

# Program TribPit SampleSize

for Estimating Survival in Tributaries

Developed by:

James Lady and John R. Skalski  
Columbia Basin Research  
School of Aquatic & Fishery Sciences  
University of Washington  
1325 Fourth Avenue, Suite 1820  
Seattle, Washington 98101-2509

Prepared for:

U.S. Department of Energy  
Bonneville Power Administration  
Division of Fish and Wildlife  
P.O. Box 3621  
Portland, Oregon 97208-3621  
Project No. 1989-107-00  
Contract No. 67518

March 2015



# Acknowledgments

---

This project is funded by the Bonneville Power Administration (BPA), US Department of Energy, under Contract No. 67518, Project No. 1989-107-00, as part of the BPA's program to protect, mitigate, and enhance fish and wildlife affected by the development and operation of hydroelectric facilities on the Columbia River and its tributaries.



# Table of Contents

Program TribPit SampleSize..... i  
Acknowledgments ..... i  
Chapter 1: Overview..... 1  
Chapter 2: Study Configuration..... 3  
Chapter 3: Data Entry ..... 5  
    Entering the Proportions..... 7  
Chapter 4: Compute Function..... 9  
    Single Value for the Release Size..... 9  
    Range of Values for the Release Size ..... 10  
Appendix: Calculations ..... 13  
    Calculate the Migration Proportions ..... 13  
    Calculate the Variance of **S** ..... 14  
    Calculate the variance of **p1**..... 16



---

# Chapter 1: Overview



Studies have shown that many juvenile salmon and steelhead stay upstream in the tributaries one or more years before migrating downstream, and traditional release-recapture models are not valid for analyzing these data. This was the motivation behind the development of Program TribPit, available at [www.cbr.washington.edu/analysis/apps/TribPit](http://www.cbr.washington.edu/analysis/apps/TribPit). The details of the model are explained in the TribPit User's Manual available at the same website.

The Program TribPit SampleSize was developed to allow project managers to calculate the required sample sizes for studies in the tributaries to achieve a required precision of cohort survival at the first detection site.





# Chapter 2: Study Configuration

When the Tributary Survival SampleSize program is started, a splash screen is displayed. It will go away in about five seconds, or the user can simply click on it to dismiss it. The user then sees the data entry dialog as shown in Figure 1.

The screenshot shows a software window titled "Tributary Survival SampleSize". The window contains a diagram with a horizontal line and several input fields. The fields are arranged as follows: R0, S11,  $\pi_1$ , S21, S12,  $\pi_2$ , S22, P1, P2, and Cohort survival. At the bottom of the window, there are two dropdown menus: "Number of age classes" set to 2 and "Number of downstream arrays" set to 2. A "Compute" button is located at the bottom right of the window.

**Figure 1. Data entry dialog when user starts the Tributary Survival SampleSize program**

At the bottom of the dialog are two user controls (circled in red in Figure 1), allowing the user to set (1) the number of age classes, and (2) the number of downstream arrays. The number of age classes can be set between 2 and 9, and the number of downstream arrays can be either 2 or 3. At initial startup, the number of age classes is 2 and the number of downstream arrays is set to 2. When the user changes the number of age classes and/or number of downstream arrays, the diagram immediately changes to reflect the new configuration. For example, see below in Figure 2.

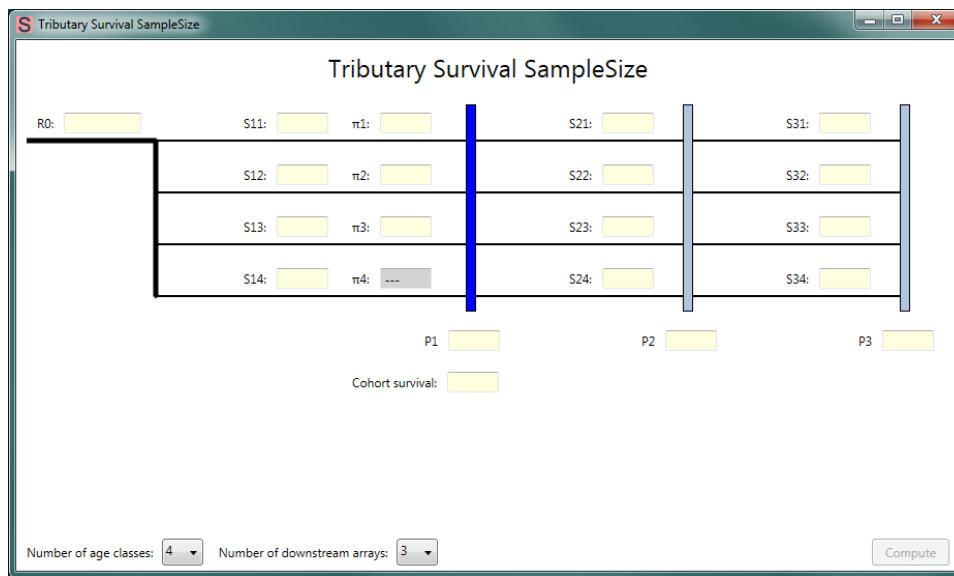


Figure 2. Data entry dialog with 4 age classes and 3 downstream array

Notice that now there are three detection sites rather than two. The first detection site is highlighted in dark blue—it is the site for which cohort survivals are being estimated. The subsequent downstream sites are in a lighter shade of blue.

# Chapter 3: Data Entry

---

The user enters data for the sample size calculation in the text boxes with a light yellow background. The values to be entered are:

$R_0$	The release size, either a single value or a range of values as explained below.
$S_{1j}$	The estimated survival probability for the first reach for age class $j$ ( $j$ being a number between 1 and the number of age classes). The survival probability must be between 0 and 1.
$\pi_j$	The estimated proportion of individuals of age class $j$ that survive to the first detection site. The proportion must be between 0 and 1. In addition, all the proportions must sum to 1. Notice that the final proportion is in gray. The user does not enter a value for the final proportion; it is automatically set to 1 minus the sum of the other proportions.
$S_{ij}$	The estimated survival probability for the downstream sites. For three downstream arrays, the user enters estimates for $S_{2j}$ and $S_{3j}$ ; for two downstream arrays, only values for $S_{2j}$ are required. The survival probabilities must be between 0 and 1.
$P_i$	The estimated detection probability for site $i$ , between 0 and 1.
Cohort survival	The estimated overall cohort survival for the first detection site, between 0 and 1.

The "Compute" button at the bottom left is grayed out (disabled) when the program is started. It is only enabled when there is a valid entry in each user input. When user input is invalid, the text box with the invalid input is outlined in red. In Figure 3, the user entered "1.1" for the survival in reach 2, age class 1. The text box is outlined in red, and the compute button is disabled.

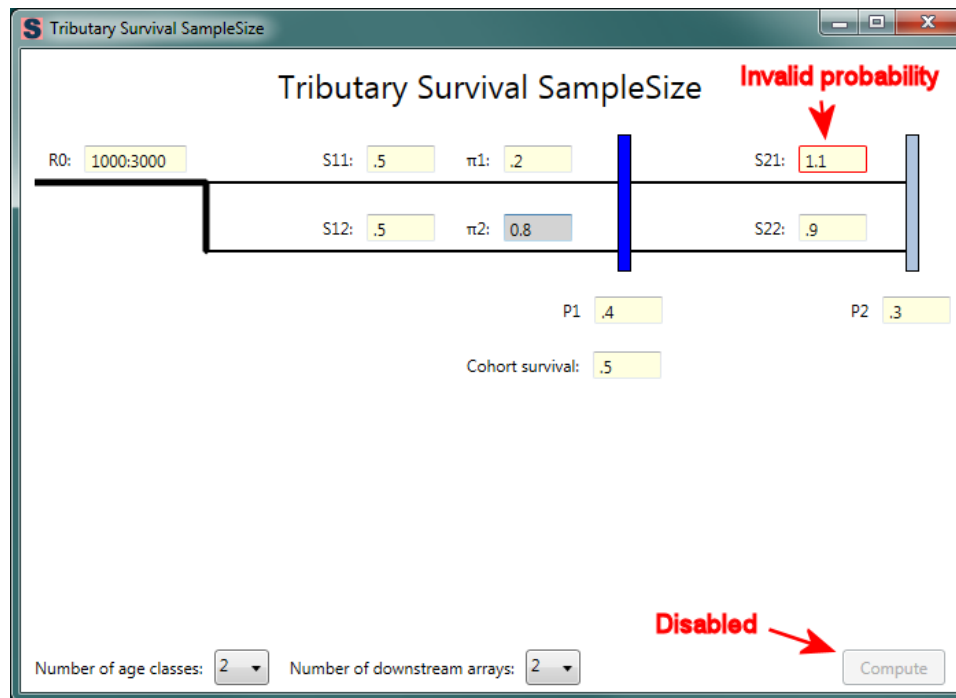


Figure 3. Data entry dialog with an invalid value for a survival probability

In Figure 4, an invalid number has been entered; hence, the compute button is disabled.

The screenshot shows a software dialog box titled "Tributary Survival SampleSize". It contains several input fields and a "Compute" button. The fields are:

- R0: 1000:3000
- S11: .5
- $\pi_1$ : .2
- S21: .5
- S12: e5 (highlighted in red with a red arrow pointing to it and the text "Not a valid number" below it)
- $\pi_2$ : 0.8 (disabled)
- S22: .9
- P1: .4
- P2: .3
- Cohort survival: .5

At the bottom, there are two dropdown menus: "Number of age classes:" set to 2 and "Number of downstream arrays:" set to 2. The "Compute" button is disabled, indicated by a red arrow and the text "Disabled" next to it.

Figure 4. Data entry dialog with an invalid number

## Entering the Proportions

The proportions ( $\pi_j$ ) represent the number of age class  $j$  that survive to the first detection site. The proportion for the final age class is in gray because it is not a user input; it is calculated automatically so that the sum of all the proportions is 1.0. Thus, not only must the proportions be between 0 and 1, but they must sum to 1.0. In Figure 5, the user has entered proportions between 0 and 1 for the first three age class, but they sum to 1.2, causing the final proportion to be set to -1.2.

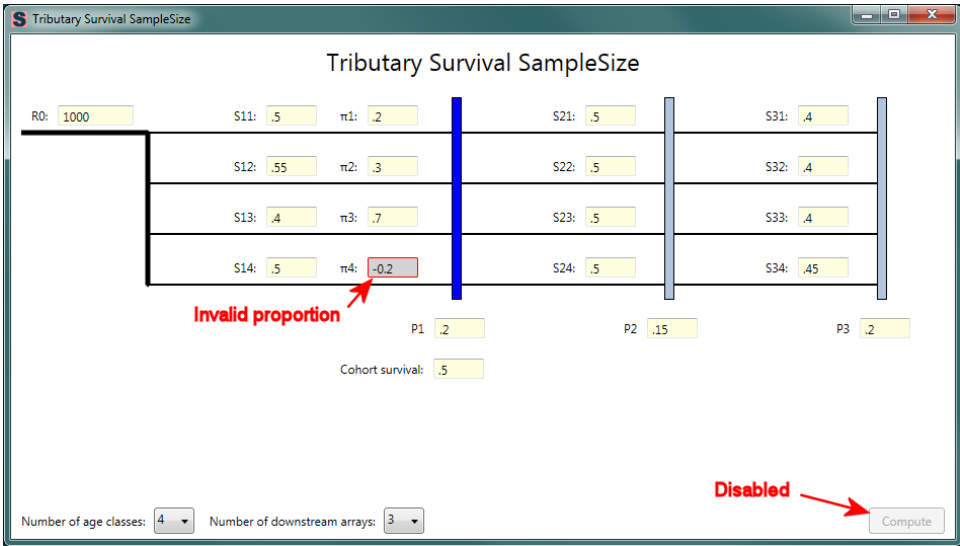


Figure 5. Data entry dialog with invalid proportions

Figure 6 shows an example with all entries valid and the compute button enabled.

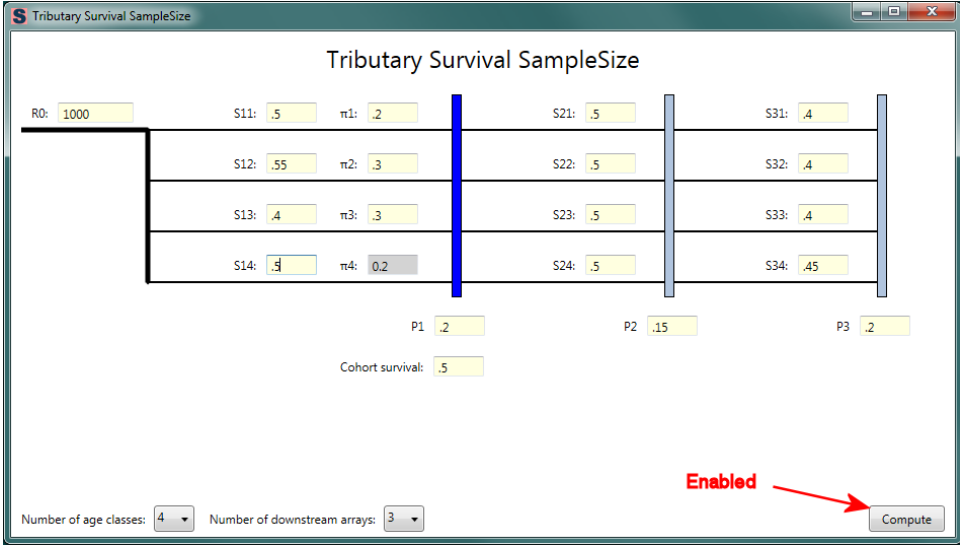


Figure 6. Data entry dialog with all entries valid and the Compute button enabled

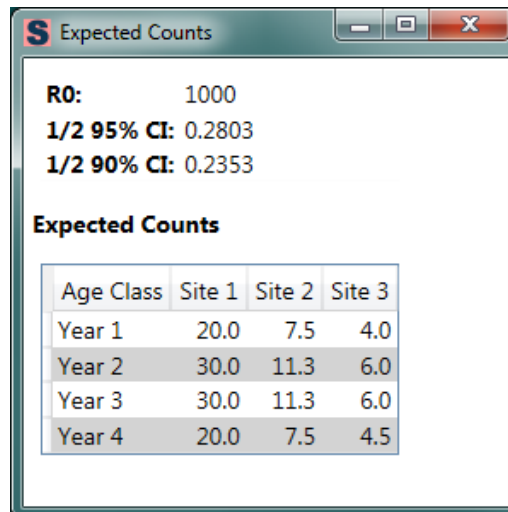
# Chapter 4: Compute Function

When the user presses the Compute button, one of two actions will be performed: (1) If a single value for the release size ( $R_0$ ) is entered, a table of expected counts is shown; (2) if a range is entered for  $R_0$ , a plot of  $\frac{1}{2}$  confidence intervals as a function of the release size is shown.

## Single Value for the Release Size

If the user enters a single value for the release size, a table will be produced as shown in Figure 7. The first three lines are:

1. The entered release size,
2.  $\frac{1}{2}$  of a 95% confidence interval,
3.  $\frac{1}{2}$  of a 90% confidence interval.



The screenshot shows a window titled "Expected Counts" with a red 'S' icon. It displays the following text:

**R0:** 1000  
**1/2 95% CI:** 0.2803  
**1/2 90% CI:** 0.2353

**Expected Counts**

Age Class	Site 1	Site 2	Site 3
Year 1	20.0	7.5	4.0
Year 2	30.0	11.3	6.0
Year 3	30.0	11.3	6.0
Year 4	20.0	7.5	4.5

Figure 7. Table produced when a single value is entered for the release size

In the example shown in Figure 7, if the user had entered, say, 0.55 for the cohort survival, a 95% confidence interval would be  $0.55 \pm 0.2803$ , and a 90% confidence interval would be  $0.55 \pm 0.2353$ .

The second part of the table shows the expected counts for each age class and site based on the user inputs.

### Range of Values for the Release Size

The user has the option of entering a range of values for  $R_0$  in the form "min:max." For example in Figure 4, the user specified a range of values from 1,000 to 3,000 for  $R_0$ . When the user presses the Compute button for a range of values for the release size, a plot appears as shown in Figure 8. The X axis is the release size  $R_0$ , and the Y axis is the corresponding values for the confidence intervals. The 95% confidence interval is represented by the blue line, and the 90% confidence interval is represented by the green line.

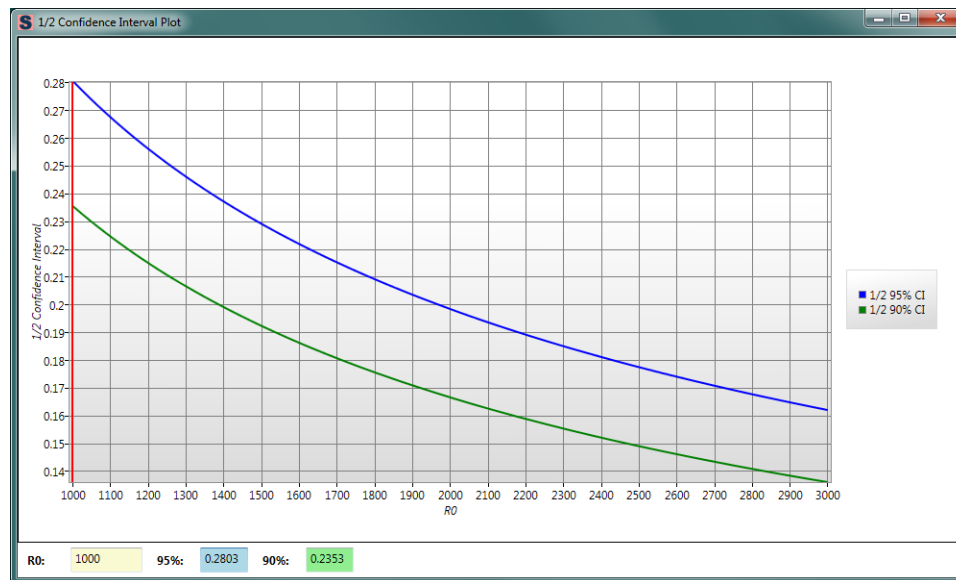


Figure 8. Plot for a range of values for the release size



## Chapter 4: Compute Function

When the plot first appears, a vertical red line is positioned at the left side at the minimum release size. When the user left-clicks on the plot, the line will move to the mouse position. The user can then drag the red “tracker” line by holding the left mouse button and moving the mouse. The text boxes at the bottom of the plot show the release size ( $R_0$ ),  $\frac{1}{2}$  of a 95% confidence interval (blue background), and  $\frac{1}{2}$  of 90% confidence interval (green background) corresponding to the current position of the red tracking line. In Figure 9, the user has positioned the red tracking line at a release size of 1700. If the cohort survival is 0.55, then a 95% confidence interval for  $R_0 = 1700$  would be  $0.55 \pm 0.215$ , and a 90% confidence interval would be  $0.55 \pm 0.1805$ .

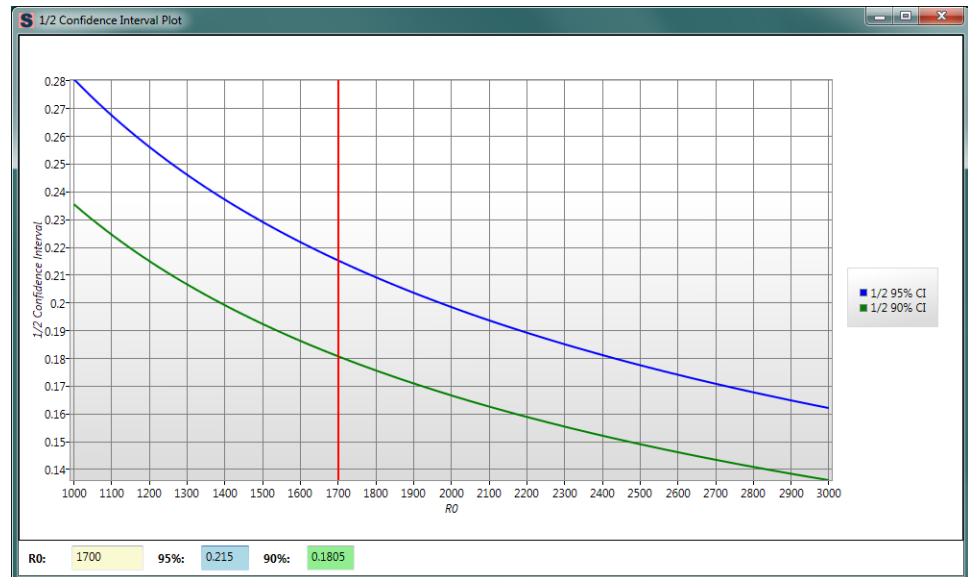


Figure 9. Plot for a range of values with the red tracking line positioned at  $R_0 = 1700$

Alternately, the user can edit the value in the  $R_0$  text box directly, causing the red tracking line to move to the specified value.



---

# Appendix: Calculations

---

## Calculate the Migration Proportions

$S$  = cohort survival to the first detection site

$S_{ij}$  = reach survival to site  $i$ , age class  $j$

$P_i$  = detection probability at site  $i$

$\pi_j$  = proportion of those reaching the first detection site in year  $j$

$R_0$  = number released

$K$  = number of age classes

Calculate  $\theta_j$  = probability of migrating in year  $j$ . Let  $\tau_j = \theta_j S_{ij}$ .

$$S = \sum_{k=1}^K \theta_k S_{1k} = \sum_{k=1}^K \tau_k \quad (1)$$

$$\pi_j = \frac{\tau_j}{\sum_{k=1}^K \tau_k} \quad (2)$$

Equations (1) and (2) can be written for 2 age classes:

$$\begin{bmatrix} \pi_1 - 1 & \pi_1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix} = \begin{bmatrix} 0 \\ S \end{bmatrix}.$$

For 3 age classes:

$$\begin{bmatrix} \pi_1 - 1 & \pi_1 & \pi_1 \\ \pi_2 & \pi_2 - 1 & \pi_2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ S \end{bmatrix}.$$

In general

$$M\vec{\tau} = \vec{y}$$

$$\vec{\tau} = M^{-1}\vec{y}.$$

so

$$\theta_j = \frac{\tau_j}{S_{1j}}.$$

### Calculate the Variance of $\hat{S}$

$$\hat{S} = \frac{n_1}{\hat{P}_1 R_0}$$

where  $n_1$  = number detected at the first detection site.

$$E(n_1) = R_0 p_1 \sum_{k=1}^K \theta_k S_{1k}$$

$$\text{Var}(\hat{S}) = \text{Var}[E(\hat{S}|p_1)] + E[\text{Var}(\hat{S}|p_1)]$$

$$E(\hat{S}|p_1) = \frac{n_1}{\hat{p}_1 R_0}$$

$$\text{Var}[E(\hat{S}|p_1)] = \frac{n_1^2}{R_0^2} \text{Var}\left(\frac{1}{p_1}\right)$$

Using the delta method

$$\text{Var}\left(\frac{1}{\hat{p}_1}\right) = \frac{1}{p_1^4} \text{Var}(\hat{p}_1)$$

$$\text{Var}[E(\hat{S}|\hat{p}_1)] = \frac{n_1^2}{R_0^2 p_1^4} \text{Var}(\hat{p}_1)$$

$$\text{Var}(\hat{S}|p_1) = \frac{1}{\hat{p}_1^2 R_0^2} \text{Var}(n_1)$$

$$\text{Var}(\hat{S}|p_1) = \frac{1}{\hat{p}_1^2 R_0^2} \cdot R_0 \cdot p(n_1) \cdot (1 - p(n_1))$$

$$\text{Var}(\hat{S}|p_1) = \frac{1}{\hat{p}_1^2 R_0} \cdot p_1 \hat{S} (1 - p_1 \hat{S})$$

$$E[\text{Var}(\hat{S}|p)] = \frac{\hat{S}}{\hat{p}_1 R_1} (1 - \hat{p}_1 \hat{S})$$

$$\text{Var}(\hat{S}) = \frac{n_1^2}{R_0^2 \hat{p}_1^4} \text{Var}(\hat{p}_1) + \frac{\hat{S}}{\hat{p}_1 R_0} (1 - \hat{p}_1 \hat{S}).$$

Substituting  $E(n_1)$  for  $n_1$

$$\text{Var}(\hat{S}) = \frac{R_0^2 \hat{p}_1^2 \hat{S}^2}{R_0^2 \hat{p}_1^4} \text{Var}(\hat{p}_1) + \frac{\hat{S}}{\hat{p}_1 R_0} (1 - \hat{p}_1 \hat{S})$$

$$\text{Var}(\hat{S}) = \frac{\hat{S}^2}{\hat{p}_1^2} \text{Var}(\hat{p}) + \frac{\hat{S}}{\hat{p}_1 R_0} (1 - \hat{p}_1 \hat{S}).$$

### Calculate the variance of $\hat{p}_1$

$\hat{p}_1$  can be written as

$$\hat{p}_1 = \frac{n_{11}}{n_{11} + n_{01}}.$$

Using the delta method

$$\begin{aligned} \text{Var}(\hat{p}_1) = \frac{(n_{11}n_{01})^2}{(n_{11} + n_{01})^4} & \left[ \frac{\text{Var}(n_{11})}{n_{11}^2} + \frac{\text{Var}(n_{01})^2}{n_{01}^2} \right. \\ & \left. + \frac{2\text{Cov}(n_{11}, n_{01})}{n_{11}n_{01}} \right] \end{aligned}$$

$$\text{Var}(n_{11}) = R_0 p(n_{11})(1 - p(n_{11}))$$

$$\text{Var}(n_{01}) = R_0 p(n_{01})(1 - p(n_{01}))$$

$$\text{Cov}(n_{11}, n_{01}) = -R_0 p(n_{11})p(n_{01}).$$

Substitute in expected values for  $n_{11}$ ,  $n_{01}$ .

## Appendix: Calculations

---

For two downstream arrays:

$$E(n_{11}) = R_0 p_1 p_2 \sum_{k=1}^K \theta_k S_{1k} S_{2k}$$

$$E(n_{01}) = R_0 (1 - p_1) p_2 \sum_{k=1}^K \theta_k S_{1k} S_{2k}$$

For three downstream arrays:

$$E(n_{11}) = R_0 p_1 \sum_{k=1}^K \theta_k S_{1k} S_{2k} (p_2 + p_3 S_{3k} - p_2 p_3 S_{3k})$$

$$E(n_{01}) = R_0 (1 - p_1) \sum_{k=1}^K \theta_k S_{1k} S_{2k} (p_2 + p_3 S_{3k} - p_2 p_3 S_{3k})$$