

## RRS2

A statistical tool to conduct an *a priori* analysis of the precision and accuracy of a maximum likelihood estimator of the relative reproductive success of hatchery-origin spawners with multiple years of data

Richard A. Hinrichsen

September 25, 2014

## SUMMARY

An *a priori* power analysis was conducted to aid the design of studies aimed at estimating the reproductive success of hatchery-born spawners relative to wild-born spawners (RRS) using parentage assignment. Power was defined as the probability of rejecting the null hypothesis of equal reproductive contributions of hatchery- and wild-born spawners. A maximum likelihood estimator of relative reproductive success and its variance were derived when multiple years of data are available. The software developed allows one to determine the sample sizes necessary to achieve a reliable estimate of RRS.

## INTRODUCTION

The presence of hatchery-born spawners is believed to influence reproductive success on the spawning grounds in many ways, yet its actual effect on recruits per spawner remains one of the greatest unknowns hampering the estimation of long-term population trends of salmonid populations. When hatchery-born fish are allowed to interbreed with wild-born fish on spawning grounds (i.e., the wild population is supplemented), the productivity of naturally spawning populations may be reduced, compromising conservation objectives (Reisenbichler and Rubin 1999). Even when hatchery spawners do not affect the reproductive success of wild-born spawners, hatchery-born spawners can still impede the estimation of trends in the wild portion of the population (Hinrichsen 2003; McClure et al. 2003). To see this, suppose that a large total recruitment per wild spawner is evident in the retrospective spawner series. If hatchery-reared spawners are as effective at producing adult progeny as wild fish, then the large total recruitment per wild spawner could have been due to a large input of hatchery-born spawners one generation ago, even though the trend in wild population abundance was downward. At the other extreme, if hatchery-born spawners produce no adult progeny (recruits), then a large total recruitment per wild-born spawner indicates a clear upward trend in wild population abundance. Reliably estimating the relative reproductive success of hatchery-born spawners (RRS) is also important for evaluating the effectiveness of captive rearing programs as a restoration strategy (Fleming 1994; Berejikian et al. 1997).

Because of its importance and advances in genetic tagging, studies of the relative reproductive success of hatchery-born spawners have increased over the last two decades. In early studies, Chilcote et al. (1986) used electrophoretic methodologies of Allendorf (1975) to estimate differential reproductive success of hatchery and wild summer-run steelhead (*Oncorhynchus mykiss*) under natural conditions. They concluded that the success of hatchery fish in producing smolt offspring was only 28% that of wild fish. This work was later expanded by Leider et al. (1990), who showed that relative reproductive success of naturally spawning hatchery steelhead compared with wild steelhead decreased from 0.75–0.788 at the subyearling stage to 0.108–0.129 at the adult stage. Later investigations by Garant et al. (2001) and Taggart et al. (2001), used a pedigree analysis whereby parents (spawners) were assigned to offspring (adult returns) using DNA analysis. This approach uses genetic differences among individuals to assess fitness. It differs from the approach of earlier experiments of Reisenbichler and McIntyre (1977), Leider et al. (1990), and Fleming et al. (2000) that used genetic markers that differentiated among groups of spawners, not individuals. The advantage of the group-based approach is that it requires few genetic markers; its disadvantage is that it requires good genetic differentiation between groups and, unlike the individual approach, does not provide individual fitness data, which can be used to estimate fitness with respect to individual traits such as run timing, size, or age (Morgan and Conner 2001). In individual-based experiments, the parents of the adult progeny (recruits) can be uniquely determined with high probability by using microsatellites, which serve as nonlethal, permanent, heritable markers (Ferguson and Danzmann 1998; Bernatchez and Duchesne 2000; Letcher and King 2001). Such experiments are expensive and exacting, requiring a high number of microsatellite loci and alleles per locus for reliable parentage assignment, but they may provide a relatively swift, accurate procedure for estimating the relative reproductive success of hatchery spawners. Recently, Hess et al. (2012), using a pedigree analysis found that mean RRS of hatchery-born females was 1.1 and that of hatchery-born males was 0.89.

The estimation of RRS is possible using the modern technique of genetic tagging called parentage-based tagging (PBT) of hatchery releases, which can be used to mark a high percentage of juveniles released. PBT involves genotyping hatchery broodstock (parents) and adding these genotypes to a database (Steele et al. 2011; Anderson and Garza 2005; Anderson and Garza 2006). Genotyped progeny of these parents collected as juveniles or adults can be assigned back to their parents, thus creating a tag identifying the origin of the parents (wild or hatchery). Software used to assign genotyped progeny to their parents, SNPPIT 1.0 developed by Anderson (2010), is available online at <http://www.mybiosoftware.com/population-genetics/6013>.

The objective of this documentation is to show how to estimate the *a priori* power of an experiment used to estimate relative reproductive success of hatchery-born spawners, RRS, where hatchery-born fish are allowed to spawn naturally with wild-born fish. I developed a likelihood-based, iterative procedure for estimating the relative reproductive success of hatchery-born spawners. This estimator forms the basis for a statistical test of whether recruits per hatchery-born spawner and recruits per wild-born spawner are equal (i.e.,  $RRS = 1$ ), which is the null hypothesis used in the power analysis.

## METHODS

There are two basic steps for collecting spawner data necessary to estimate the relative reproductive success of hatchery-origin spawners. First, genotype adults in a certain spawning area and note their origin: hatchery or wild. Enter the genotype and origin into a parent database. Second, take a random sample from the progeny of these genotyped adults and determine which female was its parent using parentage assignment. The progeny sampled may be juveniles or returning adults. Note the origin of the parent.

The method I develop in this documentation is a maximum likelihood technique (Mood et al. 1974). The variance of the MLE may be derived as the inverse of the Fisher Information Matrix, which is the negative of the expected value of the second derivative of the likelihood function (Mood et al. 1974). I begin by defining the assumptions (Table 1) and the variables used in the study, which are used to develop the probability model. For convenience, variable names and their definitions are given in Appendix A. Statistical code for the analysis, written in the R programming language, may be found in Appendix B.

Table 1.—Assumptions

|   |  |
|---|--|
| 1 | The assignment of a particular fish to a parent does not influence the assignment of any other fish to a parent (independence).  |
| 2 | Each sampled fish that is the offspring of a hatchery-origin female parent has the same probability of being assigned a hatchery-origin female parent (identically distributed). |
| 3 | Each sampled fish that is the offspring of a wild-origin female parent has the same probability of being assigned a wild-origin female parent (identically distributed).         |
| 4 | Whether an individual is sampled has no effect on the probability that another individual is sampled (independently drawn random sample).  |
| 5 | Relative reproductive success is constant across the years of study.   |

## Estimation

The assumptions in Table 1 allow one to express the total number of progeny assigned to a wild-born female parent,  $x$ , as a binomial random variable with probability density function:

$$f(x) = \prod_{t=1}^T \binom{n_t}{x_t} \left( \frac{S_{W,t}}{S_{W,t} + S_{H,t}\theta} \right)^x \left( \frac{S_{H,t}\theta}{S_{W,t} + S_{H,t}\theta} \right)^{n_t - x_t} = \prod_{t=1}^T \binom{n_t}{x_t} \frac{S_{W,t}^{x_t} S_{H,t}^{n_t - x_t} \theta^{n_t - x_t}}{(S_{W,t} + S_{H,t}\theta)^{n_t}}, \quad (1)$$

where  $n_t$  is the number of progeny sampled in year  $t$ ,  $T$  is the number of years progeny were collected during the study,  $x$  is the number of progeny assigned to a wild-origin female parent,

$S_W$  is the total number of wild-born female parents,  $S_H$  is the total number of hatchery-born female parents, and  $\theta$  is the RRS of hatchery- versus wild-origin spawners. Given the joint distribution of the observations in equation (1), it is now possible to form the log-likelihood function of the unknown parameter  $\theta$  by taking the natural log of the probability distribution and treating the result as a function of the parameter:

$$l(\theta) = \sum_{t=1}^T \left[ \log \binom{n_t}{x_t} + x_t \log(S_{W,t}) + (n_t - x_t) \log(S_{H,t} \theta) - n_t \log(S_{W,t} + S_{H,t} \theta) \right]. \quad (2)$$

To determine the MLEs of  $\theta$ , a numerical optimization routine, Fisher's Scoring Method (Jennrich and Sampson 1976), is used to maximize the log likelihood function given in equation (2). The maximization routine makes use of the derivative of the log likelihood function which is given by

$$\frac{dl}{d\theta} = \sum_{t=1}^T \left[ \frac{(n_t - x_t)}{\theta} - \frac{n_t S_{H,t}}{(S_{W,t} + S_{H,t} \theta)} \right]. \quad (3)$$

In the special case where there is a single brood year of data, the MLE may be found directly by setting the right hand side of equation (3) to zero

$$\hat{\theta} = \frac{(n_1 - x_1) S_{W,1}}{x_1 S_{H,1}}, \quad (4)$$

which is the number of progeny per hatchery spawner divided by the number of progeny per wild spawner.

The next step in deriving the theoretical formulas for precision of the MLEs is to derive the Fisher Information Matrix (which is a scalar in this case, not a full matrix). The inverse of the Fisher Information Matrix will supply the variance of the MLE of  $\theta$ . The Fisher Information

Matrix is the negative of the expected value of the second derivative of the likelihood function. The second derivative of the likelihood function is

$$\frac{d^2 l}{d\theta^2} = \sum_{t=1}^T \left[ -\frac{(n_t - x_t)}{\theta^2} + \frac{n_t S_{H,t}^2}{(S_{W,t} + S_{H,t} \theta)^2} \right]. \quad (5)$$

The Fisher Information Matrix, is given by

$$\mathbf{I} = -E \left( \frac{d^2 l}{d\theta^2} \right) = \sum_{t=1}^T \left[ \frac{n_t S_{H,t} S_{W,t}}{\theta (S_{W,t} + S_{H,t} \theta)^2} \right], \quad (6)$$

and therefore the variance of the MLE is given by

$$\text{var}(\hat{\theta}) = \left\{ \sum_{t=1}^T \left[ \frac{n_t S_{H,t} S_{W,t}}{\theta (S_{W,t} + S_{H,t} \theta)^2} \right] \right\}^{-1}. \quad (7)$$

The standard error of  $\hat{\theta}$  is then

$$SE(\hat{\theta}) = \sqrt{\text{var}(\hat{\theta})}, \quad (8)$$

and the coefficient of variation of  $\hat{\theta}$  is

$$CV(\hat{\theta}) = \frac{SE(\hat{\theta})}{\hat{\theta}}. \quad (9)$$

The MLE of RRS is calculated using Fisher's Scoring Method, which is similar to Newton's method for solving nonlinear equations (Press et al. 1992), but modified to use the expected Hessian (i.e.,  $-\mathbf{I}$ ) in place of the usual Hessian (Jennrich and Sampson 1976). In this case, the equation describing the iteration used in Fisher's Scoring Method becomes

$$\hat{\theta}_{l+1} = \hat{\theta}_l + \mathbf{I}^{-1}(\hat{\theta}_l) \nabla l(\hat{\theta}_l). \quad (10)$$

where  $\hat{\theta}_{l+1}$  is the maximum likelihood estimator of  $\theta$  on iteration  $l+1$ ,  $\hat{\theta}_l$  is the maximum likelihood estimator of  $\theta$  on iteration  $l$ ,  $\mathbf{I}^{-1}(\hat{\theta}_l)$  is the inverse of the Fisher Information Matrix evaluated at  $\hat{\theta}_l$ , and  $\nabla l(\hat{\theta}_l)$  is the gradient or score function, which represents the first derivative of the log-likelihood function evaluated at  $\hat{\theta}_l$ . The algorithm proceeds by making a first initial guess at the MLE, then iterating equation (10) until the estimate changes by no more than a given error tolerance (e.g.,  $10^{-5}$ ). As an initial guess, we use the sample mean of the single-year estimate of RRS given by equation (4).

## Monte Carlo Simulation

As an alternative approach to estimating the precision of  $\hat{\theta}$ , Monte Carlo simulation is used. Additionally, this approach yields an estimate of accuracy (bias). Monte Carlo estimates of precision and accuracy do not rely on asymptotic theory as in the previous sections. That is, Monte Carlo estimates of precision and bias will work with small sample sizes ( $n$ ). The Monte Carlo method proceeds by drawing NSIM random samples from the probability distribution of  $x$  given by equation (1), then calculating the MLE of RRS for each Monte Carlo sample. This yields NSIM replications of the MLE, denoted by  $\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_{NSIM}^*$ . The Monte Carlo estimate of the standard error of the MLE of the proportion of hatchery-origin spawners is then equal to the square root of the sample variance of the Monte Carlo replications:

$$SE^*(\hat{\theta}) = \sqrt{\sum_{k=1}^{NSIM} \frac{(\hat{\theta}_k^* - \bar{\hat{\theta}}^*)^2}{NSIM - 1}}, \quad (11)$$

where  $\bar{\hat{\theta}}^*$  is the sample mean of the Monte Carlo replications. The coefficient of variation of  $\hat{\theta}$  is

$$CV^*(\hat{\theta}) = \frac{SE^*(\hat{\theta})}{\hat{\theta}}. \quad (12)$$

The relative bias is calculated as

$$bias^*(\hat{\theta}) = \frac{\bar{\hat{\theta}}^* - \theta}{\theta}. \quad (13)$$

### Estimation of log(RRS)

The estimation outlined above was for the relative reproductive success RRS itself. The webtool RRS focusses on log(RRS), not on RRS. However, it is a simple matter to calculate the MLE of  $\delta = \log(\text{RRS})$  and its variance using the MLE and variance of RRS itself. Using maximum likelihood theory, it is easily demonstrated that

$$\hat{\delta} = \log(\hat{\theta}). \quad (14)$$

The variance of log(RRS) is found using the Delta method (Seber 1982), which uses a Taylor series approximation to find the variance of a transformed random variable:

$$\text{var}(\hat{\delta}) = \text{var}(\log(\hat{\theta})) = \frac{\text{var}(\hat{\theta})}{\theta^2}. \quad (15)$$

When using Monte Carlo simulations to estimate SE, CV, and relative bias, simply substitute  $\delta$  for  $\theta$  in the equations (11)-(13).

## Power

Power is the probability of rejecting the null hypothesis of  $\log(\text{RRS}) = 0$  when the actual  $\log(\text{RRS})$  is an alternative value. Power depends on the true value of  $\log(\text{RRS})$ , the probability of a type I error (usually called the alpha value), and the standard error of the estimator. By maximum likelihood theory, the MLE of  $\delta$  is asymptotically normally distributed. Thus, power may be written as

$$\Pi(\delta) = \Phi(-z_{\alpha/2} - \delta / \text{se}(\hat{\delta})) + 1 - \Phi(z_{\alpha/2} - \delta / \text{se}(\hat{\delta})); \quad (16)$$

where  $\Phi(z)$  is the cumulative distribution function of a random variable that follows a standard normal distribution (a normal distribution with mean zero and standard deviation 1), and  $z_{\alpha/2}$  is the critical value such that  $\alpha / 2$  probability lies to the right of the value  $z_{\alpha/2}$  in a standard normal distribution.

The equations for power may be used to formulate an experimental design that has a high chance of delivering the best accuracy and precision given some design constraints. The power, as measured by SE of  $\hat{\delta}$  depends crucially on sample size ( $n$ ), which may be manipulated to give lowest possible SE. As  $n$  increases, precision will always increase. Investigators often choose a design such that power of 0.8 is achieved (e.g, Peterman and Bradford 1987, Liermann and Roni 2008).

## APPLICATIONS

### Single year of data.

To illustrate the use of power to design a RRS study, consider the following example. Assume that there are equal numbers of hatchery-origin and wild-origin spawners and that there are 200 of each (i.e.,  $S_H = S_W = 200$ ). Suppose that our goal is to reliably detect a difference in reproductive success between hatchery- and wild-origin spawners when RRS is less than 0.8 or



greater than 1.2. What sample size should be used? To answer this question, I varied the sample size  $n$  from 100 to 1000, used a target of 80% power (probability of rejecting the null hypothesis that  $\text{RRS} = 1$ ) and fixed the probability of Type I error is fixed at  $\alpha = 0.05$ . The true RRS was set to the alternative values of 0.8 and 1.2 and power was plotted as a function of sample size,  $n$ . Note that when the true values of RRS are 0.8, and 1.2, that the true values of  $\delta$  are  $\log(0.8)$  and  $\log(1.2)$ , respectively. The results of this analysis are shown in Figure 1. The plot shows that the sample size,  $n$ , must be at least 953 to achieve power of 0.80 when  $\text{RRS} = 1.2$ . When  $\text{RRS}=0.8$ ,  $n=639$  will suffice.

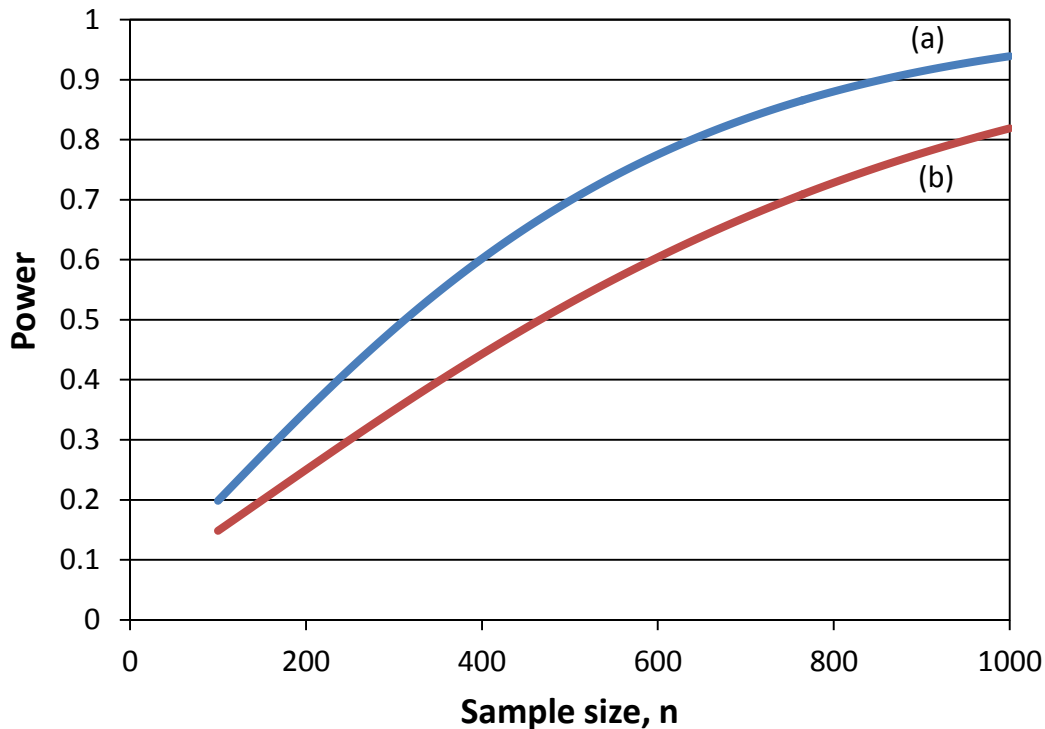


Figure 1. Plots of power for two different design scenarios (a)  $\text{RRS} = 0.8$  (b)  $\text{RRS}=1.2$ .

### Multiple years of data.

Consider an example where there are multiple years of data. In each year, assume there are equal numbers of hatchery-origin and wild-origin spawners and that there are 200 of each (i.e.,  $S_{H,t} = S_{W,t} = 200$ ). Assume that the true value of RRS is 0.80 and that the number of years of study is allowed to vary from 1 to 10 in steps of 1; that is, the alternative values of  $T$  are  $T=1,2,\dots,10$ . Assuming that number of offspring assigned to each of these years is given by  $n_t = 100$ . How many years of data need to be collected in order to achieve a power of 0.80? A plot of power versus number of brood years of study shows that a power of 0.80 can be achieved with 7 years (Figure 2).

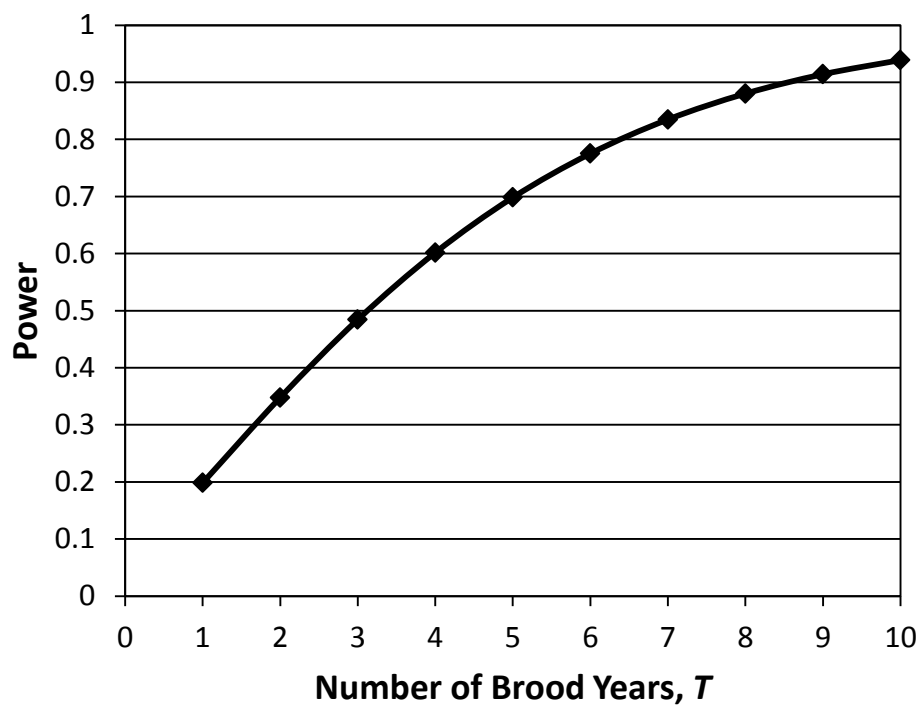


Figure 2. Plot of power varying the number of brood years of study,  $T$ .

## **ACKNOWLEDGEMENTS**

This work was supported by Bonneville Power Administration. Brian Maschhoff implemented this statistical analysis as a web-based tool at [www.onefishtwofish.net](http://www.onefishtwofish.net). Amber Parsons reviewed this documentation and tested the web-based tool RRS2. The views expressed are solely those of the author and are not intended to represent the views of any organization with which the author is affiliated.

## REFERENCES

- Allendorf, F.W. 1975. Genetic variability in a species possessing extensive gene duplication: genetic interpretation of duplicate loci and examination of genetic variation of populations of rainbow trout. Ph.D. thesis, University of Washington, Seattle, WA.
- Anderson, E.C. 2010. Computational algorithms and user-friendly software for parentage-based tagging of Pacific salmonids. Final report submitted to the Pacific Salmon Commission's Chinook Technical Committee (US Section). 46 p. Available: [http://swfsc.noaa.gov/uploadedFiles/Divisions/FED/Staff\\_Pages/Eric\\_Anderson/PBT\\_PS\\_C\\_final\\_report.pdf](http://swfsc.noaa.gov/uploadedFiles/Divisions/FED/Staff_Pages/Eric_Anderson/PBT_PS_C_final_report.pdf).
- Anderson, E. C., and J. C. Garza. 2005. A description of full parental genotyping. Report submitted to the Pacific Salmon Commission, Vancouver, British Columbia. 11p. Available: <http://swfsc.noaa.gov/publications/FED/00675.pdf>.
- Anderson E. C., and J. C. Garza. 2006. The power of single-nucleotide polymorphisms for large-scale parentage inference. *Genetics* 172: 2567–2582.
- Berejikian, B.A., Tezak, E.P., Schroder, S.L., Knudsen, C.M., and J.J. Hard. 1997. Reproductive behavioral interactions between wild and captively reared coho salmon (*Oncorhynchus kisutch*). *ICES Journal of Marine Science* 54: 1040–1050.
- Bernatchez, L., and P. Duchesne. 2000. Individual-based genotype analysis in studies of parentage and population assignment: how many loci, how many alleles? *Canadian Journal of Fisheries and Aquatic Sciences* 57: 1–12.
- Chilcote, M.W., Leider, S.A., and J.J. Loch. 1986. Differential reproductive success of hatchery and wild summer-run steelhead under natural conditions. *Transactions of the American Fisheries Society* 115: 726–735.
- Ferguson, M.M., and R.G. Danzmann. 1998. Role of genetic markers in fisheries and aquaculture: useful tools or stamp collecting? *Canadian Journal of Fisheries and Aquatic Sciences* 55: 1553–1563.
- Fleming, I.A. 1994. Captive breeding and the conservation of wild salmon populations. *Conservation Biology* 8: 886–888.
- Fleming, I.A., Hindar, K., Mjølnerod, I.B., Jonsson, B., Balstad, T., and A. Lamberg. 2000. Lifetime success and interactions of farm salmon invading a native population. *Proceedings of the Royal Society B: Biological Sciences* 267: 1517–1524.
- Garant, D., Dodson, J.J., and L. Bematchez. 2001. A genetic evaluation of mating system and determinants of individual reproductive success in Atlantic salmon (*Salmo salar L.*). *Journal of Heredity* 92: 137–145.

- Hess, M. A., C. D. Rabe, J. L. Vogel, J. J. Stephenson, D. D. Nelson, and S. R. Narum. 2012. Supportive breeding boosts natural population abundance with minimal negative impacts on fitness of a wild population of Chinook Salmon. *Molecular Ecology* 21:5236–5250.
- Hinrichsen, R. A. 2003. The power of experiments for estimating the relative reproductive success of hatchery-origin spawners. *Canadian Journal of Fisheries and Aquatic Sciences* 60:864–872.
- Jennrich, R. I., and P. F. Sampson. 1976. Newton-Raphson and related algorithms for maximum likelihood variance component estimation. *Technometrics* 18:11-17.
- Letcher, B.H., and T.L. King. 2001. Parentage and grandparentage assignment with known and unknown matings: application to Connecticut River Atlantic salmon restoration. *Canadian Journal of Fisheries and Aquatic Sciences* 58: 1812–1821.
- Leider, S.A., Hulett, P.L., Loch, J.L., and M.W. Chilcote. 1990. Electrophoretic comparison of the reproductive success of naturally spawning transplanted and wild steelhead trout through the returning adult stage. *Aquaculture* 88: 239–252.
- Liermann, M., and P. Roni. 2008. More sites or more years? Optimal study design for monitoring fish response to watershed restoration. *North American Journal of Fisheries Management*. 28:935-943.
- McClure, M.M., E.E. Holmes, B.L. Sanderson, and C.E. Jordan. 2003. A large-scale multispecial status assessment: anadromous salmonids in the Columbia River basin. *Ecological Applications* 13:964-989.
- Mood, A. M., F. A. Graybill, and D. C. Boes. 1974. Introduction to the theory of statistics, 3rd edition. McGraw-Hill, New York.
- Morgan, M.T., and J.K. Conner. 2001. Using genetic markers to directly estimate male selection gradients. *Evolution*, 55: 272–281.
- Peterman, W. M., and M. J. Bradford. 1987. Statistical power of trends in fish abundance. *Canadian Journal of Fisheries and Aquatic Sciences* 44: 1879:1889.
- Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. 1992. Numerical recipes in FORTRAN. Cambridge University Press, New York.
- Reisenbichler, R.R., and J.D. McIntyre. 1977. Genetic differences in growth and survival of juvenile hatchery and wild steelhead trout, *Salmo gairdneri*. *Journal of the Fisheries Research Board of Canada* 34: 123–128.

- Reisenbichler, R.R., and S.P. Rubin. 1999. Genetic changes from artificial propagation of Pacific salmon affect the productivity and viability of supplemented populations. *ICES Journal of Marine Science*. 56: 459–466.
- Seber, G. A. F. 1982. The estimation of animal abundance and related parameters, second edition. Chapman, London and Macmillan, New York.
- Steele, C., M. Ackerman, J. McCane, M. Campbell, M. Hess, N. Campbell, and S. Narum. 2011. Parentage-based tagging of Snake River hatchery steelhead and Chinook salmon. 2010 Annual Report. Idaho Department of Fish and Game and Columbia River Inter-Tribal Fish Commission IDFG Report Number 11-111. June 2011. Available: <https://collaboration.idfg.idaho.gov/FisheriesTechnicalReports/Res11-111Steele2010%20Parentage%20Based%20Tagging%20Snake%20River%20Steelhead%20Salmon.pdf>
- Taggart, J.B., McLaren, I.S., Hay, D.W., Webb, J.H., and A.F. Youngson. 2001. Spawning success in Atlantic salmon (*Salmo salar* L): a long-term DNA profiling-based study conducted in a natural stream. *Molecular Ecology* 10: 1047–1060.

## APPENDIX A. NAMES OF VARIABLES

Table A.1.—Variables.

| R code     | Mathematical derivation | Definition   |
|------------|-------------------------|--|
| Sw         | $S_{W,t}$               | Number of wild-born female spawners (one for each year $t$ ).  |
| Sh         | $S_{H,t}$               | Number of hatchery-born female spawners (one for each year $t$ ).  |
| n          | $n_t$                   | The sample size of progeny that are genotyped and assigned back to spawning females (one number for each year $t$ ). |
| Rw         | $x_t$                   | Number of sampled progeny that are assigned to a wild-born mother (one for each year $t$ ).                          |
| theta      | $\theta$                | Relative reproductive success (RRS) of hatchery-origin spawners  |
| delta      | $\delta$                | The natural log of RRS   |
| n.a.       | $l(\theta)$             | Log-likelihood function of RRS   |
| SE.delta   | $SE(\hat{\delta})$      | Standard error of the estimate of log(RRS)   |
| CV.delta   | $CV(\hat{\delta})$      | Coefficient of variation of the estimate of log(RRS)   |
| BIAS.delta | $bias^*(\hat{\delta})$  | Monte Carlo relative bias which is the bias divided by the true value of $\hat{\delta}$ .                            |

## APPENDIX B. R-CODE

Table B.1. R-code

---

```

#Program to calculate power of an experiment aimed at estimating the relative reproductive success (RRS)
#of hatchery-origin spawners using multiple brood years.
#AUTHOR: Richard A. Hinrichsen, Ph.D.
#DATE: 6-21-2014
#FILE: rrs2-6-21-2014.s
#input variables
#Sw number of wild-origin spawning females (one for each year)
#Sh number of hatchery-origin spawning females (one for each year)
#n is the sample size of progeny (one for each year)
#delta is log(RRS)

#top level function
rrs2.main<-
function(Sw=c(200,200),Sh=c(200,200),n=c(800,200),alpha=0.05,delta=log(.8),MONTE=FALSE,NSIM=1000){
  check.inputs(Sw,Sh,n,alpha,delta,MONTE,NSIM)
  if(!MONTE){res<-rrs2(Sw=Sw,Sh=Sh,n=n,alpha=alpha,delta=delta)}
  if(MONTE){res<-rrs2.monte(Sw=Sw,Sh=Sh,n=n,alpha=alpha,delta=delta,NSIM)}
  final.res<-list(MONTE=res$MONTE,
    NSIM=res$NSIM,
    Sw=res$Sw,
    Sh=res$Sh,
    n=res$n,
    alpha=res$alpha,
    delta=res$delta,
    SE.delta=res$SE.delta,
    CV.delta=res$CV.delta,
    BIAS.delta=res$BIAS.delta,
    power=res$power)
  return(final.res)
}

#check that inputs are valid
check.inputs<-function(Sw,Sh,n,alpha,delta,MONTE,NSIM){
  if(!is.logical(MONTE))stop("MONTE must be TRUE or FALSE")
  if(MONTE){
    if(floor(NSIM)!=NSIM){stop("NSIM must be a positive integer")}
    if(NSIM<=0){stop("NSIM must be a positive integer")}}
  #check dimension of inputs
  k1<-length(Sw);k2<-length(Sh);k3<-length(n)
  mytest<-abs(k1-k2)+abs(k2-k3)
  if(mytest>0) stop("dimensions of Sw, Sh, and n must match")
  if(!is.numeric(Sw)){stop("Sw must be a number")}
  if(!is.numeric(Sh)){stop("Sh must be a number")}
  if(!is.numeric(n)){stop("n must be a number")}
  if(sum(floor(n)-n)){stop("Each n must be a positive integer")}
  if(sum(n<=0)){stop("Each n must be a positive integer")}
  if(sum(floor(Sw)-Sw)){stop("Each Sw must be a positive integer")}
  if(sum(Sw<=0)){stop("Each Sw must be a positive integer")}
  if(sum(floor(Sh)-Sh)){stop("Each Sh must be a positive integer")}
  if(sum(Sh<=0)){stop("Each Sh must be a positive integer")}

```

---



---

```

if(alpha<=0){stop("alpha must be between zero and 1.0")}
if(alpha>=1){stop("alpha must be between zero and 1.0")}
if(!is.double(alpha)){stop("alpha must be a double")}
if(!is.double(delta)){stop("delta must be a double")}
return(NULL)
}

#This uses theoretical formulas from Hinrichsen (2003)
rrs2<-function(Sw,Sh,n,alpha,delta){
  theta<-exp(delta)
  q<-qnorm(1-alpha/2)
  INF<-sum((n*Sh*Sw)/(theta*(Sw+Sh*theta)^2))
  thetavar<-1/INF
  deltavar<-thetavar/(theta*theta)
  se<-sqrt(deltavar)
  power<-(1-pnorm(q*se,mean=delta,sd=se))+pnorm(-q*se,mean=delta,sd=se)
  myres<-list(MONTE=FALSE,
    NSIM=NA,
    Sw=Sw,
    Sh=Sh,
    n=n,
    alpha=alpha,
    delta=delta,
    SE.delta=se,
    CV.delta=se/delta,
    BIAS.delta=NA,
    power=power)
  return(myres)
}

#return MLE of delta and its SE
#use Fisher's Scoring Method
get.estimate<-function(Sw,Sh,n,Rw){
  NTRIAL<-100
  theta<-mean(Sw*(n-Rw)/(Rw*Sh))
  tol<-1.e-5

  for(ii in 1:NTRIAL){
    INF<-sum((n*Sh*Sw)/(theta*(Sw+Sh*theta)^2))
    df<-sum((n-Rw)/theta-n*Sh/(Sw+Sh*theta))
    delx<-(1/INF)*df
    theta<-theta+delx
    errx<-sum(abs(delx))/abs(theta)
    if(errx<=tol)break
  }
  if(ii==NTRIAL){
    warning("maximum number of iterations was reached")
    return(list(delta=NA,se=NA))
  }
  delta<-log(theta)
  thetavar<-1/INF
  deltavar<-thetavar/(theta*theta)
  se<-sqrt(deltavar)
  return(list(delta=delta,se=se))
}

```

---

---

```

}

#calculate SE and statistical power using Monte Carlo simulation
rrs2.monte<-function(NSIM,Sw,Sh,n,alpha,delta){
  theta<-exp(delta)
  q<-qnorm(1-alpha/2)
  prob<-Sw/(Sw+Sh*theta)
  nprob<-length(prob)
  Rw<-rep(NA,nprob)
  deltas<-rep(NA,NSIM)
  ses<-rep(NA,NSIM)

  for(ii in 1:NSIM){
    Rw<-rep(NA,nprob)
    for(jj in 1:nprob){
      Rw[jj]<-rbinom(n=1,size=n[jj],prob=prob[jj])
    }
    res<-get.estimate(Sw,Sh,n,Rw)
    deltas[ii]<-res$delta
    ses[ii]<-res$se
  }

  power<-abs(deltas/ses)>q
  power<-sum(power)/NSIM
  se<-sqrt(var(deltas,na.rm=T))
  mymean<-mean(deltas,na.rm=T)
  BIAS.delta<-(mymean-delta)/delta
  myres<-list(MONTE=TRUE,
    NSIM=NSIM,
    Sw=Sw,
    Sh=Sh,
    n=n,
    alpha=alpha,
    delta=delta,
    SE.delta=se,
    CV.delta=se/delta,
    BIAS.delta=BIAS.delta,
    power=power)
  return(myres)
}

```

---