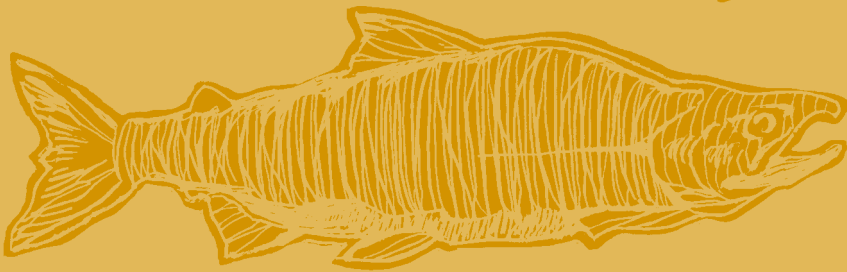# SampleSize **3** USER'S MANUAL

SAMPLE SIZE CALCULATIONS FOR FISH AND WILDLIFE SURVIVAL STUDIES

COLUMBIA BASIN RESEARCH

SCHOOL OF AQUATIC AND FISHERY SCIENCES

UNIVERSITY OF WASHINGTON

Program SampleSize 3

Sample Size Calculations for Fish and Wildlife Survival Studies

Developed by:
Trish Lockhart, Jim Lady, and John R. Skalski
Columbia Basin Research
School of Aquatic & Fishery Sciences
University of Washington
1325 Fourth Avenue, Suite 1515
Seattle, Washington 98101-2540

February 2023

Acknowledgments

Table of Contents

Chapter 1: Introduction

Release-recapture studies are commonly used to estimate survival-related parameters in fish and wildlife studies. In designing such a study, one must determine the quantity of tagged individuals needed to achieve the desired precision without wasting resources by using larger release sizes than necessary. Program SampleSize was developed to fill this need. It allows the user to input a range of values for a release size or estimated parameters and calculate the resulting precision of survival projections. It also graphically displays the calculated precision as a function of the range of values, allowing the user to observe the sensitivity of the precision over the range.

There are six study design types implemented in Program SampleSize:
1. Single Release
2. Paired Release
3. Virtual Paired Release
4. Ricker Two-Release
5. ViRDCt (Full model)
6. ViRDCt (Reduced model)

Note that previous versions of Program SampleSize included two study types not included in the current version: (1) transport in-river ratio and (2) balloon-tag passage survival estimate. These study types are no longer explicitly included because the transport in-river ratio study is an example of a Ricker study, and the variance for a balloon-tag study is the same as the variance for the Ricker model.

Chapter 2 will describe the basics in using Program SampleSize, version 3.0 The subsequent chapters will describe the specifics in using SampleSize for the specific models mentioned above.

SampleSize User's Manual

Chapter 2: Using Program SampleSize

Figure 1 shows Program SampleSize at startup, displaying the "Welcome" screen.
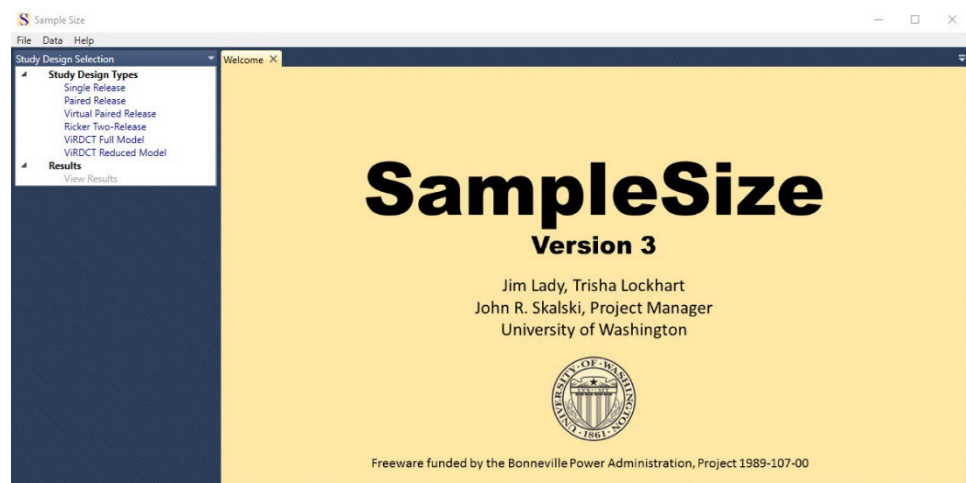
Figure 1. Program SampleSize at startup, displaying the Welcome screen

On the left side under the heading "Study Design Selection," the user can click on any one of the four study design types. Clicking on one of the selections will bring up a tabbed dialog for that study design.

Figure 2 shows the program after clicking on "Single Release," under "Study Design Types," bringing up the tabbed window for a single release release-recapture study.
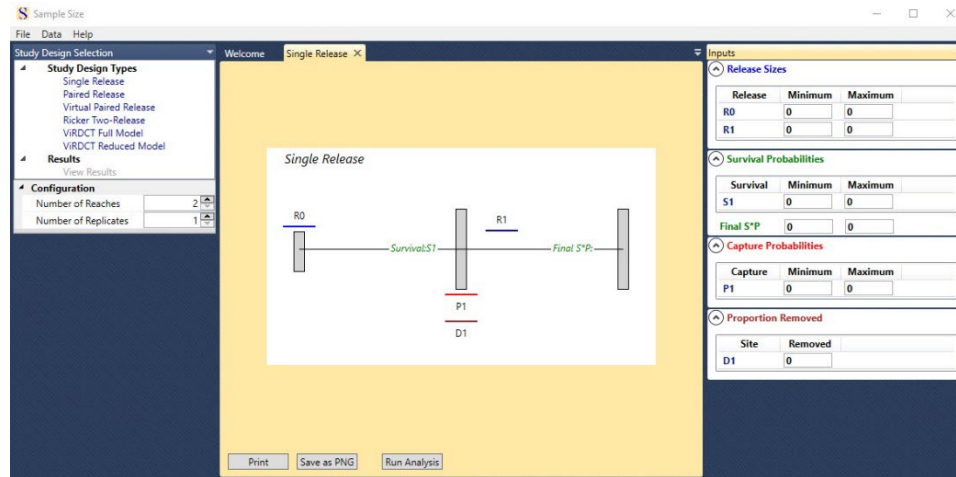
Figure 2. Program SampleSize with Single Release study design selected

The user may have more than one study tab open at a time. In Figure 3, the user has both a Single Release study and a "Paired Release" study open, with the Paired Release tab currently active.

Figure 3. Program SampleSize with multiple study designs open

The user may close the current tab by clicking the "X" on the tab. They may also open the context menu for other options by right-clicking on the tab as shown in Figure 4.
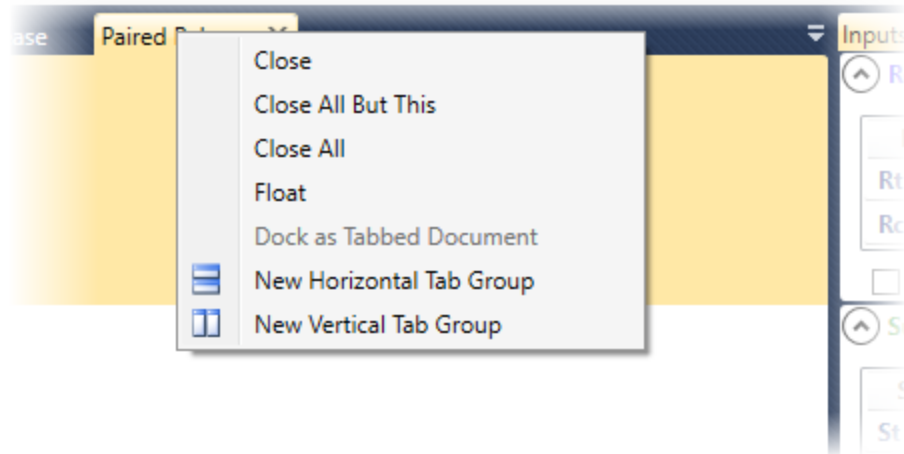
Figure 4. The context menu for the study design tabs

On each study tab, there are three buttons at the bottom (shown in Figure 5) that allow the user to print the study diagram, save it as a PNG file, or run the analysis once all inputs have been entered.
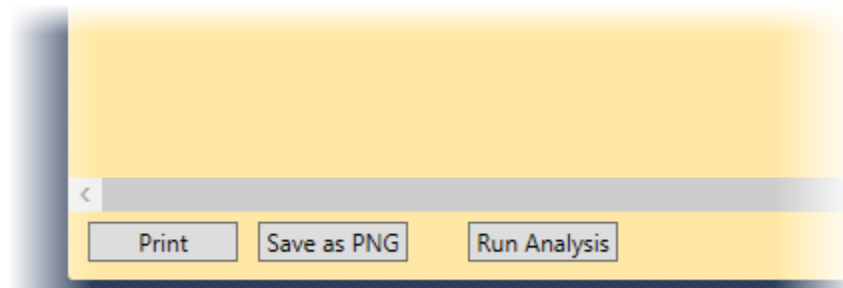
Figure 5. The buttons at the bottom of a study design diagram for printing or saving the diagram and for running the analysis

Once a study tab is active, one or two additional areas appear on the SampleSize dialog: (1) a "Configuration" section, if applicable, under the Study Design section on the left and (2) the "Inputs" section on the right side.

Figure 6 shows the Configuration section for a single-release study. In this case, the user may specify the "Number of Reaches" and the "Number of Replicates" (explained in Chapter 3).

Figure 6. The Configuration menu for a Single Release study

2.1 Input Section

Figure 7 shows the "Inputs" section for a Single Release study. In this section, the user inputs all release sizes, estimates of survival probabilities, capture probabilities, and the proportion removed from each site (when applicable). Release sizes must be integers, and survival and capture probabilities as well as proportions removed must be values between 0.0 and 1.0.
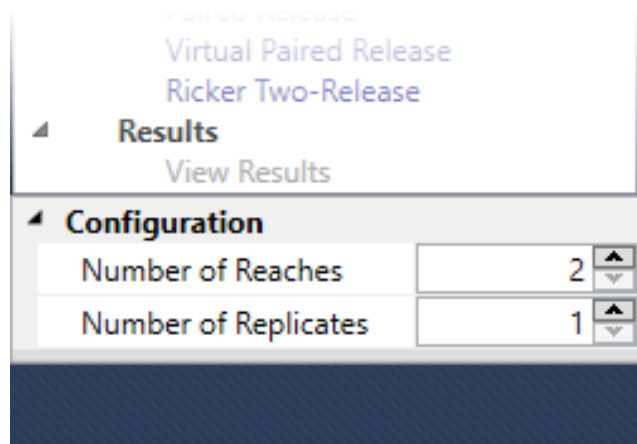
Most inputs have a place to enter both a minimum and a maximum value — $R0$, $R1$, $S1$, and $P1$ in Figure 7. This allows the user to explore the sensitivity of the resulting precision to the input value with a range. Note that only one input can have a range; when both a minimum and maximum for a given input are entered, the "Maximum" input box disappears for the other inputs.

In order to cancel a range for an input, the user must enter a "0" (zero) in the Maximum entry box for the input.

Figure 7. The Inputs section for a Single Release study

2.2 File Operations

The "File" menu on the upper left of the SampleSize Program allows the user to save the current configuration for all study designs that are open and to restore them from a file at a later time.

2.3 Analysis Results

2.3.1 Analysis with No Range

Figure 8 shows the study diagram and input section filled in for a Single Release study without a range specified. Note that the input values are displayed on the diagram upon being entered by the user.
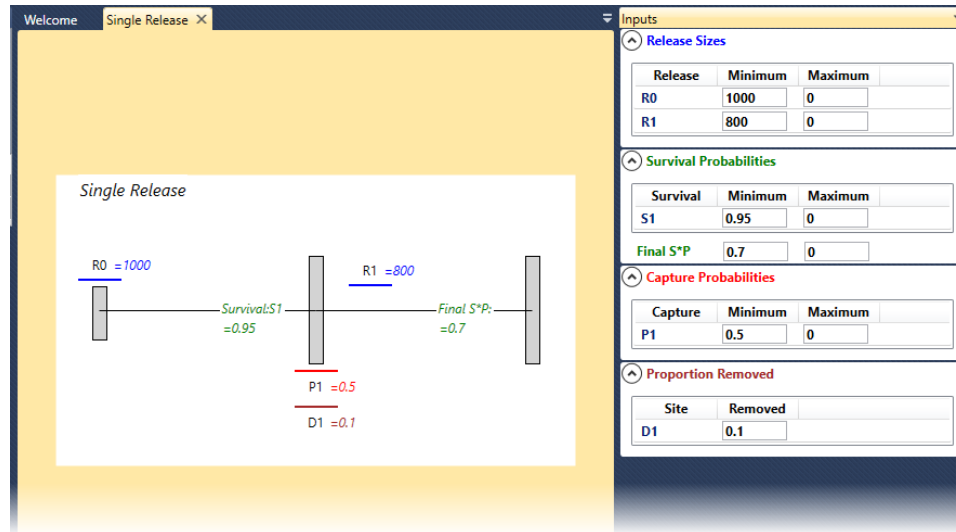
Figure 8. The study diagram and input section with no range specified

When the "Run Analysis" button is pressed, a report appears in a separate window as shown in Figure 9 that lists all of the user-supplied inputs and a results table with:

1. Standard error of the parameter of interest (*S*1 in this case)
2. One-half of a 95% confidence interval
3. One-half of a 90% confidence interval

At the bottom of the report are two buttons, allowing the user to print the report or save it as a comma-separated value (CSV) file.

2.3.2 Analysis with a Range

Figure 10 shows the inputs with a range entered for R0; when the range is entered, the Maximum entry box for the other inputs go away.

When the Run Analysis button is pressed, a results window appears as shown in Figure 11. There are two tabs on the results window when a range is entered: (1) the "Results Table" and (2) the "Results Graph."

Figure 9. The Results Table window when no range is specified
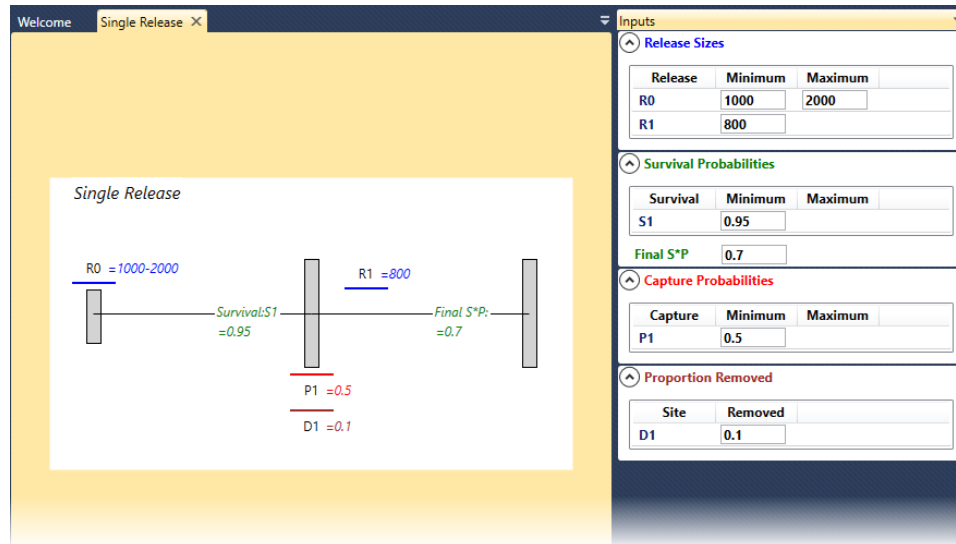
Figure 10. The Single Release study diagram and Inputs section when a range is specified
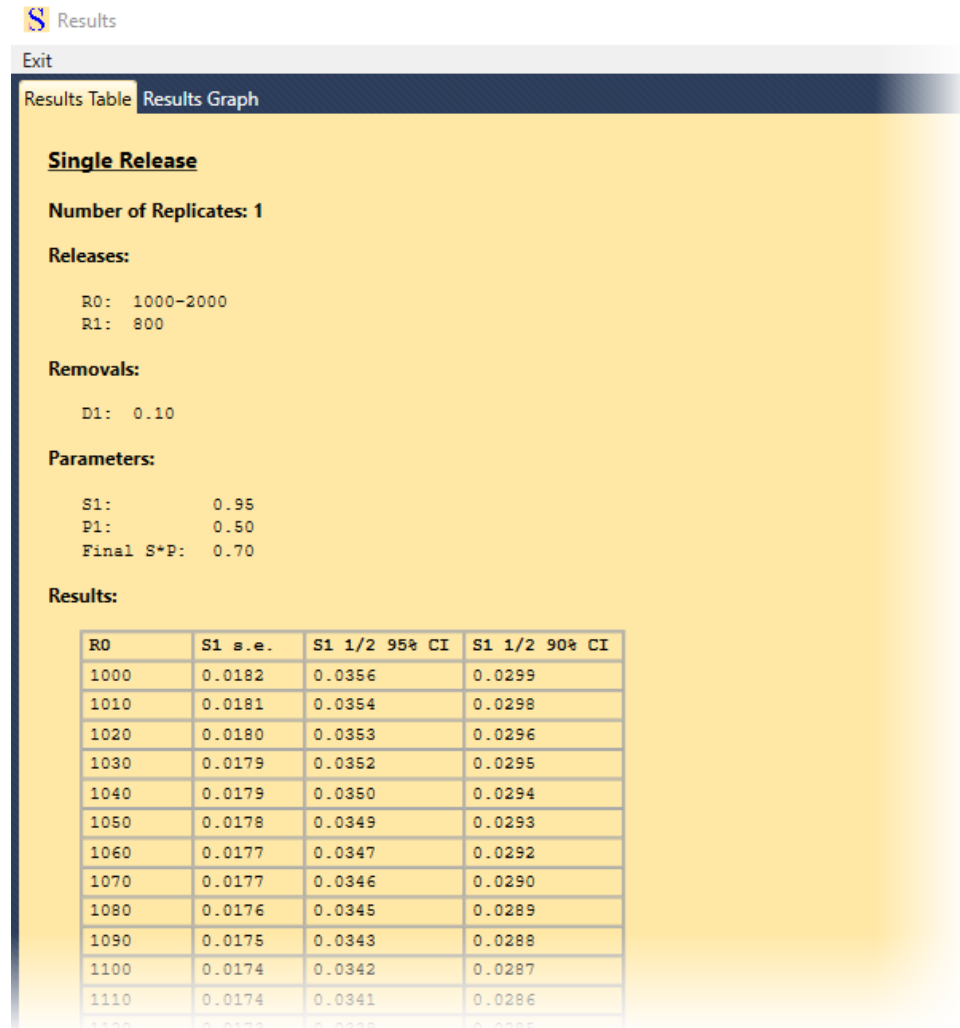
Figure 11. The Results Table window when a range is specified

Results Table

The Results Table tab is active by default when the results window appears. It is similar to the results report when no range is entered, except there is an additional column in the table for the values of the input with a range. The Results Table shows the results for each value of the input along one-hundred evenly spaced values along the range. In the table in Figure 11, a range of 1,000–2,000 was entered for *R*0, thus the results table shows results for R0 in values that increment by ten.

Results Graph

Figure 12 shows the Results Graph tab. The input with the range (R0 in this case) is on the *X*-axis and one-half of a confidence interval (CI) is on the *Y*-

axis. The red line (on top in Figure 12) represents a 95% CI and the blue line represents a 90% CI. When the user moves the mouse along the graph horizontally, a vertical line follows the mouse and shows CI values for 95% and 90%. The user can click on any point on the graph, and a fixed line will appear showing the CI values at that point.

In Figure 12, the user has clicked on the value $R0 = 1400$, and a line has appeared showing one-half of a 95% CI as 0.0309 and one-half of a 90% CI as 0.0259. The mouse currently rests over the value $R0 = 1710$.
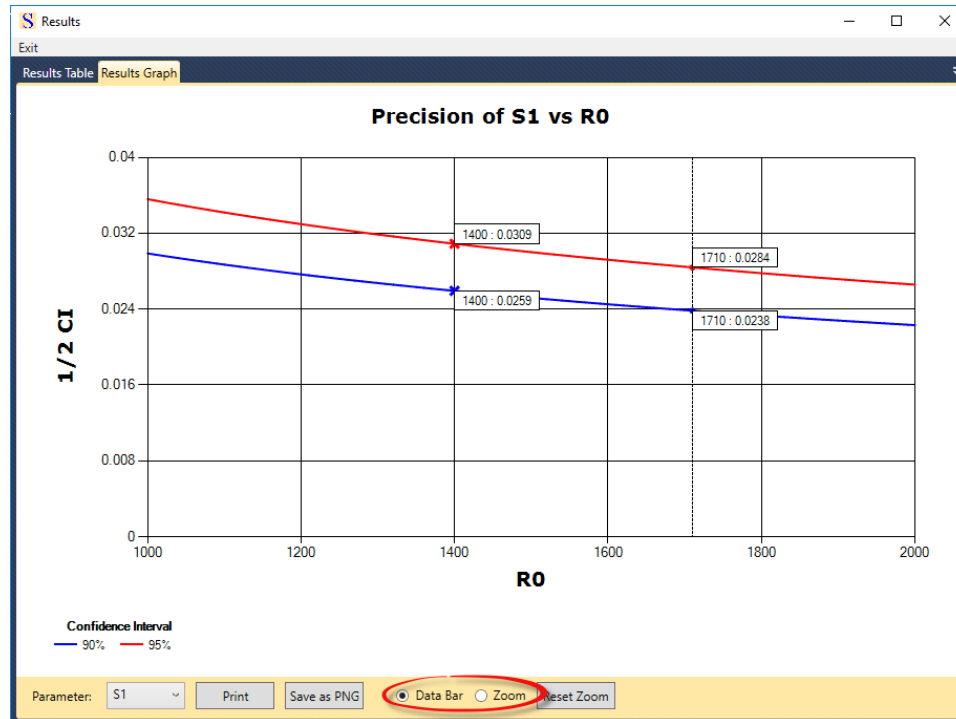
Figure 12. The Results Graph

At the bottom of the graph are two option buttons (circled in Figure 12) labeled "Data Bar" and "Zoom." The Data Bar button is selected by default, enabling the behavior just described. If the user clicks on Zoom, the user can hold the left mouse button and drag over a portion of the graph in order to focus on that section. In Figure 13, the user has dragged the mouse over a portion of the graph indicated in gray between $R0 = 1200$ and $R0 = 1600$. When the mouse button is released, the Results Graph shows only that portion, as shown in Figure 14. The user can restore the graph to its initial state by clicking on the "Reset Zoom" button at the bottom.
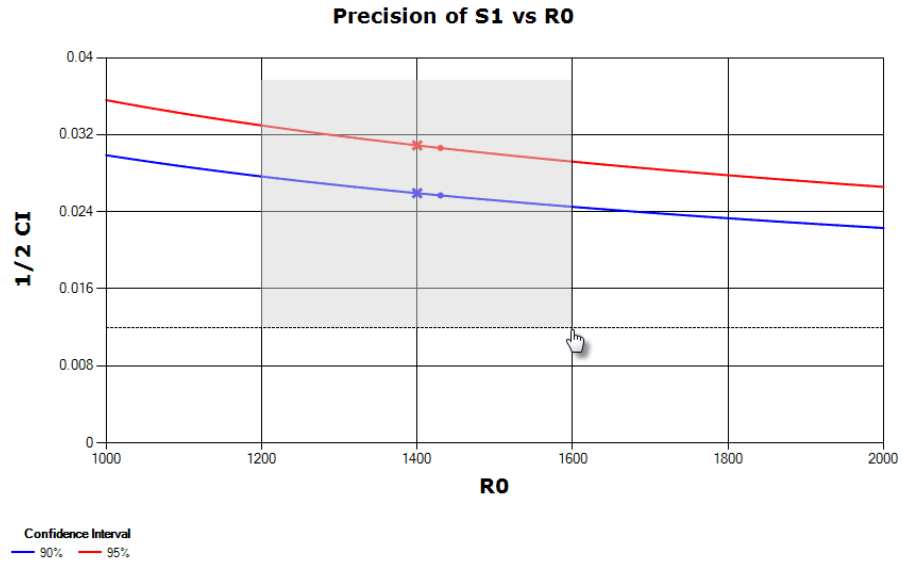
Figure 13. The user selecting a section of the graph to zoom in on

Figure 14. The results graph after the user the user has zoomed in on a section

At the bottom are buttons to print the graph or to save it as a PNG file. There is also a box labeled "Parameter" showing the parameter of interest that the CIs apply to. For a Single Release design only, this is a combo box where the user can select the survival parameter of interest (described in Chapter 3).
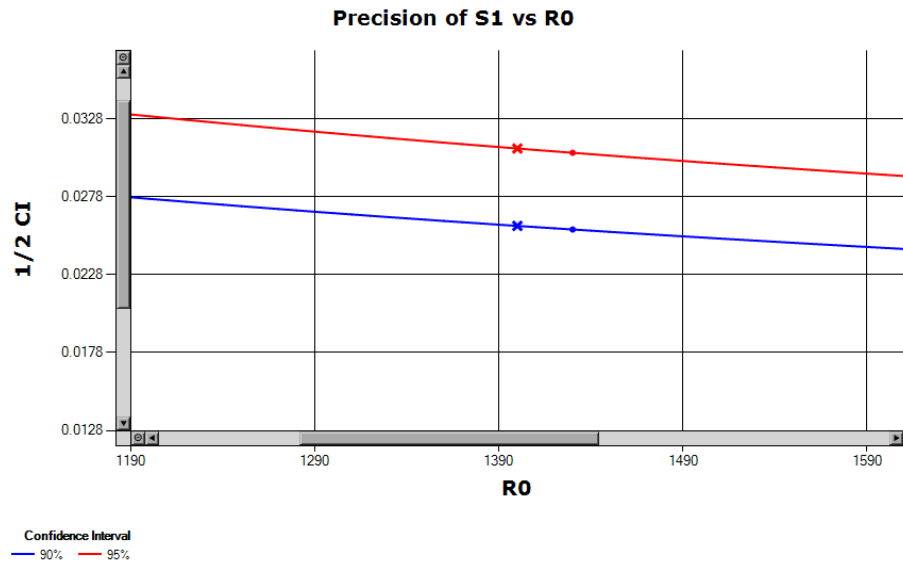
Chapter 3: Single Release Study

When the Single Release study design is selected, the user can specify the number of reaches and the number of replicates under the "Configuration" heading at the left of the SampleSize dialog.

In Figure 15, the user has selected "4" reaches and "2" replicates. The diagram and inputs are adjusted for the number of reaches; the resulting variance calculations will be adjusted for the number of replicates specified.

When a diagram becomes too large to show in entirety (e.g., when adding more reaches in the Single Release design), a scroll bar appears at the bottom, allowing the user to see the rest of the diagram.
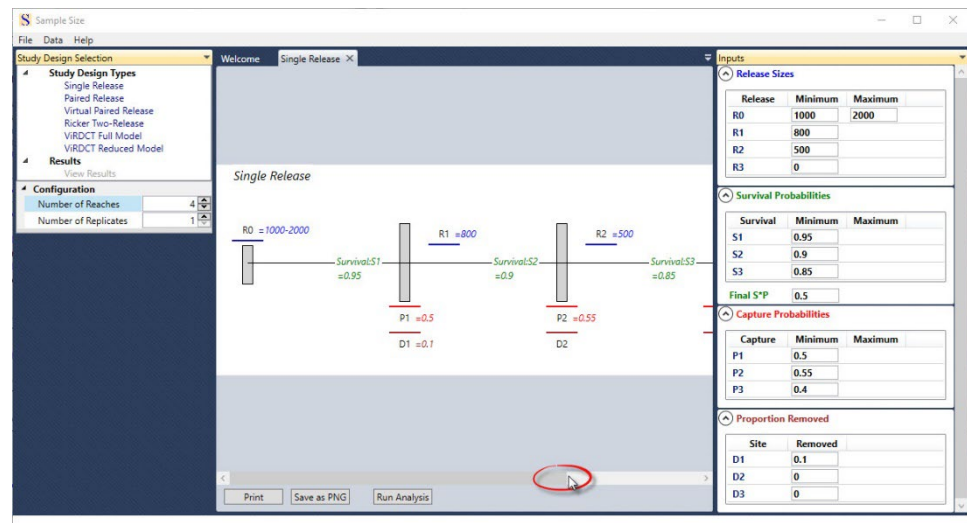
Figure 15. The Single Release study diagram with a scroll bar at the bottom

3.1 Single Release Results Graph

When there are three or more reaches in a Single Release design, the user can select the survival parameter of interest. Figure 16 shows the lower left

portion of the Results Graph for a Single Release design with four downstream reaches and three survival parameters: $S1$, $S2$, and $S3$. The user can select the parameter of interest, and the graph will display the CIs for the chosen parameter.
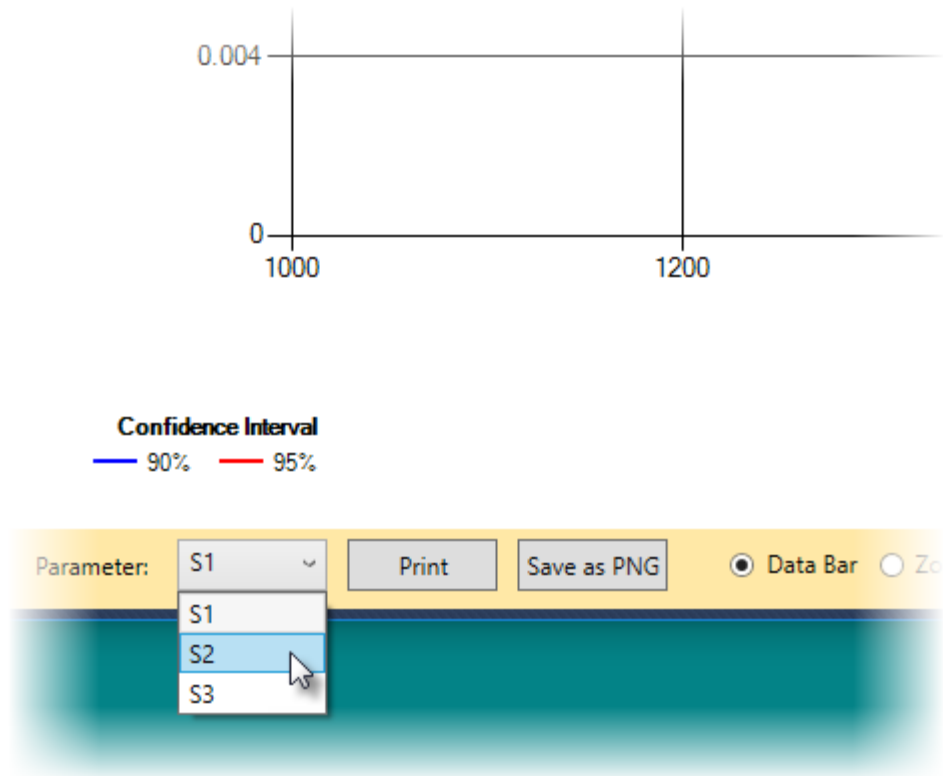
Figure 16. Selecting the parameter of interest on the Results Graph for a Single Release study with four reaches

Chapter 4: Paired Release Study

Figure 17 shows the diagram for a Paired Release study. As with the Single Release study, the user can specify the number of reaches and the number of replicates. Unlike the Single Release study, the results will only show the CIs for the treatment survival (*St*).
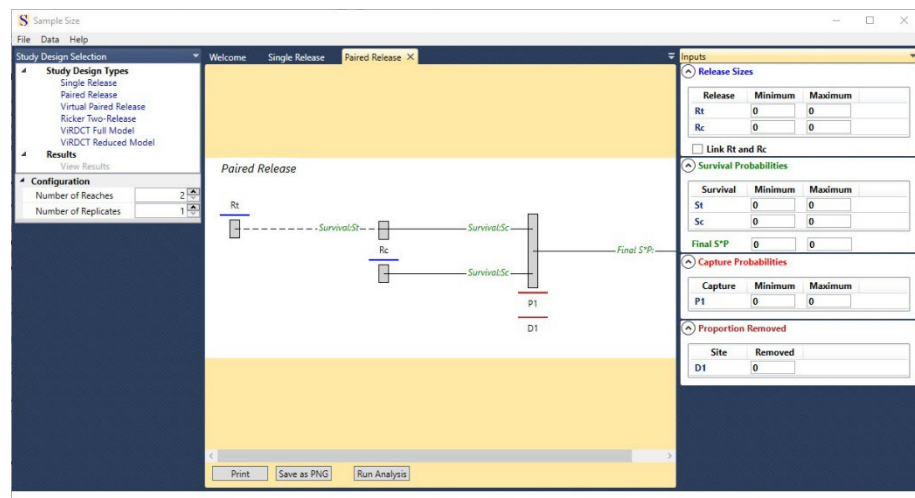
Figure 17. Program SampleSize with the Paired Release study design tab active

Program SampleSize allows the user to link the treatment release and control release values under the Inputs section as shown in Figure 18. When the "Link *Rt* and *Rc*" check box is checked, the control release (*Rc*) will be linked to the value of the treatment release (*Rt*). If a range is then specified for *Rt*, then both the treatment and control releases will be varied over the range.

Figure 18. Linking the treatment and control releases with a Paired Release study

Chapter 5: Virtual Paired Release Study

Figure 19 shows the diagram for a "Virtual Paired Release" study. There are no configuration options for a Virtual Paired Release study design, and the one parameter of interest for which CIs will be calculated and displayed is the dam passage survival (*Sdam*). Analogous to the Paired Release study design, release *R*2 and *R*3 can be linked with the same behavior as linking *Rt* and *Rc* in the Paired Release study design (Figure 20).
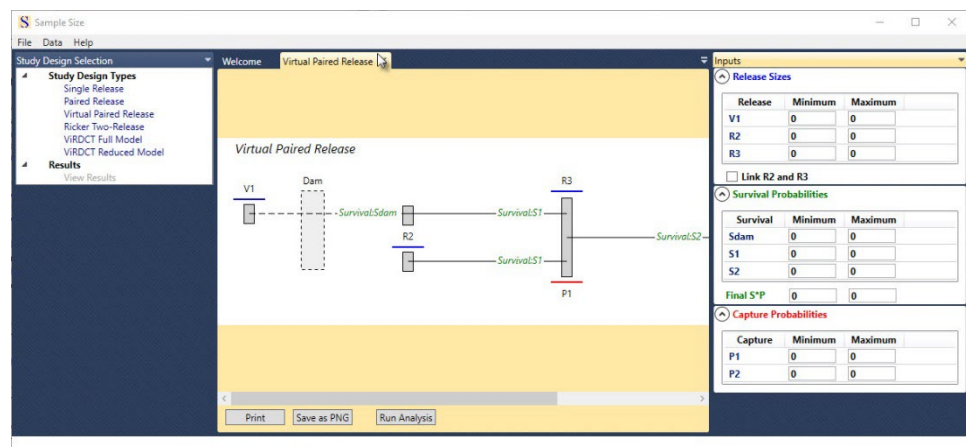
Figure 19. Program SampleSize with the Virtual Paired Release study design tab active

Figure 20. Linking release *R*2 and *R*3 in a Virtual Paired Release study

Chapter 6: Ricker Two-Release Study

Figure 21 shows the diagram for the "Ricker Two-Release" study. There are no configuration options, and the one parameter of interest for which CIs are calculated is the treatment survival (*St*). The treatment and control releases (*Rt* and *Rc*) are linked in sharing the same downstream capture probability, *P*.
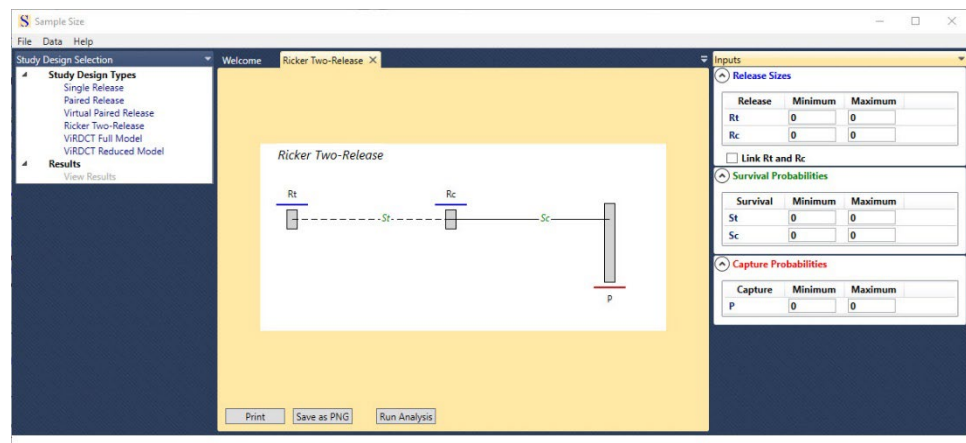
Figure 21. Program SampleSize with the Ricker Two-Release study design tab active

The survival probability for the treatment (*Rt*) is *St* x *Sc*, and the survival probability for the control release is *Sc*. Thus, the treatment survival probability is calculated as the ratio of the two.

Chapter 7: ViRDCt (Full model)

Figure 22 shows the diagram for the ViRDCt Full Model. There are no configuration options, and the one parameter of interest for which CIs are calculated is the dam passage survival (*SD*).

Figure 22: Program SampleSize with the ViRDCt Full Model study design tab active.

There are two release for the ViRDCt model:

*V*1 = Virtual release, consisting of fish detected at the dam face,
*D*1 = Dead fish released in the dam tailrace.

The parameters for the full ViRDCt model are as follows:

SD	=	dam passage survival,
λ	=	joint probability of survival between tailrace and tailwater arrays, and being detected at the tailwater array,
ω	=	probability of a dead fish from D_1 arriving at the tailrace array,
Ψ	=	joint probability that a dead fish is washed down to the tailwater array from the tailrace array and is detected at the tailwater array,
*P*1	=	probability of an alive V_1 fish being detected at the tailrace array,
PD	=	probability of detecting a dead fish at the tailrace array.

Chapter 8: ViRDCt (Reduced model)

Figure 23 shows the diagram for the "ViRDCt Reduced Model" study. There are no configuration options, and the one parameter of interest for which CIs are calculated is the dam passage survival (SD).
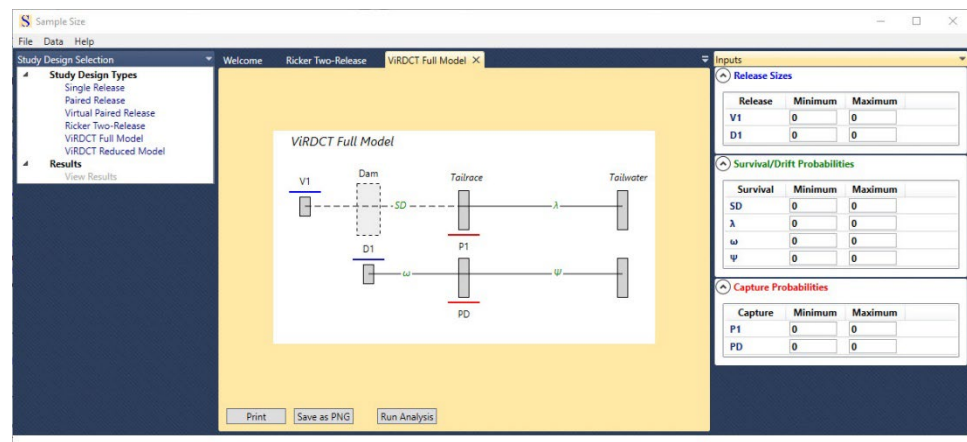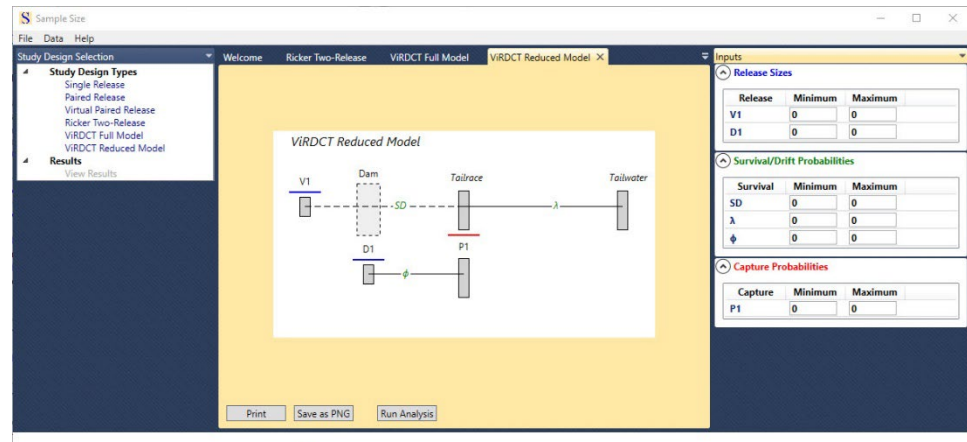
Figure 23: Program SampleSize with the ViRDCt Reduced Model study design tab active.

The reduced model has two released, $V1$ and $D1$ as described in the previous section for the full ViRDCt model. The parameters for the reduced ViRDCt model are:

SD = dam passage survival,

λ = joint probability of survival between tailrace and tailwater arrays, and being detected at the tailwater array,

ϕ = joint probability of a dead released fish (D_1) arriving at the tailrace array and being detected at that array,

$P1$ = probability of an alive $V1$ fish being detected at the tailrace.

Appendix: Variance Calculations

A.1: Single Release Study

The variance calculations for the Single Release study are adapted from Burnham et al. (1987).

For survival parameter S_i,

$Var(S_i) = S_i{}^2(f_1 + f_2 + f_3)$ for $i = 1$ to $K - 1$ where

$K =$ the number of reaches,

$$f_1 = \frac{1}{E(r_i)} - \frac{1}{R_i}$$

$$f_2 = (1 - P_{i+1})^2 \left(\frac{1}{E(r_{i+1})} - \frac{1}{R_{i+1}} \right)$$

$$f_3 = P_{i+1}(1 - P_{i+1}) \frac{(1 - A_{i+1})^2}{A_{i+1}T_{i+1}}$$

$$E(r_i) = R_i(1 - \chi_i)$$

$\chi_i =$ The probability of not being detected after occasion i

$$\chi_{K-1} = 1 - S_K P_K$$

$\chi_i = 1 - S_{i+1} + S_{i+1}(1 - P_{i+1})\chi_{i+1}$ for $i = 0 \ to \ K - 1$

$R_i =$ The number released from site i, including both those initially released and those previously released, detected and not removed

$P_i =$ Capture probability at site i

$$A_i = \frac{r_i}{R_i}$$

$$T_i = m_i + z_i$$

m_i = Number of previously released individuals detected at site i

$$z_i = \sum_{j=0}^{i-1} (r_j - m_{j+1})$$

For a replicated survival study with n replicates, survival in reach i is estimated by $\hat{\bar{S}}_i$:

$$\hat{\bar{S}}_i = \frac{\sum_{j=1}^{n} \hat{S}_{ij}}{n}$$

where $\hat{S}_{ij}$ is the reach-i survival estimate for replicate j ($j = 1, \dots, n$). The variance of $\hat{\bar{S}}_i$ is composed of the population variability in survival for reach i and the average sampling variability for that reach:

$$\text{Var}\left(\hat{\bar{S}}_i\right) = \frac{1}{n}\left[\sigma_{S_i}^2 + \frac{\sum_{j=1}^{n} \text{Var}(\hat{S}_{ij}|S_{ij})}{n}\right]$$

where $\sigma_{S_i}^2$ is the natural variation in S_i. The quantity $\text{Var}(\hat{S}_{ij}|S_{ij})$ is the measurement error of $\hat{S}_{ij}$ given S_{ij} for replicate j.

In turn,

$$\text{Var}\left(\hat{\bar{S}}_i\right) = \frac{\sigma_{S_i}^2}{n} + \frac{\sum_{j=1}^{n} \text{Var}(\hat{S}_{ij}|S_{ij})}{n^2}$$

$$= \frac{\sigma_{S_i}^2}{n} + \frac{\overline{\text{Var}(\hat{S}_i|S_i)}}{n}.$$

So then

$$\text{SE}\left(\hat{\bar{S}}_i\right) = \sqrt{\frac{\sigma_{S_i}^2 + \overline{\text{Var}(\hat{S}_i|S_i)}}{n}}. \tag{A.1}$$

If $\sigma_{S_i}^2 = 0$, then $\text{SE}\left(\hat{\bar{S}}_i\right) = \dfrac{\sqrt{\overline{\text{Var}(\hat{S}_i|S_i)}}}{\sqrt{n}}$ or $\dfrac{\sqrt{\sum_{j=1}^{n} \text{Var}(\hat{S}_{ij}|S_{ij})}}{n}$.

If it is further assumed that $\text{Var}(\hat{S}_{ij}|S_{ij}) = \text{Var}(\hat{S}_i|S_i)$ for all replicates j ($j = 1, \ldots, n$), then the expected standard error for reach i becomes

$$\text{SE}\left(\hat{\bar{S}}_i\right) = \frac{\sqrt{n\text{Var}(\hat{S}_i|S_i)}}{n}$$

$$\text{SE}\left(\hat{\bar{S}}_i\right) = \frac{\text{SE}(\hat{S}_i|S_i)}{\sqrt{n}} \quad\quad\quad (A.2)$$

Either Equation (A.1) or (A.2) is correct depending on sample size assumptions.

Note that this development is for sample size calculations rather than for data analysis. For data analysis, a bias correction is necessary because $\widehat{\text{Var}}\left(\hat{\bar{S}}_i\right) \neq \text{Var}\left(\hat{\bar{S}}_i\right)$.

A.2 Paired Release Study

Let $S_1 = S_t S_c$

Then the variances of S_1 and S_c are calculated as with the Single Release study in A.1, treating each release as independent of the other.

$$S_t = \frac{S_1}{S_c}$$

$Var(S_t)$ is calculated using the delta method and the exact formula for the variance of a produce of two independent variables (Seber, 1982)

$$Var(S_t) = S_t^{\,2}\left(\frac{Var(S_1)}{S_1^{\,2}} + \frac{Var(S_c)}{S_c^{\,2}} + \frac{Var(S_1)Var(S_2)}{S_1^{\,2}S_c^{\,2}}\right)$$

A.3 Virtual Paired Release Study

Let $\quad \phi_1 = Sdam \; x \; S_1$
$\quad\quad \phi_2 = S_1 S_2$
$\quad\quad \phi_3 = S_2$

Then $Sdam = \dfrac{\phi_1}{\phi_2/\phi_3} = \dfrac{\phi_1 \phi_3}{\phi_2}$

Treating all three releases as independent from each other, ϕ_1, ϕ_2 and ϕ_3 are calculated as with a Single Release study in A.1. Thus the variance using methods described in A.2 is as follows

$$Var(Sdam) = \left(\frac{1}{\phi_2{}^2} + \frac{Var(\phi_2)}{\phi_2{}^4} \right) \cdot$$

$$\left[\phi_1{}^2 Var(\phi_3) + \phi_3{}^2 Var(\phi_1) + Var(\phi_1)Var(\phi_3) \right] +$$

$$\frac{(\phi_1\phi_3)^2}{\phi_2{}^4} \, Var(\phi_2)$$

A.4 Ricker Two-Release Study

$$Var(S_t) = \frac{S_t}{S_c \cdot P} \left(\frac{(1 - S_c S_t P)}{R_t} + \frac{(1 - S_c P)S_t}{P_c} \right)$$

A.5 ViRDCt Full and Reduced Model

For the two ViRDCt models, there are no closed form variance estimates. Thus, the variances are calculated numerically from the likelihoods.

The likelihood for the Full ViRDCt model:

$$L = \begin{pmatrix} V_1 \\ n_{11} & n_{01} & n_{10} & n_{00} \end{pmatrix} (S_D \cdot P_1 \cdot \lambda + (1 - S_D)\omega P_D \cdot \psi)^{n_{11}}$$

$$\cdot (S_D(1 - P_1)\lambda + (1 - S_D)\,\omega(1 - P_D)\,\psi)^{n_{01}}$$

$$\cdot (S_D P_1(1 - \lambda) + (1 - S_D)\omega P_D(1 - \psi))^{n_{10}}$$

$$\cdot (S_D(1 - P_1)(1 - \lambda) + (1 - S_D)(1 - \omega + \omega(1 - P_D)(1 - \psi))^{n_{00}}$$

$$\cdot \begin{pmatrix} D_1 \\ d_{11} & d_{01} & d_{10} & d_{00} \end{pmatrix} \cdot (\omega P_D \psi)^{d_{11}} \cdot (\omega(1 - P_D)\psi)^{d_{01}}$$

$$\cdot (\omega P_D(1 - \psi))^{d_{10}}(1 - \omega + \omega(1 - P_D)(1 - \psi))^{d_{00}}$$

The full model has six parameters and six minimum sufficient statistics.

The likelihood for the Reduced ViRDCt model:

$$L = \begin{pmatrix} V_1 \\ n_{11} & n_{01} & n_{10} & n_{00} \end{pmatrix} (S_D P_1 \lambda)^{n_{11}} (S_D(1 - P_1)\lambda)^{n_{01}}$$

$$\cdot \, (S_D P_1 (1 - \lambda) + (1 - S_D)\phi)^{n_{10}}$$

$$\cdot \, (S_D (1 - P_1)(1 - \lambda) + (1 - S_D)(1 - \phi))^{n_{00}}$$

$$\cdot \, \binom{D_1}{d_1 \quad d_0} \phi^{d_1} (1 - \phi)^{d_0}$$

The reduced model has four parameters and four minimum sufficient statistics.

References

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

G. A. F. Seber. *The Estimation of Animal Abundance*. MacMillan, New York, New York, 1982.