McCutcheon and A.J. Novotny. 1987. Evaluation of Transportation of Juvenile Salmonids and Related Research on the Columbia and Snake Rivers, 1986. Annual report to U.S. Army Corps of Engineers, Contract # DACW68-84-H0034. 123 p. (Available from Northwest Fisheries Science Center, 2725 Montlake Blvd. E., Seattle, WA 98112-2097).

Matthews, G.M., D.L. Park, J.R. Harmon and T.E. Ruehle. 1988. Evaluation of Transportation of Juvenile Salmonids and Related Research on the Columbia and Snake Rivers, 1987. Annual report to U.S. Army Corps of Engineers, Contract # DACW68-84-H0034. 161 p. (Available from Northwest Fisheries Science Center, 2725 Montlake Blvd. E., Seattle, WA 98112-2097).

McCullagh, P. and J.A. Nelder. 1991. Generalized Linear Models, 2nd ed. Chapman and Hall, London, pp:106-107, 188-119, 471.

Appendix A

Table A1: The chinook salmon smolt release and adult recovery data used in the T/C ratio analysis. The total adult recoveries are the total number adult salmon uniquely observed throughout the experimental time-frame. Repeated observations (i.e. salmon captured at a dam trap, jaw-tagged and released to be recovered again further up-stream) were removed.

year	batch code	Controls			Transports		
		CWT tag	number of smolt marked	total adult recoveries	CWT tag	number of smolt marked	total adult recoveries
1986	1	231729	5620	4	231846	5235	1
1986	2	231845	5054	0	231848	4936	1
1986	3	231847	5168	1	231850	5209	0
1986	4	231849	5243	0	231852	5014	0
1986	5	231851	5329	1	231854	5119	0
1986	6	231853	5158	1	231856	5106	0
1986	7	231855	5043	1	231858	5011	0
1986	8	231857	5111	0	231860	5099	1
1986	9	231859	5079	3	231861	5032	1
1986	10	231919	3472	3	231920	3513	6
1987	11	231949	7365	13	232008	4957	7
1987	12	231950	7501	12	232009	5000	3
1987	13	231951	7500	13	232010	5000	11
1987	14	231952	7500	12	232011	5003	4
1987	15	231953	7501	10	232012	5000	23
1987	16	231954	7505	1	232013	5002	11
1987	17	231955	7501	3	232014	5000	11
1987	18	231956	5529	7	232015	3525	8
1988	19	232226	7504	18	232236	5002	14
1988	20	232227	7500	19	232237	5002	17
1988	21	232228	7503	9	232238	5002	7
1988	22	232229	7534	8	232239	5011	5
1988	23	232230	7503	6	232240	5002	10
1988	24	232231	7482	3	232241	5002	2
1988	25	232232	7501	1	232242	5001	5
1988	26	232233	7505	3	232243	5003	3
1988	27	232234	7502	2	232244	5002	6
1988	28	232235	7502	3	232245	5002	4

Appendix B

Residual Plots for Alternative Analyses

Quantile-quantile (qq) plots of the residuals from each of the analysis methods aid in comparing methods and determining the appropriateness in applying each particular method to the data. Each qq-plot is the standardized residuals from the regression versus a theoretical normal distribution. If the residuals have a normal distribution, the points will plot along an approximately straight line. Any extreme departures from the normality assumption will be apparent by a departure from linearity. The residuals from two methods involving data transformation may be looked at directly, as they are assumed to have a normal distribution (Figures B1a,b; Figures B2a,b). General linear model regressions, however, give residuals in the same scale as the dependent variable, in this case, a binomial or poisson distribution. Anscombe residuals are used for the qq-plots when this occurs. Anscombe residuals are transformed regression residuals that are asymptotically normal in distribution (Cox & Snell (1968)). When binomial error is assumed (Figures B1c, e and f; Figures B2c, e and f) the residuals are transformed using the equation:

$$h(Y_{ij}) = \left[\phi\left(\frac{Y_{ij}}{m_{ij}}\right) - \phi\left\{\theta_{ij} - \frac{1}{6}(1-2\theta_{ij})/m_{ij}\right\}\right] / \{\theta_{ij}^{1/6}(1-\theta_{ij})^{1/6}/\sqrt{m_{ij}}\} \qquad \text{(B1)}$$

where h(Y_{ij}) = the transformed Anscombe residual for the i^{th} batch, j^{th} treatment,
Y_{ij}, = the observed number of recaptured adults from the i^{th} batch, j^{th} treatment,
m_{ij} = the released number of smolt for the i^{th} batch, j^{th} treatment,
θ_{ij} = the expected percentage, $\hat{p}_{ij}$ for the i^{th} batch, j^{th} treatment, and

$$\phi(u) = \int_0^u t^{-1/3}(1-t)^{-1/3}dt, (0 \le u \le 1).$$

$\phi(u)$ is also known as the incomplete beta function with the shape parameters $B\left(\frac{2}{3}, \frac{2}{3}\right)$.

When the poisson error is assumed (Figures B1d and B2d), the residuals have a simpler transformation:

$$h(Y_{ij}) = \left\{Y_{ij}^{2/3} - \left(\mu_{ij} - \frac{1}{6}\right)^{2/3}\right\} / \left(\frac{2}{3}\mu_{ij}^{1/6}\right) \qquad \text{(B2)}$$

where h(Y_{ij}) = the transformed Anscombe residual for the i^{th} batch, j^{th} treatment,
Y_{ij}, = the observed number of recaptured smolt for the i^{th} batch, j^{th} treatment,
m_{ij} = the released number of smolt for the i^{th} batch, j^{th} treatment, and
μ_{ij} = the expected frequency for the i^{th} batch, j^{th} treatment.

The residual plot (Figure B1) are qq-normal plots of the six different methods presented in this analysis. These are all very similar in both their degree of “normality” and the pattern that each model’s residuals exhibit. Another way of viewing the difference in the fits between the models, the residuals are plotted as an estimated distribution against the standard normal distribution (Figure B2). The departure from normality is more apparent, though not enough to disprove the assumption of the normality distribution of the residuals. The multi-model characteristics of the residuals are due more to the small number of samples than to an actual tri-model distribution.

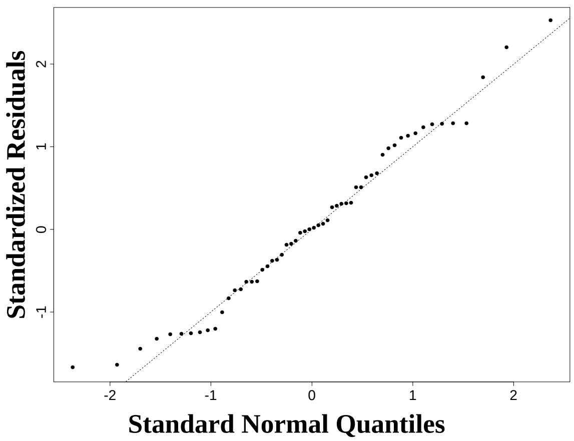
Untransformed Residual Plots

A comparison of the weighted raw residual plots ($(y_i - \hat{y}_i)\sqrt{w_i}$) of each of the model regressions is multi-purpose. Raw residual plots will reveal any bias occurring in the fit (i.e. fitted responses are always higher or lower than the observed responses), or any non-linearity of the data (residuals get smaller or larger as the value of the fitted response increases). The weighted residuals are used to show the amount of influence that a particular data point exerts over the fit. The residuals in the plot (Figure B3) were not standardized, so the vertical scales vary widely in each case. Nothing appears to stand out as unusual in the first five plots, though a symmetric pattern is apparent in both the poisson error (log link), and the maximum likelihood residual plots. Surprisingly, each of these model predict almost the same expected number of adult recaptures for smolt transported as those used as controls in the same release batch, though there is definitely a transportation effect. Further investigation shows that this is the influence of the greater number of controls released over the transported smolt in 1987 and 1988--a ratio approximately equal to the effect of transportation on the survival of the smolt. The reason this symmetry is apparent in these two plots is that the scale of the fitted response, the counts of recovered adult fish, is big enough to make the difference between a control response and a transportation response appear minimal.

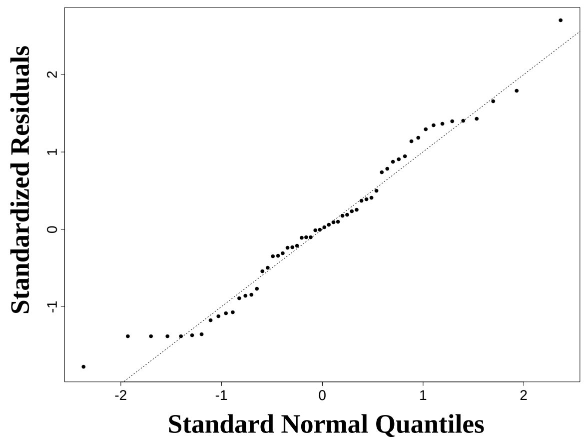
In the maximum likelihood residual plot, a left-to-right “shot-gun” pattern appears. The lower left hand diagonal border represents the limitation that the fitted response be non-negative. Since no fitted count can be less than zero, the size of residuals are restricted for expected counts near zero. The variation is also increasing as the expected count increases, which is appropriate for the type of error assumed. A binomial error has a variation of *p(1-p)*, which increases as *p* goes from zero to 0.5 and then decreases for $p > 0.5$. Since the number of marked smolt are approximately the same, the increase in counts comes from an increase in expected *p (*each batch had it’s own *p* estimated*)*, and therefore will have a greater variance as the expected count gets higher.

Figure B1: QQ-normal plots of residuals from each method of determining the T/C ratio.

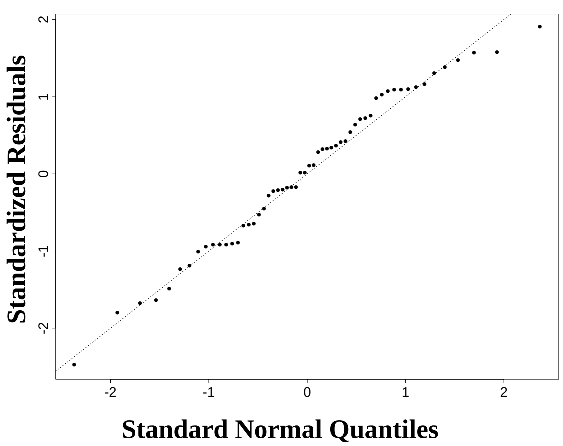
a. arc-sine square-root transformation



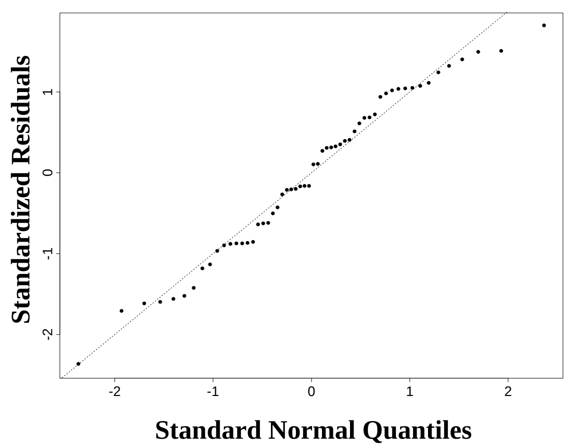
b. logit transformation



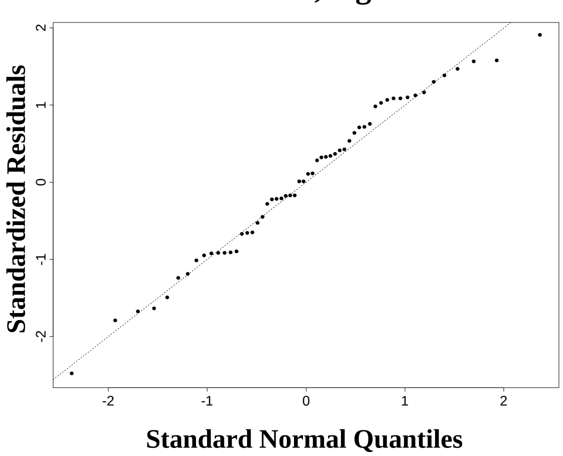
c. binomial GLM, logit link function



d. poisson GLM, log link function



e. binomial GLM, log link function



f. binomial maximum likelihood

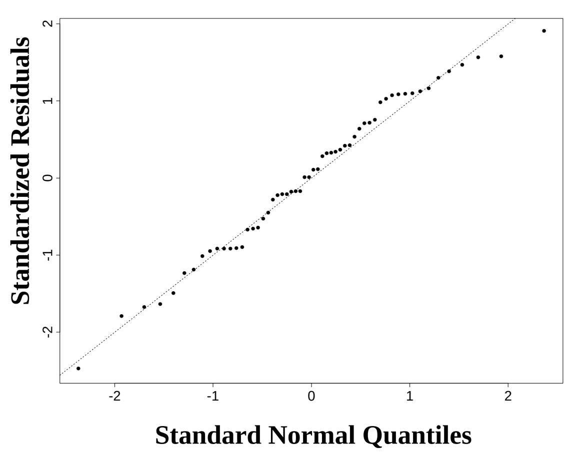


Figure B2: Plots of empirical distribution of the standardized residuals from each model versus the standard normal distribution.

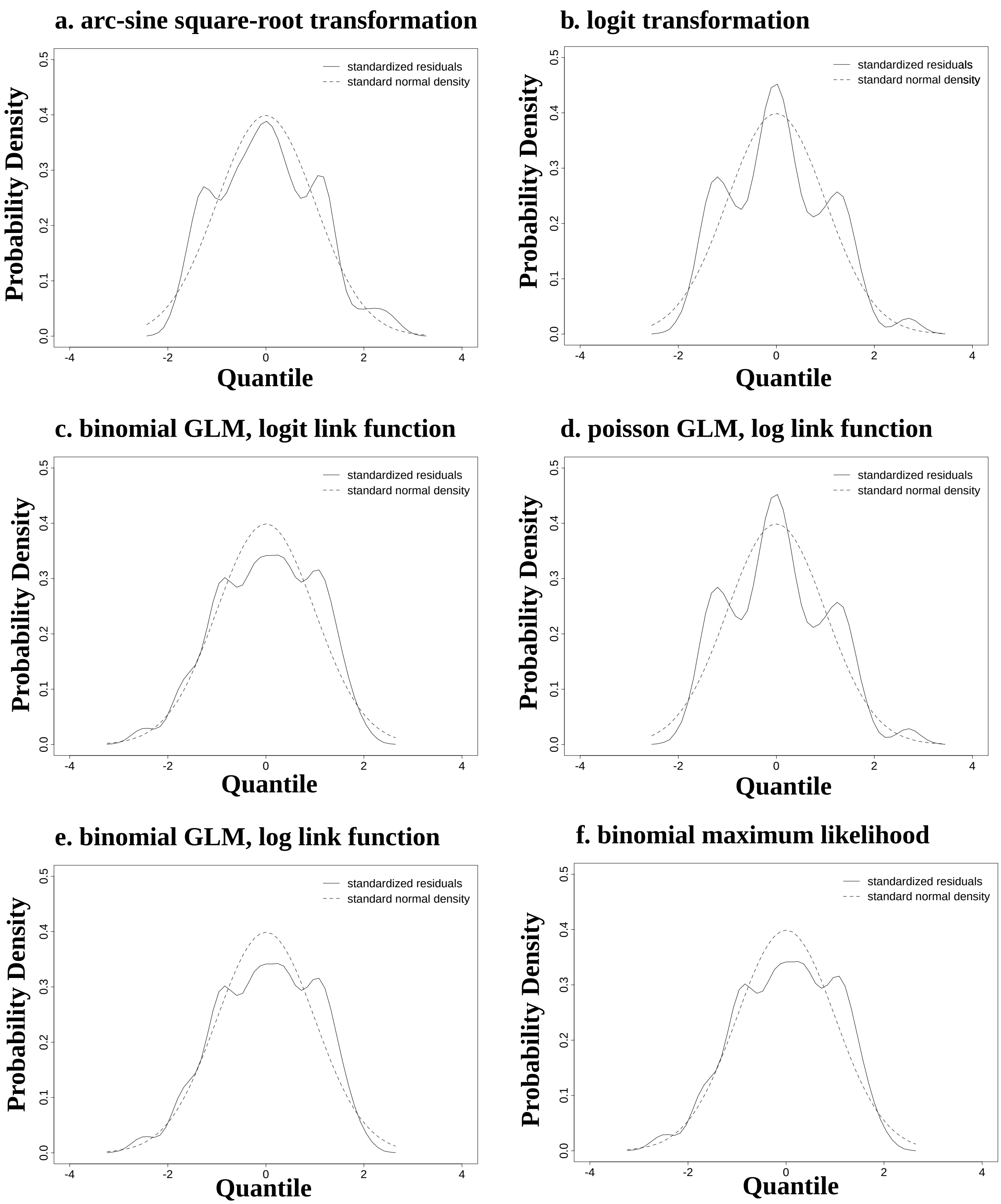


Figure B3: Plots of weighted residuals for (a) arc-sine square-root and (b) logit transformations; general linear model regressions using a (c) logit link, binomial error distribution, (d) log link, poisson error distribution, and (e) log link, binomial error distribution. (f) Residuals for the maximum likelihood, binomial error distribution are unweighted.