T:M – a tool to estimate the power of detecting a difference in smolt-to-adult ratios (SARs) between transported and migrant salmon

Richard A. Hinrichsen

August 27, 2011

Introduction

After 1975, when the federal hydropower system was in place on the Snake and Columbia rivers, the number of adult salmon returning to spawn declined, in part because of reduced survival of juvenile fish during their seaward migration (Raymond 1979, 1988). To aid juveniles during their seaward migration, smolts are collected at Snake River dams and transported (barged) downstream of Bonneville Dam (river km 234) and released (Ebel et al. 1973; Ebel 1980). An important question that emerges is "how effective is transporting juveniles at increasing survival to adulthood compared to allowing the juveniles to migrate in-river to the sea?" To answer this question, smolt-to-adult ratios (SARs) are calculated for these two groups and compared.

Calculation of the SARs is made possible by tagging smolts with passive integrated transponder (PIT) tags. Each tag has a unique numerical code, and those codes are detected as fish pass through detection systems installed in the juvenile bypass systems of most main-stem dams and in adult fish ladders at some of the mainstem dams. An SAR of a release group is simply the number of adult recoveries divided by number released. The SARs are calculated for transported fish and fish migrating in river.

Equations for estimating the power of detecting a difference between the survivals of transported and migrant juvenile salmon are derived. The SAR for transported juveniles divided by SAR for the migrant juveniles is known as the T:M ratio, which is used to gauge the difference in survival (Muir et al. 2006). When a T:M ratio is greater than one, then transported juveniles have higher estimated smolt-to-adult survival than the migrant juveniles.

The website www.onefishtwofish.net contains a web-based tool that implements this power analysis. The assumptions of the analysis are given in Table 1. The code for implementing

this power analysis (Appendix A) was implemented in R, a system for statistical computation and graphics (Venerables et al. 2010).

Table 1.—Assumptions used in the power analysis.

A1	The number of tagged transported smolts and the number of tagged in-river migrant smolts are fixed.
A2	The number of adult recoveries from each tagged group of juveniles is a random variable that follows a binomial distribution.
A3	The probability that a smolt survives and is recovered as adult is a constant for a given release group.
A4	The event that each individual fish is recovered as an adult is independent of the event that any other individual is recovered as an adult.
A5	The probability that an adult is detected if it is alive (detection rate) is 100%.
A6	The null hypothesis is that the true smolt-to-adult survival of transported fish is equal to that of migrant fish.
A7	The estimator log(T:M ratio) is approximately normally distributed.

Methods

To derive equations for estimating the power to detect a difference in the smolt-to-adult survival of transported and in-river migrant groups, maximum likelihood theory is used (Mood et al. 1974). This requires deriving variance equations for the estimators. I begin by deriving a formula for the variance of SAR and log(SAR). These formulas are then used to derive an estimate of the variance of the T:M ratio. Assume that the number of adult recoveries, n, follows a binomial distribution where the true SAR (denoted θ) is the success probability and N, the number of juvenile released, is fixed. Then, it is well known that the maximum likelihood estimator of the success probability is the SAR and the variance of the SAR is equal to

$$\text{var}(SAR) = \theta(1-\theta)/N \tag{1}$$

The log of the survival estimate also has an asymptotic normal distribution, and Burnham et al. (1987) have shown that log (survival) is better approximated by a normal distribution than survival, especially when the coefficient of variation is large. Peterman (1981) examined ocean survival of three different salmon species and found that assuming a normal distribution for log(survival rate) was more consistent with the data than assuming a normal distribution for survival rate itself. Therefore, the power analysis is based on the log(survival) instead of the untransformed survival.

Applying the Delta method to calculate the variance of a transformation (Seber 1982),

$$\operatorname{var}(\log(\hat{\theta})) \cong \operatorname{var}(\hat{\theta}) / \theta^2 = \left(CV(\hat{\theta}) \right)^2 \tag{2}$$

This formula is applied to the SAR to yield

$$\operatorname{var}(\log(SAR)) \cong \frac{(1-\theta)}{\theta N} \tag{3}$$

I next derive the estimate of the variance of a T:M ratio, which is the ratio of the SAR for transported fish divided the SAR for in-river migrating fish. Let SAR_T represent the SAR of transported fish and let SAR_M represent the SAR of in-river migrating fish. Then the T:M ratio is equal to

$$T:M \; ratio = SAR_T / SAR_M \tag{4}$$

Using the Delta method on a ratio of random variables, and assuming that the covariance between the SARs of the transported fish and in-river migrants is zero, the estimate of the variance of the T:M ratio is

$$\text{var}(T:M\ ratio) = \text{var}(SAR_T / SAR_M) \tag{5}$$

$$\cong \begin{bmatrix} \dfrac{1}{\theta_M} & -\dfrac{\theta_T}{\theta_M^2} \end{bmatrix} \begin{bmatrix} \text{var}(SAR_T) & 0 \\ 0 & \text{var}(SAR_M) \end{bmatrix} \begin{bmatrix} \dfrac{1}{\theta_M} \\ -\dfrac{\theta_T}{\theta_M^2} \end{bmatrix}$$

$$= \frac{\text{var}(SAR_T)}{\theta_M^2} + \frac{\text{var}(SAR_M)\theta_T^2}{\theta_M^4}$$

$$= \frac{\theta_T(1-\theta_T)}{\theta_M^2 N_T} + \frac{\theta_T^2 \theta_M(1-\theta_M)}{\theta_M^4 N_M}$$

$$= \left(\frac{\theta_T}{\theta_M}\right)^2 \left(\frac{(1-\theta_T)}{\theta_T N_T} + \frac{(1-\theta_M)}{\theta_M N_M}\right),$$

where is N_T the number of transported juveniles, θ_T is the probability that a transported juvenile survives to be detected as an adult, N_M is the number of juveniles that are in-river migrants, and θ_M is the probability that an in-river migrant juvenile survives to be detected as an adult. Applying the Delta method once again, this time to the log of the T:M ratio, yields

$$\text{var}(\log(T:M\ ratio)) = \frac{(1-\theta_T)}{\theta_T N_T} + \frac{(1-\theta_M)}{\theta_M N_M}. \tag{6}$$

The variance of the estimate in equation (6) is next used to derive the statistical power, which represents the probability of rejecting the null hypothesis of "no treatment" effect when the true treatment effect is not zero. In this case, the treatment effect , defined as δ, is the (true) value of log(T:M ratio) that would be obtained if the sample size were infinite. The estimate of the treatment effect, denoted $\hat{\delta}$, is equal to the observed value of $\log(T:M\ ratio)$.

Power is a function of the true treatment effect, the probability of a type I error (usually called the significance level and denoted by α), and the standard error of the estimator. By maximum likelihood theory, the estimator of the treatment effect is asymptotically normally distributed. Therefore, power is approximately

$$\Pi(\delta) = \Phi(-z_{\alpha/2} - \delta/se(\hat{\delta})) + 1 - \Phi(z_{\alpha/2} - \delta/se(\hat{\delta})) , \tag{7}$$

where $\Phi(z)$ is the cumulative distribution function of a standard normal random variable, and $z_{\alpha/2}$ is the critical value such that $\alpha/2$ probability of a standard normal distribution lies to the right of the value $z_{\alpha/2}$, δ is the true value of the treatment effect, and $se(\hat{\delta})$ is the square root of the variance of $\hat{\delta}$, which may be calculated from equation (6). Experimenters usually choose a design that will deliver a power of 0.80 or greater.

Example 1. —As an example, I estimate power in the simple case when there are equal numbers of PIT- tagged transported and in-river migrant juveniles. The true log(T:M ratio) takes on the values of 0.50, 0.75, 1.25 and 1.50, while the true SAR of the transported group is varied from 0.003 to 0.03. This is within the range of SARs observed for juveniles transported Lower Granite Dam (river km 695) on the Snake River. I ask the question, "How many total juveniles must be tagged in order to reliably detect a significant difference in the SARs of transported and in-river migrants?" To answer this question I set the significance level at 0.05 and seek the release size necessary to deliver a power of 0.80.

The results of this *a priori* power analysis are given in Figure 1. Notice that for a fixed true T:M ratio, as the SAR_T declines, power can decline greatly. Furthermore, the number of tagged transported juveniles strongly affected power and, in most cases, increasing this number to 50,000 would be sufficient to guarantee power greater than 0.80.

Evaluating the log-transformation. —One important assumption of the analysis is that log(T:M ratio) is approximately normally distributed. To demonstrate the effect of the log transformation on the T:M ratio I plot histograms of Monte Carlo replications of the log-transformed and untransformed T:M ratios. These replications were generated using true SAR_T values of 0.003 and 0.03. I set the true T:M ratio to 1.25, N_T to 5,000, and N_M to 5,000. With these inputs, 1,000 Monte Carlo replications were generated by sampling from binomial distributions. The results of these simulations are in Figure 2. The log-transformed replications were more normally distributed than the untransformed replications when SAR_T was relatively small (0.003). When SAR_T was relatively large (0.03), then the histograms of both the T:M ratio and log(T:M ratio) replications appeared to be approximately normal.
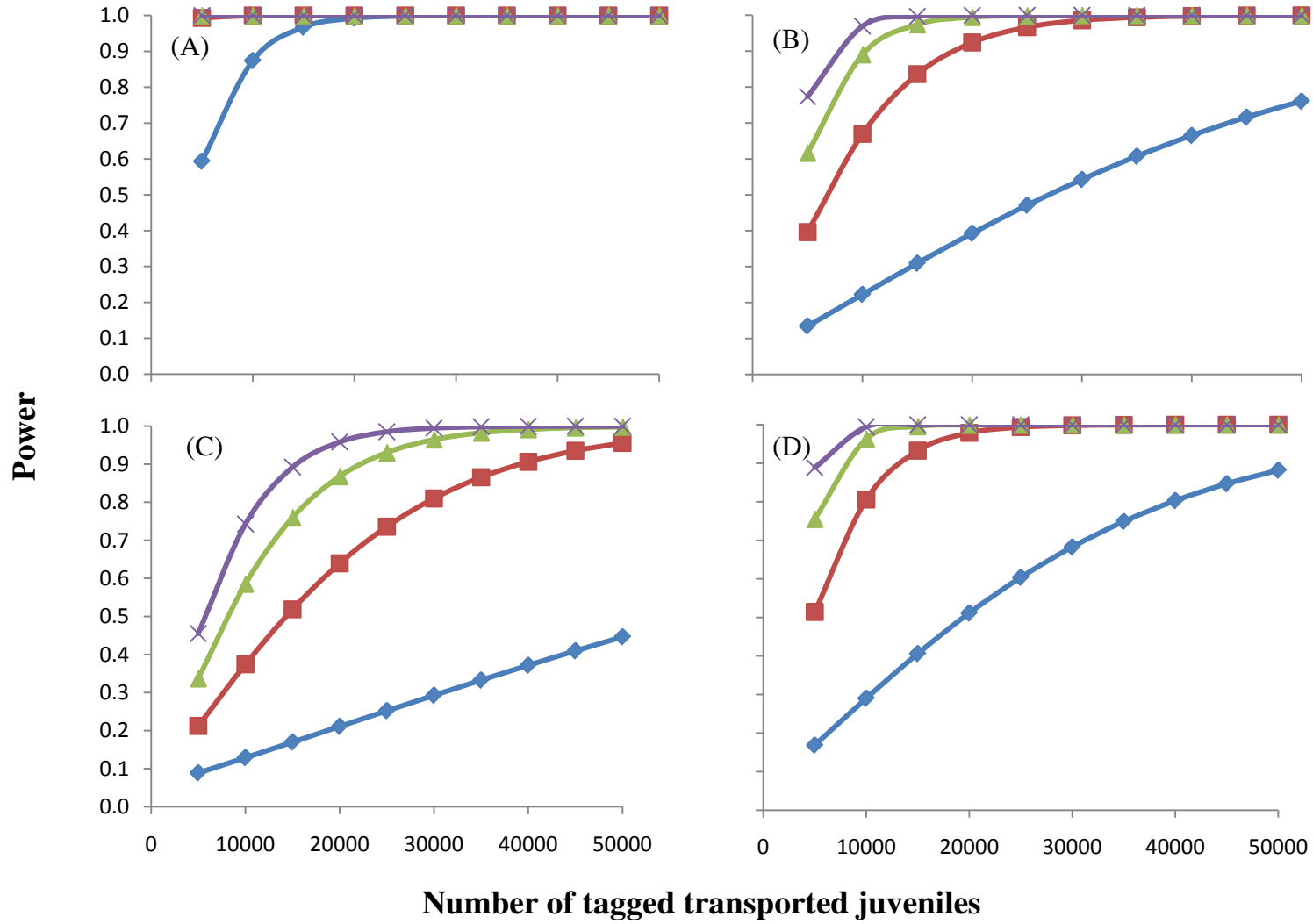
Figure 1.— Power to detect a log(T:M ratio) that is significantly different from zero. The true T:M ratio was assumed to take on the values (A) 0.50 (B) 0.75 (C) 1.25 and (D) 1.50. The values of true SAR_T were 0.003 (diamonds), 0.012 (squares), 0.021 (triangles) and 0.030 (Xs). The number of tagged in-river migrant juveniles was equal to the number of tagged transported juveniles.
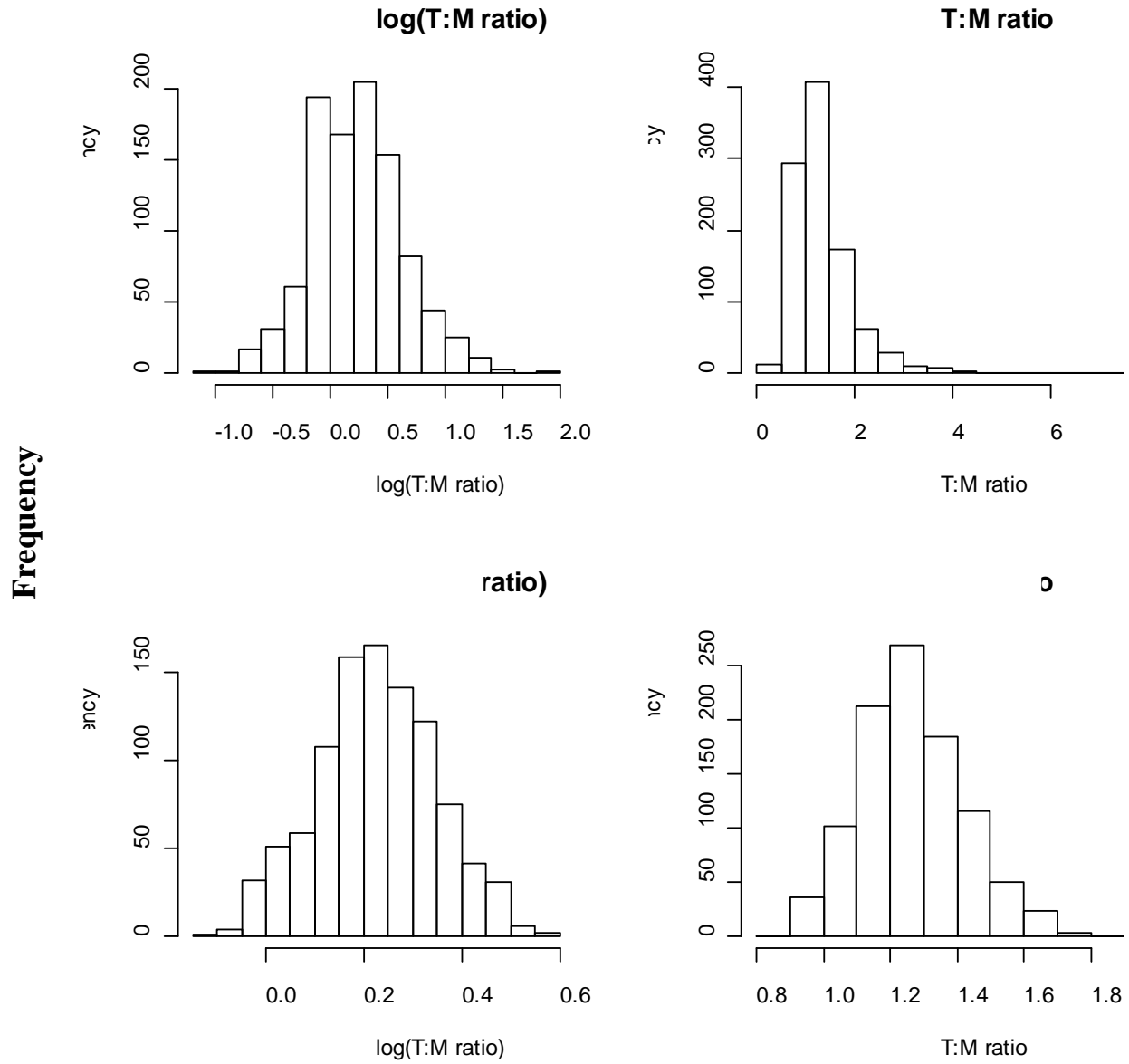
Figure 2. — Histograms of log-transformed and untransformed Monte Carlo replications of the T:M ratio. True SAR_T (SAR of transported juveniles) was set to the values of 0.003 (top two histograms), and 0.03 (bottom two histograms). Notice that when true SAR_T is 0.003, log(T:M ratio) is more normally distributed than the untransformed T:M ratio.

References

Burnham, K. P., D. R. Anderson, G. C. White, C. Brownie, and K. H. Pollock. 1987. Design and analysis methods for fish survival experiments based on release-recapture. American Fisheries Society Monograph No. 5. Bethesda, Maryland, USA. 437pp.

Ebel, W. J. 1980. Transportation of Chinook salmon, *Oncorhynchus tshawytscha*, and steelhead, *Salmo gairdneri*, smolts in the Columbia River and effects on adult returns. U.S. National Marine Fisheries Service Fishery Bulletin 78:491–506.

Ebel, W. J., D. L. Park, and R. C. Johnson. 1973. Effects of transportation on survival and homing of Snake River chinook salmon and steelhead trout. U.S. National Marine Fisheries Service Fishery Bulletin 71:549–563.

Mood, A.M., F.A. Graybill, and D.C. Boes.1974. Introduction to the theory of statistics, third edition. McGraw-Hill Book Company. New York, New York.

Muir, W.D., D.M. Marsh, B.P. Sandford, S.G. Smith, and J.G. Williams. 2006. Post-hydropower system delayed mortality of transported Snake River stream-type chinook salmon: unraveling the mystery. Transactions of the American Fisheries Society 135:1523-1534.

Peterman, R.M. 1981. Form of random variation in salmon smolt-to-adult relations and its influence on production estimates. Canadian Journal of Fisheries and Aquatic Sciences. 38: 1113-1119.

Raymond, H. L. 1979. Effects of dams and impoundments on migrations of juvenile chinook salmon and steelhead from the Snake River, 1966 to 1975. Transactions of the American Fisheries Society 108(6):505–529.

Raymond, H. L. 1988. Effects of hydroelectric development and fisheries enhancement on spring and summer chinook salmon and steelhead in the Columbia River basin. North American Journal of Fisheries Management 8:1–24.

Seber, G. A. F. 1982. The estimation of animal abundance and related parameters, second edition. Chapman, London and Macmillan, New York.

Venerables, W.N., Smith, D.M., and R Development Core Team. 2010. An Introduction to R. Notes on R: A Programming Environment for Data Analysis and Graphics Version 2.11.1 (2010-05-31). http://www.r-project.org/

Acknowledgements

Appendix A. R code used to estimate power to detect a significant difference between smolt-to-adult ratios (SARs) of transported and in-river migrant juveniles

```
#R-code used to estimate the power of an experiment
#to detect a significant difference between the
#smolt-to-adult SARs of transported juveniles
#and the smolt-to-adult survival rate of juveniles
#migrating in-river.

#inputs
#SART -- SAR of transported juveniles
#NT -- the number of tagged transported juveniles
#SARM -- SAR of in-river juvenile migrants
#NM -- number of tagged in-river juvenile migrants

tmpower<-function(SART=.01,NT=1000,SARM=.02,NM=1000,alpha=0.05){
if((SART>1)|(SARM<0)){
  warning("The SARs are survival rates and so must be less than one")
  return(NULL)}
if((SARM>1)|(SARM<0)){
  warning("The SARs are survival rates and so must be less than one")
  return(NULL)}
if(NM<=0){
  warning("The number of juveniles released must exceed zero")
  return(NULL)}
if(NT<=0){
  warning("The number of juveniles released must exceed zero")
  return(NULL)}
if((alpha>1)|(alpha<0)){
  warning("The significance level (alpha) must be between zero and one")
  return(NULL)}
if((NT*SART<5)|(NM*SARM<5)){
 warning("An expected cell count is less than 5" )
 return(NULL)}
delta<-log(SART/SARM)
q<-qnorm(1-alpha/2)
tmvar<-(1-SART)/(SART*NT)+(1-SARM)/(SARM*NM)
se<-sqrt(tmvar)
power<-(1-pnorm(q*se,mean=delta,sd=se))+pnorm(-q*se,mean=delta,sd=se)
return(list(SART=SART,NT=NT,SARM=SARM,NM=NM,alpha=alpha,delta=delta,se=se,
cv=se/delta,power=power))
}
#outputs
```

#delta is the true log(SART/SARM)
#se is the standard error of the estimate
#cv is the CV of the estimate
#power is the probability of rejecting the null hypothesis of delta=0.