

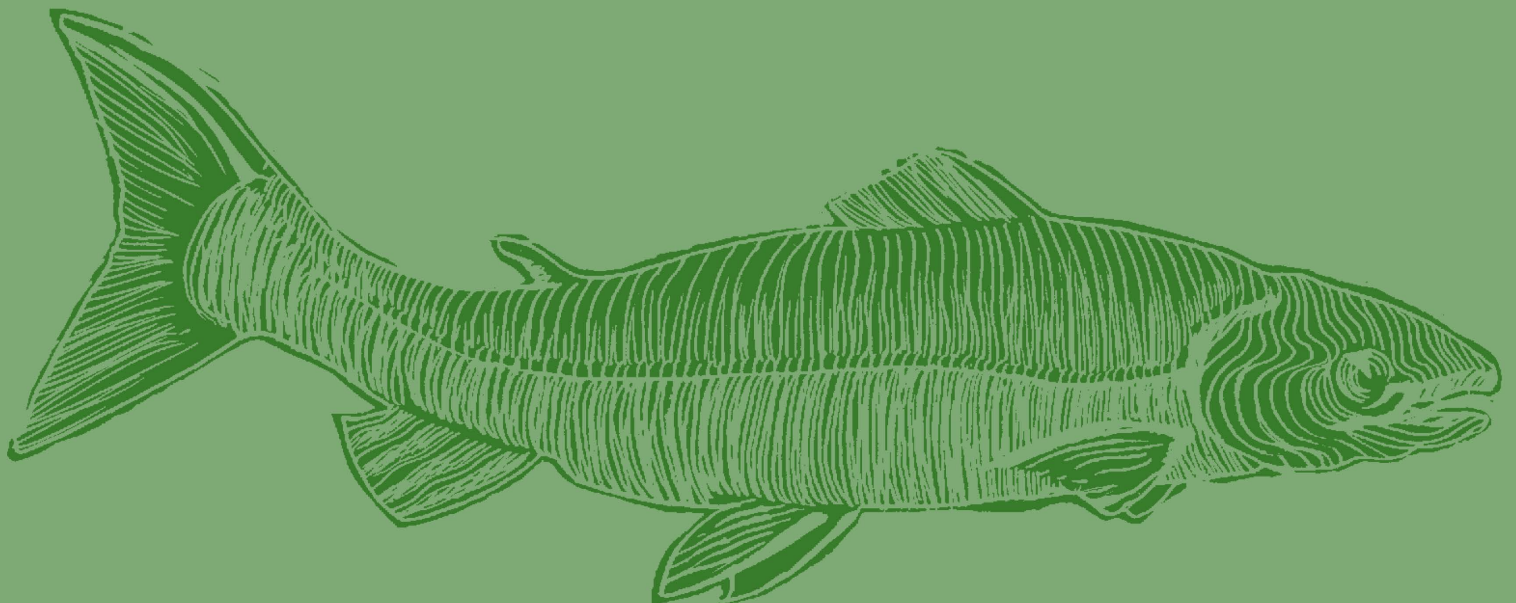
SURPH 3.5 USER'S MANUAL

STATISTICAL SURVIVAL ANALYSIS FOR FISH AND WILDLIFE TAGGING STUDIES

COLUMBIA BASIN RESEARCH

SCHOOL OF AQUATIC AND FISHERY SCIENCES

W UNIVERSITY *of* WASHINGTON



SURPH 3.5

SURvival under Proportion Hazards

Developed by:

James Lady, Peter Westhagen, 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. 59002 (Original), 84060 REL 13 (Current)

April 2026

Acknowledgements

This project is funded by the Bonneville Power Administration (BPA), U.S. Department of Energy, 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

Acknowledgements.....	1
Chapter 1: Introduction.....	6
Chapter 2: Using SURPH.....	7
2.1 Navigation Panel.....	9
2.2 SURPH Menus.....	9
2.2.1 File Menu.....	10
2.2.2 Edit Menu.....	10
2.2.3 View Menu.....	13
2.2.4 Help Menu.....	14
2.3 Literature Cited.....	14
Chapter 3: Data.....	13
3.1 Data Diagnostics.....	13
3.1.1 M-Array.....	15
3.1.2 Burnham Tests.....	16
3.1.3 Individual Covariate Graph.....	20
3.2 Manly-Parr Estimates.....	21
3.3 CJS Estimates.....	22
3.4 CJS Profile Likelihoods.....	25
3.5 Literature Cited.....	31
Chapter 4: Models.....	32
4.1 Using Effects Dialog and Effects Control Buttons.....	32
4.2 Modeling Main Effects.....	34
4.3 Modeling Group and Individual Covariate Effects.....	39
4.4 Joint Survival and Capture Modeling.....	41

4.5 Two-stage Modeling.....	44
4.5.1 Detection-Only Modeling	44
4.5.2 Survival Modeling	47
Chapter 5: Results.....	48
5.1 View Results.....	48
5.1.1 Probability Report.....	49
5.1.2 Parameter Report.....	50
5.2 Survival Plots	52
5.2.1 Individual Covariate Survival.....	52
5.2.2 Group Survival Plot	54
5.3 Diagnostic Plots	55
5.3.1 Model Estimates vs. CJS Estimates	55
5.3.2 Residuals vs. Capture Histories.....	56
5.3.3 Residual - Normal QQ Plot.....	58
5.4 Advanced Export Functions	59
5.5 Literature Cited.....	60
Chapter 6: Hypothesis Testing.....	61
6.1 Likelihood Ratio Test	61
6.2 Analysis of Deviance.....	63
6.2.1 Using the Defaults	63
6.2.2 Select Existing Results.....	64
6.2.3 Analysis of Deviance Dialog.....	65
6.2.3 Adding and Removing Models from the ANODEV Table	66
6.3 Literature Cited.....	73
Chapter 7: Models Across Multiple Datasets	74
Chapter 8: Analysis of Examples	80

8.1 Paired Release Example.....	80
8.1.1 Create the Models	81
8.1.2 Test the Hypotheses	86
8.1.3 Increase Precision by Reducing the Number of Parameters	88
8.1.4 Results Diagnostics	92
8.2 Effect of Timing of Releases from a Hatchery Example	94
8.2.1 Detection-only Modeling	95
8.2.2 Survival Modeling with Analysis of Deviance.....	101
8.3 Relationship of Migration Distance to Survival	105
8.3.1 Analysis of Deviance.....	106
8.3.2 Group Survival Plots.....	108
8.4 Relationship of Length to Survival	110
8.4.1 Initial Data Analysis	110
8.4.2 Modeling the Detection Probabilities	111
8.4.3 Modeling the Survival Probabilities.....	114
8.4.4. Individual Covariate Plot	116
8.5 Literature Cited	119
Appendix A: SURPH Input Data File.....	120
Capture Histories.....	123
Appendix B: SURPH Statistical Theory	125
B.1 Release-Recapture Theory	125
B.1.1 Definition	125
B.1.2 An Intuitive Derivation of the Release-Recapture Parameters	126
B.1.3 Multinomial Likelihood Model	127
B.1.4 Multinomial Likelihood for a Multiple-Population Study.....	130
B.1.5 Model Assumptions	131

B.1.6 Cormack-Jolly-Seber Parameter Estimates.....	132
B.2 Modeling Survival Main Effects.....	133
B.2.1 Modeling Main Effects	134
B.2.2 Modeling Group-Covariate Effects.....	138
B.2.3 Modeling Individual-Covariate Effects	140
B.2.4 Modeling the Product Probabilities.....	143
B.2.5 Link Functions.....	144
B.3 Detection Probability Estimation with Manly-Parr Estimates.....	146
B.3.1 Closed-Form Estimates.....	146
B.3.2 Modeling Detection Probabilities with Manly-Parr Estimates	147
B.4 Diagnostic Plot Calculations.....	148
B.4.1 Individual Covariate Plots	148
B.4.2 Group Covariate Plots	150
B.5 Analysis of Deviance.....	151
B.5.1 Analysis of Deviance Preferred for Group Covariates.....	155
B.6 Literature Cited.....	156

Chapter 1: Introduction

Program SURPH (SURvival under Proportional Hazards)

SURPH was developed to analyze data from release-recapture studies of animal populations—in particular, to relate the survival estimates from release-recapture studies to individual-based and group-based covariates.

This document describes SURPH 3.5. A PDF version of this document and the latest version of the SURPH program are available on the [Columbia Basin Research SURPH page](#).

Chapter 2 explains the user interface of SURPH; Chapters 3 through 7 give a detailed description of how to use SURPH to analyze release-recapture data, while Chapter 8 walks the user through the analyses of some sample datasets to illustrate the use of SURPH. Appendix A gives a definition of release-recapture input data file, and Appendix B explains the statistical theory behind SURPH.

Please note that in-application dialogs and user interface elements mentioned in this manual are represented with **Arial Bold**. User-defined (changeable) elements within the application are listed in **Courier New Bold** and are particular to the specific example/reference.

Chapter 2: Using SURPH

Figure 2.1 shows the SURPH user interface at startup with no data loaded and the main components in their default position. The main components are:

- The **Navigation Panel** on the left side shows the list of available actions in a hierarchical structure and is explained in more detail in Section 2.1.
- The **Dataset Summary** on the upper right side summarizes the current dataset.
- The **Model Definition Summary** summarizes the current model. By default, it shares the same space as the dataset summary. **Dataset Summary** and **Model Definition Summary** are selected using the tabs at the bottom of their shared window.
- The **Effects Control Buttons** below the **dataset summary** and the **Model Definition Summary** provide shortcut keys for configuring models and will be explained in Chapter 4.
- The **Output** window along the bottom of the user interface shows all status, warning, and error messages as they occur.
- The **Workspace** refers to the shaded, middle portion of the SURPH user interface. This is where all dialogs, reports, and plots are shown in SURPH.

There is also a series of menus across the top of the SURPH user interface that, for the most part, duplicate actions on the **Navigation Panel**.

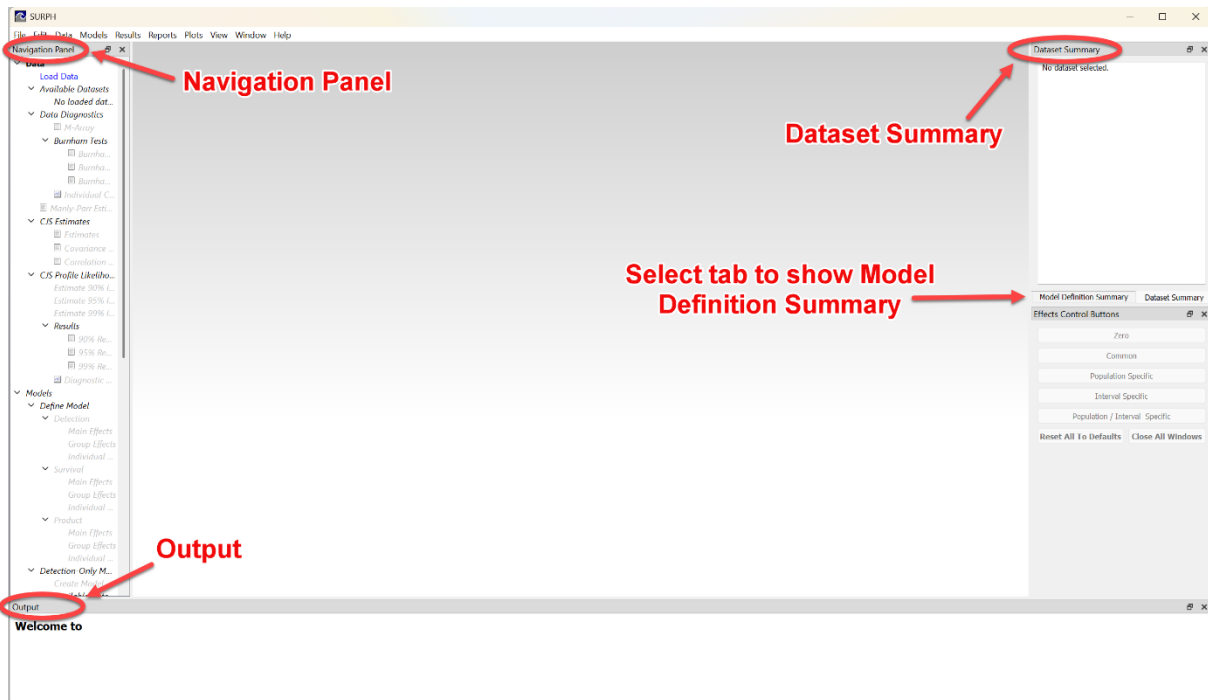


Figure 2.1. Main SURPH user interface at initial startup.

Any of the above components may be resized by dragging their edges or moved from their default positions by dragging their title bars to the desired location.

Figure 2.2 shows an alternative setup with the **Output** window and **Dataset Summary** dialog dragged completely outside of the SURPH main user interface.

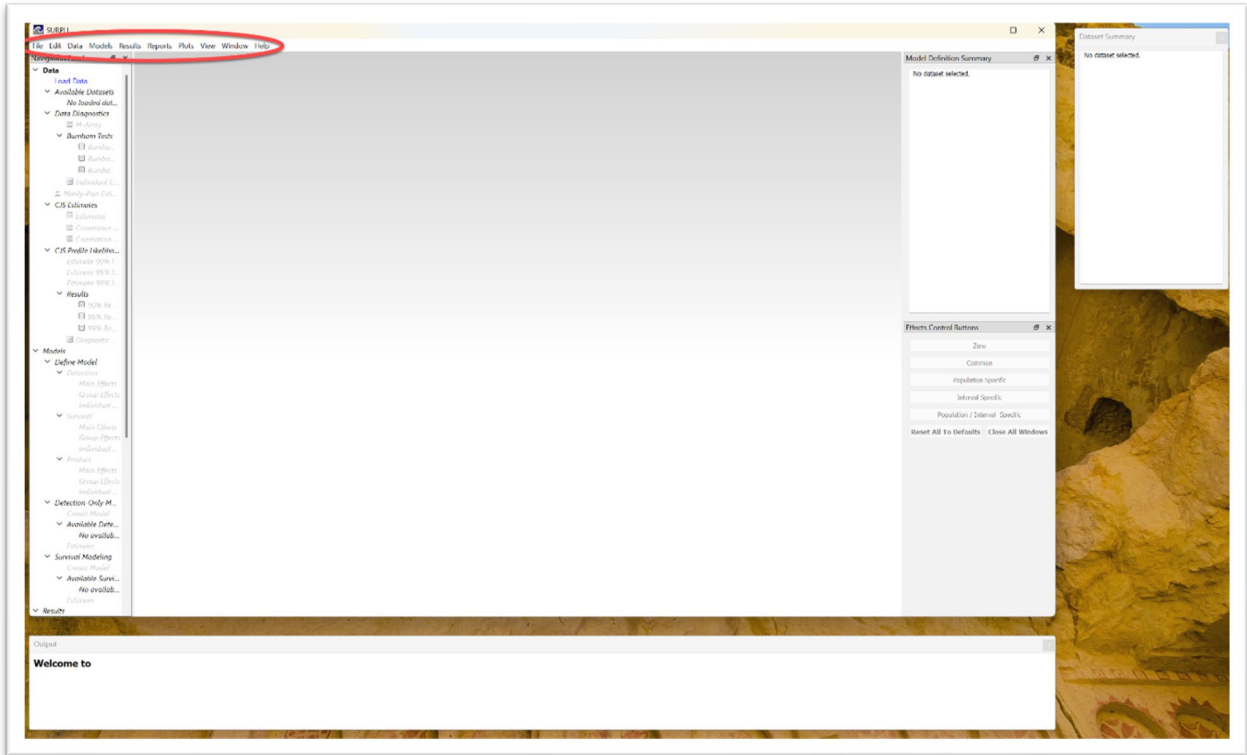


Figure 2.2. An example of an alternative configuration for SURPH that makes use of the full Windows Desktop. The Menu commands across the top are circled for clarity.

2.1 Navigation Panel

The **Navigation Panel** is organized hierarchically and reflects the general flow of work from top to bottom. The four main sections are:

1. **Data** for the loading and examining of release-recapture data into SURPH (Chapter 3).
2. **Models** for defining and creating models (Chapter 4).
3. **Results** for estimating parameters, viewing the results, and looking at result diagnostics (Chapter 5).
4. **Hypothesis Testing** for comparing results (Chapter 6).

2.2 SURPH Menus

The SURPH menus run along the top of the SURPH main user interface. For the most part, they repeat commands found on the **Navigation Panel**, with the exception of those described below.

2.2.1 File Menu

2.2.1.1 Save as

The **Save as** command under the **File** dropdown menu allows the user to save the currently selected report or plot to a file. Reports are saved in Rich Text Format (.rtf) and plots are saved in Portable Network Graphics format (.png). The dimensions of the .png file for plots can be changed on the **Edit Settings** dialog found on the dropdown options after clicking **Edit** on the main SURPH menu (Section 2.2.2.1).

2.2.1.2 Print and Print Preview

The **Print** and **Print Preview** commands found after clicking on File allow the user to print or preview the print of the currently selected report or plot.

2.2.2 Edit Menu

2.2.2.1 Edit Settings

Selecting **Edit Settings** under the **Edit** menu brings up the **Settings** dialog as shown in Figure 2.3. It allows the user to change:

- The number precision in reports for both the estimates and the standard errors. This controls the number of digits after the decimal point.
- The dimensions of a plot in pixels when saved using the **Save as** command under the **File** menu.

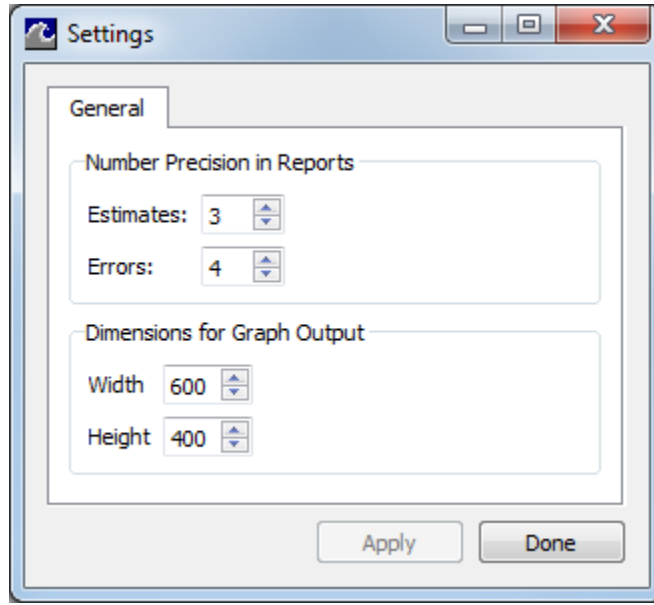


Figure 2.3: SURPH **Settings** menu.

2.2.2.2 *Optimizer Setup*

Figure 2.4 shows the **Optimizer Setup** dialog found by clicking on **Optimizer Settings** from the **Edit** dropdown menu. The optimizer is used to numerically optimize the SURPH likelihood to estimate the parameters. Four optimizers are available:

- Fletcher
- Quasi-Newton (FBSG)
- Direction Set (Powell's)
- Simplex

Experience suggests that Fletcher, a quasi-Newton optimizer, is the best choice in SURPH, except when calculating the profile likelihoods as described in Section 3.4.

Each optimizer has its own configuration settings. The user-editable fields shown in Figure 2.4 (Maximum iterations, Precision, and Proportional step size) are specific to Fletcher. In most cases, the optimizer setting can be left at their default settings. Details of the Quasi-Newton, Direction Set, and Simplex can be found in Press et al. (2007).

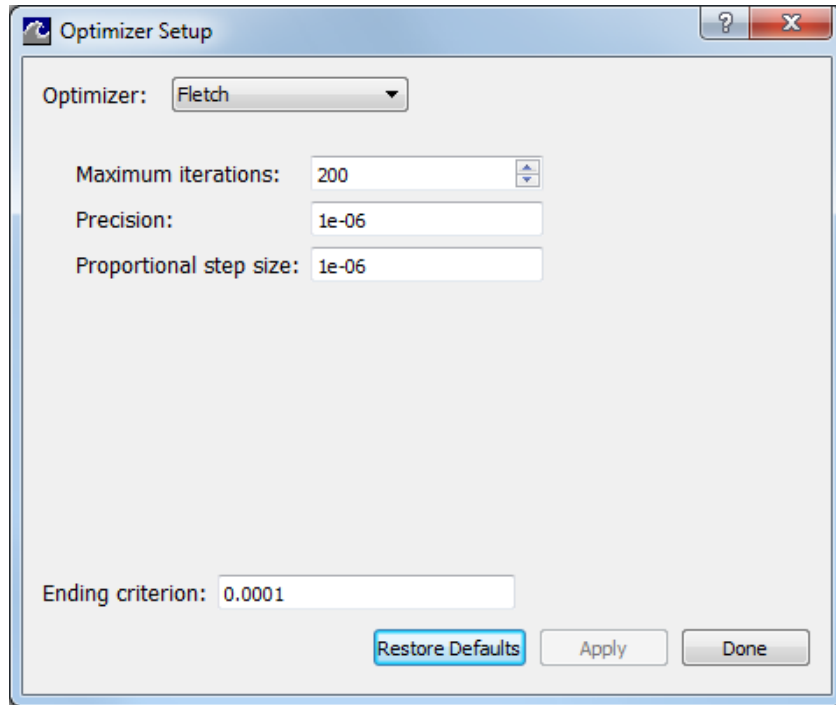


Figure 2.4: SURPH **Optimizer Setup**.

When SURPH estimates the parameters, it uses the resulting estimates as seeds and re-estimates the parameters. It repeats this process until the resulting log-likelihood of the current iteration does not vary from the previous iteration by more than the value in **Ending criterion** at the bottom of the **Optimizer Setup** dialog.

2.2.2.3 *Link Functions*

Link functions are used to model survival-related parameters as a function of the main effects and covariate data. The available link functions can be found by clicking on **Edit** on the main menu. Their use is described in Chapter 4 and more details are given in Appendix B.2.5. The Link Functions dialog is shown in Figure 2.5.

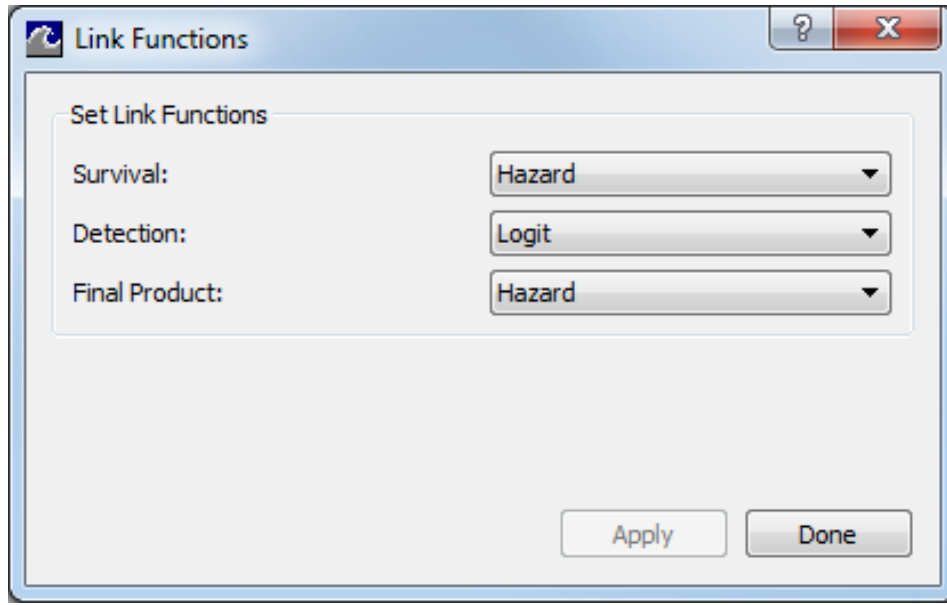


Figure 2.5: SURPH **Link Functions** dialog.

2.2.2.4 *Reset All*

The **Reset All** command on the **Edit** menu will reset SURPH by deleting all results, models, and datasets currently stored in SURPH. If selected, SURPH will show a confirmation dialog asking if the user wants to continue with the Reset All command or to cancel.

2.2.3 *View Menu*

The **View** dropdown menu allows the user to selectively hide or show the main components of the SURPH user interface. Figure 2.6 shows the default settings with all components checked and thus visible. A component may be hidden by unchecking it in the **View** menu.

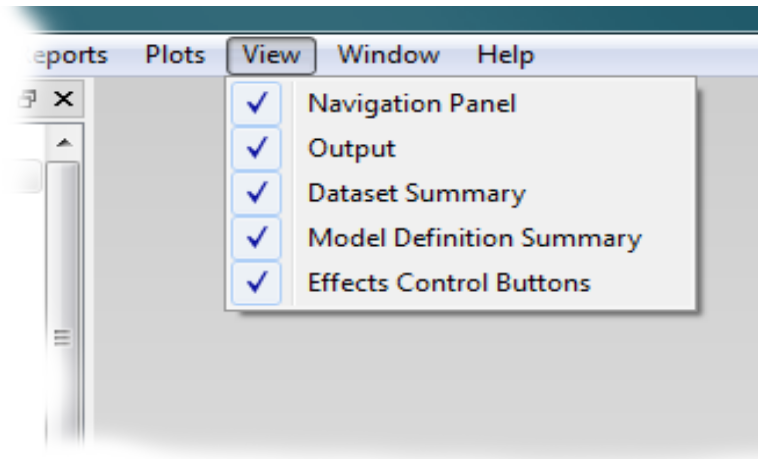


Figure 2.6. The **View** menu, used for showing/hiding the main components of the SURPH user interface.

2.2.4 Help Menu

There are three commands under the **Help** menu. The **About** command brings up a dialog showing the current version of SURPH. **Check for Updates** will look to see if there are newer versions of SURPH available for download. **Configure Updates** allows you to choose if SURPH should automatically download and install updates, how often the program should check for updates, and where it should store your downloaded updates.

2.3 Literature Cited

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. 2007. *Numerical recipes: the art of scientific computing*. New York: Cambridge University Press.

Chapter 3: Data

At initial startup with no dataset loaded, the **Load Data** action under the **Data** heading on the **Navigation Panel** is the only action available. Double-clicking on it will bring up a dialog asking the user to select the SURPH input data file. The format of the input data file is described in Appendix A. The user is asked to enter the name of the dataset, defaulting to the name of the input data file.

Once loaded, the dataset name appears under **Available Datasets** with a green check mark indicating that it is the current dataset. More than one dataset can be loaded at a time. By default, the last dataset loaded is the current dataset; the user may change the current dataset by double-clicking on its name. Figure 3.1 shows the **Navigation Panel** with multiple datasets loaded and the one named `distance` selected as the current dataset.

When a dataset is selected, the **Dataset Summary** window is updated to show the name of the dataset, its description, the number of populations and intervals, and the number tagged in each population, as shown in Figure 3.2.

There are four main sections under the **Data** header in the **Navigation Panel**: (1) **Data Diagnostics**, (2) **Manly-Parr Estimates**, (3) **CJS Estimates**, and (4) **CJS Profile Likelihoods**.

3.1 Data Diagnostics

Under **Data Diagnostics**, SURPH provides an **M-Array** report that summarizes the data, three **Burnham Tests** that test model assumptions for release-recapture data, and an **Individual Covariate Graph** for datasets with individual covariate data.

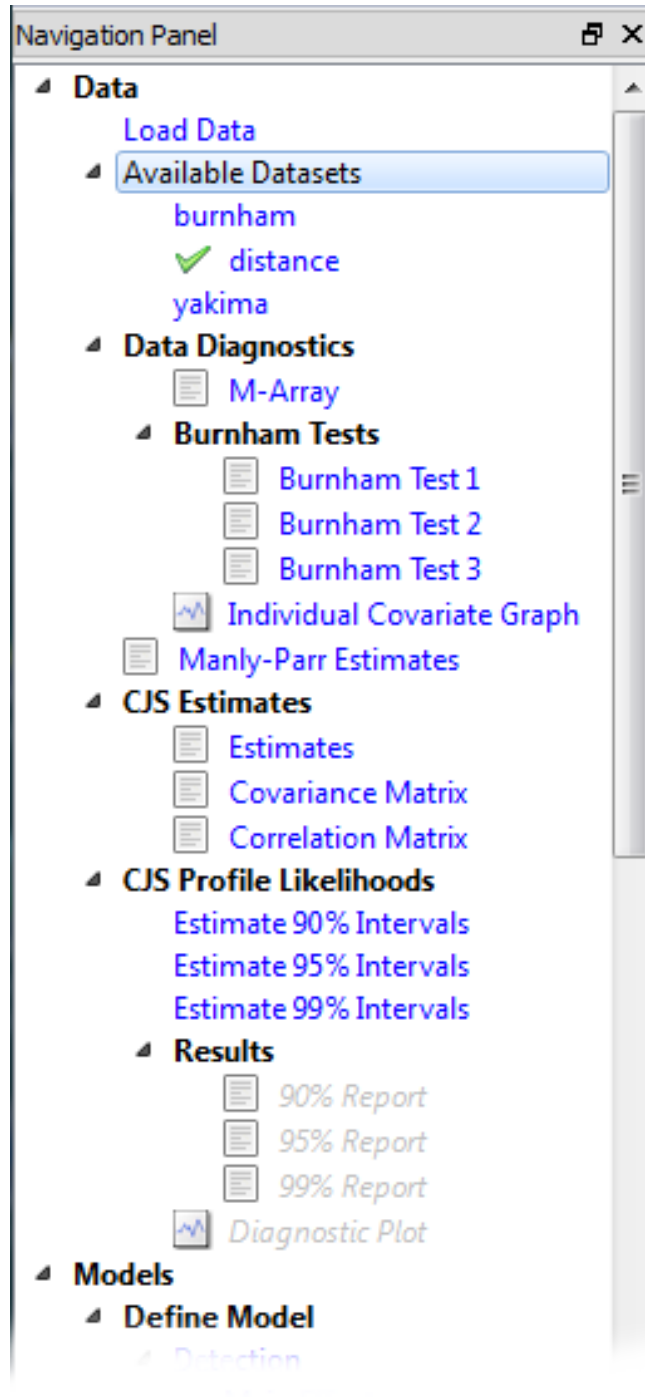


Figure 3.1. The **Data** section of the **Navigation Panel** with multiple datasets loaded and **distance** as the active dataset.

Dataset Summary	
<u>distance</u>	
Yearling Chinook Releases from Snake River Hatcheries - 1993	
Populations: 10	
Intervals: 3	
Population	Total Tagged
1	1467
2	1460
3	1445
4	1171
5	999
6	1991
7	1501
8	2993
9	2985
10	799

Figure 3.2. An example of the **Dataset Summary** report.

3.1.1 M-Array

Double-clicking on **M-Array** on the **Navigation Panel** will bring up the M-Array report with an M-Array for each population, as shown in Figure 3.3. The M-Array is taken from Burnham et al. (1987) and gives a summary of the release-recapture data. For population 1 in Figure 3.3 as an example, 1,467 tagged individuals were released initially. Of these releases, 422 were next detected at occasion 1 with 211 of them removed (censored). At occasion 2, there were 201 individuals that were first detected there without having been previously detected at occasion 1. Of those 201 detections, 76 were removed. Figure 3.3 shows that 719 of the 1,467 were never detected again. The second row of the M-Array shows the detections of the 211 tagged individuals that were detected at occasion 1 and not removed there: 78 were subsequently detected at occasion 2, 49 were detected at occasion 3 without prior detection at occasion 2, and 84 were not

detected after occasion 1.

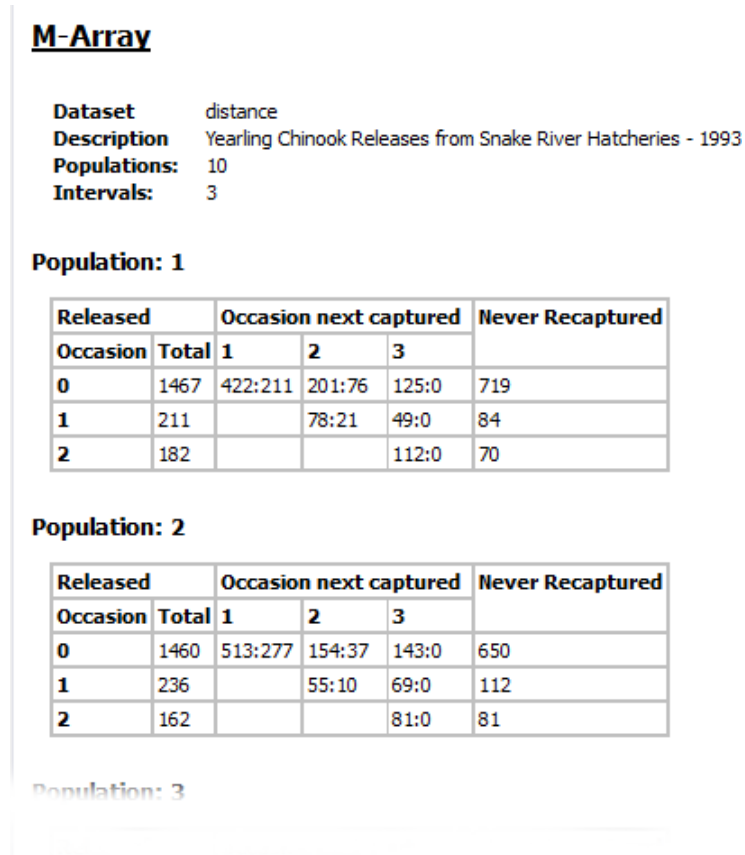


Figure 3.3. Top portion of an **M-Array** report. There is an M-Array for each population.

3.1.2 Burnham Tests

Burnham Test 1, **Burnham Test 2**, and **Burnham Test 3** are taken from Burnham et al. (1987). Burnham Test 1 compares survival parameters across populations and Burnham Tests 2 and 3 test for violation of model assumptions.

Each Burnham test consists of a series of contingency tables with an associated χ^2 statistic and *P*-value. Two tables from a **Burnham Test 1** are shown in Figure 3.4. Their structure and content are specific to each test and are explained in detail in Burnham et al. (1987).

TEST 1.R0

	Detected after 0	Not detected after 0
Treatment detected at 0	748	719
Control detected at 0	810	650

Chi-square=5.75 (df=1) P=0.02

TEST 1.R1

	Detected after 1	Not detected after 1
Treatment detected at 1	127	84
Control detected at 1	124	112

Chi-square=2.34 (df=1) P=0.13

TEST 1.R2

	Detected after 2	Not detected after 2
Treatment detected at 2	112	70
Control detected at 2	81	81

Chi-square=4.18 (df=1) P=0.04

TEST 1.T1

	Detected after 0	Not detected after 0
Treatment detected at 0	748	719
Control detected at 0	810	650

Chi-square=7.47 (df=1) P=0.01

TEST 1.T2

	Detected after 1	Not detected after 1
Treatment detected at 1	127	84
Control detected at 1	124	112

Chi-square=12.15 (df=1) P=0.00

Figure 3.4. Some contingency tables from a **Burnham Test 1** example.

At the bottom of each test report is a summary table showing the χ^2 values (labeled **Chi-square** in the tables) for each test and the associated *P*-values. Figure 3.5 shows the summary table for a **Burnham Test 1**. The overall χ^2 is 31.88 with a *P*-value of 0.00.

TEST	Chi-square	df	P
TEST 1.R0	5.75	1	0.02
TEST 1.R1	2.34	1	0.13
TEST 1.R2	4.18	1	0.04
TEST 1.T1	7.47	1	0.01
TEST 1.T2	12.15	1	0.00
TEST 1	31.88	5	0.00

Figure 3.5. Summary table for a **Burnham Test 1** example.

3.1.2.1 Burnham Test 1

Burnham Test 1 compares two populations, one identified as the treatment group and the other identified as the control group. It tests H_0 : all survival and detection parameters are the same across treatment groups, versus H_A : at least some parameters differ between or among groups. As shown in Figure 3.6, there are controls at the top of the report allowing the user to select the treatment population and the control population.

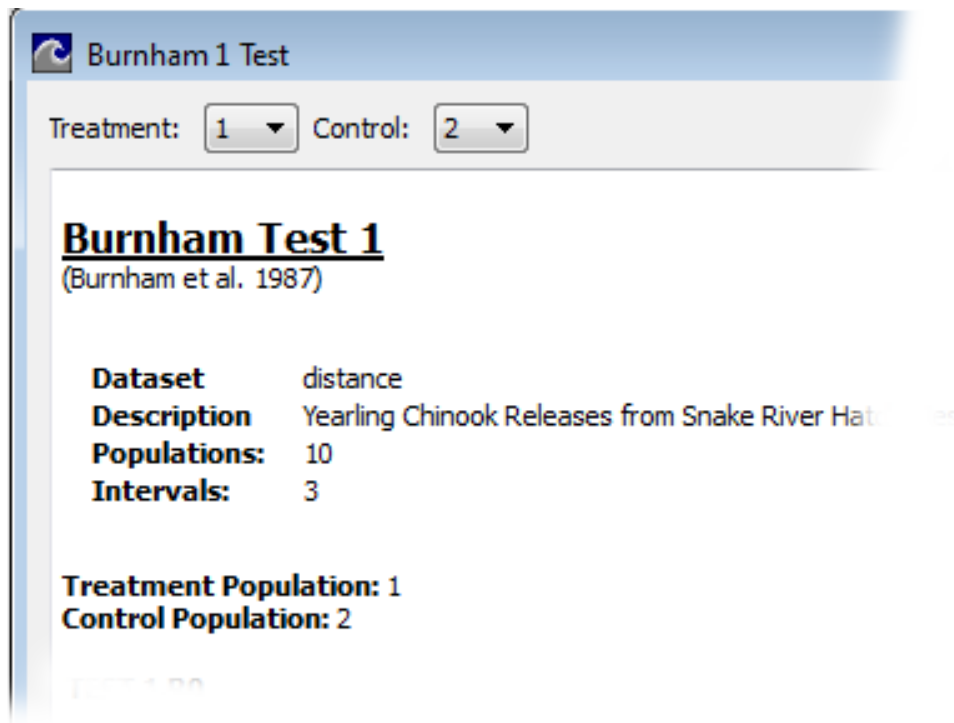


Figure 3.6. Top portion of a **Burnham Test 1** report, showing controls allowing user to select treatment population and control population.

3.1.2.2 Burnham Tests 2 and 3

Burnham Tests 2 and 3 look at one population and test for the violation of model assumptions. Both reports have a control at the top to allow the user to select the population of interest, as shown for **Burnham Test 2** in Figure 3.7. Burnham Tests 2 and 3 are not available if there are fewer than three detection sites after the initial release event.

The specifics of the tests are:

- **Test 2:** Test for goodness-of-fit of the model to the data.

H_0 : The survival and detection parameters are specific to sampling occasions within each group.

H_A : The model does not fit the data; there may be tagging effects or differential behavior.

- **Test 3:** Tests for parameters that are specific to individual capture histories.

H_0 : The survival and detection parameters do not depend on the capture histories of fish released on any release occasion.

H_A : Some of the parameters are dependent on the capture histories of fish in a given release.

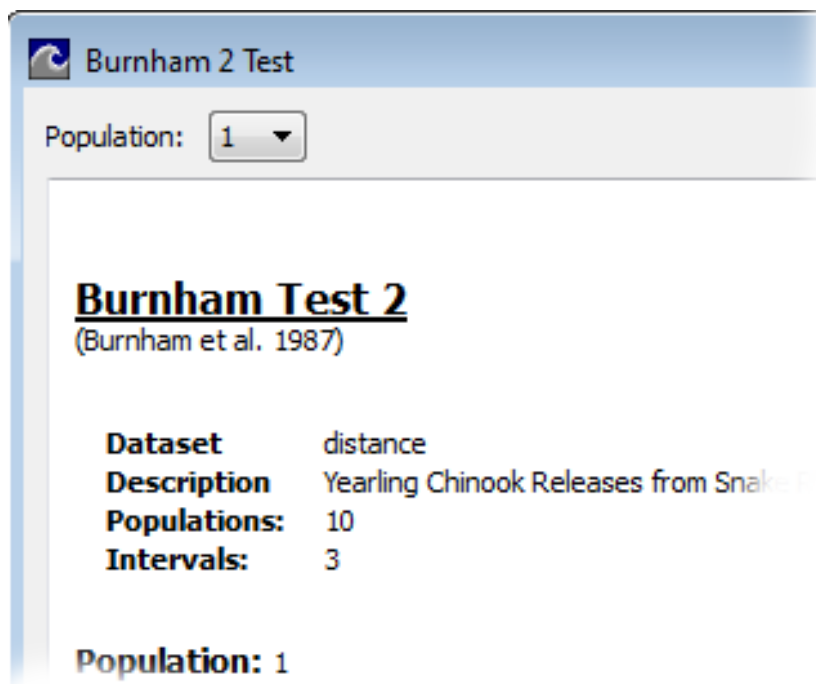


Figure 3.7. Top portion of a **Burnham Test 2** report, showing control allowing user to select the population of interest.

3.1.3 Individual Covariate Graph

If the active dataset includes one or more individual covariates, SURPH 3 allows the user to explore the individual covariate data. Figure 3.8 shows the individual covariate plot for a dataset with (at least) two individual covariates: **length** and **weight**. Under the **Data Selector** section at the lower left of the plot, the user has selected **length** under the **Covariate** selector, prompting a histogram and the empirical Cumulative Distribution Function (CDF) plots to be shown for the individual covariate **length**.

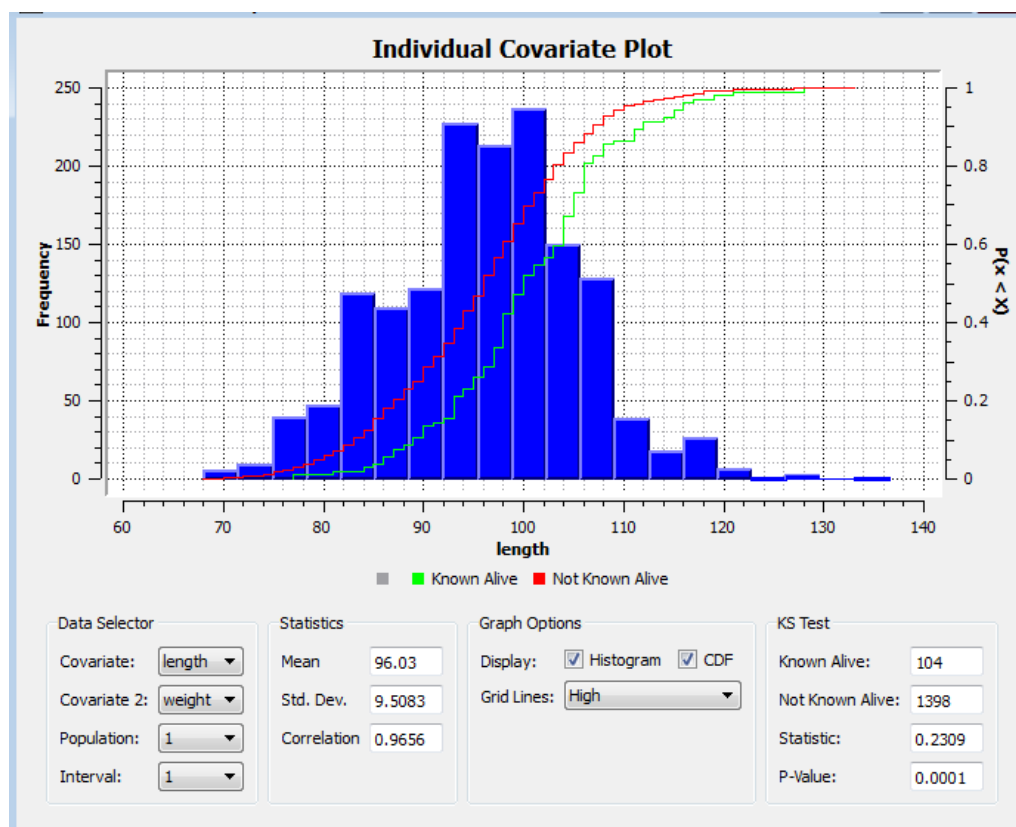


Figure 3.8. Individual Covariate Plot.

Under **Statistics**, the **Mean** and **Std. Dev.** (standard deviation) of the selected covariate **length** are shown for the selected population. The individual covariate **weight** is selected for **Covariate 2** and, thus, the correlation between **length** and **weight** is shown for the selected **Population** and **Interval**. The high correlation between the two individual covariates suggests that for creating models that include individual covariate

data, the **weight** may not add much information to a model that already includes **length**.

There are two CDF plots displayed for a given individual covariate on the **Individual Covariate Plot** (Figure 3.8): Individuals known alive (green) and those not known alive (red). The **KS Test** section refers to the Kolmogorov-Smirnov (K-S) Test (Conover, 1980) for comparing two probability distributions. The two distributions being compared are (1) those known to be alive for the given population and period and (2) those not known to be alive (i.e., not detected at this period or later) for the chosen covariate (**length** in this case). A significant *P*-value suggests that the two distributions are not from the same population, indicating that the chosen covariate may have a significant effect on survival and/or detection. Under the heading **Graph Options**, there are check boxes allowing the user to hide the **Histogram** and/or **CDF** by checking or unchecking the boxes. You can also choose what level of **Grid Lines** are displayed on the plot, from high to none.

3.2 Manly-Parr Estimates

The **Manly-Parr Estimates** report displays the Manly-Parr estimates of detection probability. Developed by Manly and Parr (1968), they provide estimates of detection probability that are independent of the survival process. These estimates may guide the user when selecting the appropriate **Detection Only** model in the modeling process (Section 4.5).

Figure 3.9 shows an example of a **Manly-Parr Estimates** report. The estimates in the final table—**Estimates adjusted for removal**—are unbiased. Details of the Manly-Parr estimates are given in Appendix B.3.

Manly-Parr Estimates

Dataset burnham
Description Example Data Set from Burnham "Blue Book"
Populations: 2
Intervals: 5

Unadjusted Estimates

Population	p1	p2	p3	p4
treatment	0.040 (128/3174)	0.011 (32/2957)	0.072 (93/1288)	0.027 (19/691)
control	0.040 (136/3427)	0.010 (31/3198)	0.074 (102/1376)	0.034 (26/758)

Removal proportions

Population	p1	p2	p3	p4
treatment	0.028 (29/1029)	0.056 (14/249)	0.048 (85/1762)	0.042 (26/616)
control	0.030 (33/1104)	0.038 (10/260)	0.032 (62/1924)	0.043 (28/644)

Estimates adjusted for removal

Population	p1	p2	p3	p4
treatment	0.041	0.011	0.076	0.029
control	0.041	0.010	0.076	0.036

Figure 3.9. Example of **Manly-Parr Estimates** report of detection probabilities.

3.3 CJS Estimates

The Cormack-Jolly-Seber (CJS) estimates (Cormack, 1964; Jolly, 1965; Seber, 1965) are the “best” estimates of the survival-related parameters for release-recapture data. They are described in detail in Appendix B.1.6. Figure 3.10 shows the report that is brought up in the workspace when the user double-clicks on **CJS Estimates**, showing the CJS estimates for the current dataset. In this case, there are ten populations and three intervals. The **s1** column is the survival estimate for the first interval and **p1** is the corresponding detection probability for the first sampling event. The **final product** is the

product of survival and capture probability for the third and final interval (λ in Appendix B.1.6). The **overall s** is the total survival probability—in this case, the product of **s1** and **s2**.

Selecting **Covariance Matrix** or **Correlation Matrix** under **CJS Estimates** will bring up the covariance matrix or correlation matrix corresponding to the CJS estimates, as shown in Figures 3.11 and 3.12. The reports show one matrix for each population.

CJS Estimates

Dataset distance
Description Yearling Chinook Releases from Snake River Hatcheries - 1993
Populations: 10
Intervals: 3

Population	s1	s2	p1	p2	final product	overall s
1	0.657 (0.0273)	0.746 (0.0480)	0.438 (0.0228)	0.497 (0.0282)	0.615 (0.0361)	0.490 (0.0237)
2	0.739 (0.0307)	0.790 (0.0654)	0.476 (0.0238)	0.330 (0.0277)	0.500 (0.0393)	0.583 (0.0395)
3	0.835 (0.0606)	0.616 (0.0738)	0.334 (0.0275)	0.283 (0.0286)	0.438 (0.0411)	0.515 (0.0432)
4	0.668 (0.0428)	0.708 (0.0863)	0.436 (0.0319)	0.299 (0.0343)	0.474 (0.0507)	0.473 (0.0450)
5	0.672 (0.0230)	0.870 (0.0408)	0.492 (0.0232)	0.574 (0.0284)	0.663 (0.0330)	0.584 (0.0241)
6	0.669 (0.0250)	0.759 (0.0468)	0.432 (0.0199)	0.405 (0.0242)	0.541 (0.0310)	0.508 (0.0248)
7	0.563 (0.0279)	0.731 (0.0518)	0.403 (0.0245)	0.468 (0.0286)	0.639 (0.0348)	0.411 (0.0211)
8	0.503 (0.0171)	0.804 (0.0438)	0.457 (0.0183)	0.399 (0.0220)	0.541 (0.0274)	0.405 (0.0183)
9	0.680 (0.0171)	0.866 (0.0414)	0.494 (0.0155)	0.355 (0.0186)	0.521 (0.0253)	0.589 (0.0243)
10	0.264 (0.0238)	1.153 (0.1946)	0.440 (0.0454)	0.248 (0.0481)	0.432 (0.0747)	0.305 (0.0482)

Figure 3.10. **CJS Estimates** report for a dataset with ten populations and three periods.

CJS Covariance Matrix

Dataset distance
Description Yearling Chinook Releases from Snake River Hatcheries - 1993
Populations: 10
Intervals: 3

Population 1:

	s1	s2	p1	p2
s1	0.000747396	-0.000863824	-0.000395854	0
s2		0.00230491	0.00057591	-0.000530117
p1			0.000519355	0
p2				0.000797768

Population 2:

	s1	s2	p1	p2
s1	0.000945462	-0.0011706	-0.000523869	0
s2		0.00428106	0.000754109	-0.000998424
p1			0.00056877	0
p2				0.000797768

Figure 3.11. Top portion of **CJS Covariance Matrix** report. There is one matrix for each population.

CJS Correlation Matrix

Dataset distance
Description Yearling Chinook Releases from Snake River Hatcheries - 1993
Populations: 10
Intervals: 3

Population 1:

	s1	s2	p1	p2
s1	1.0000	-0.6581	-0.6354	0.0000
s2		1.0000	0.5264	-0.3909
p1			1.0000	0.0000
p2				1.0000

Population 2:

	s1	s2	p1	p2
s1	1.0000	-0.5819	-0.7144	0.0000
s2		1.0000	0.4833	-0.5511

Figure 3.12. Top portion of **CJS Correlation Matrix** report. There is one matrix for each population.

3.4 CJS Profile Likelihoods

The **CJS Estimates** section provides point estimates of the survival-related parameters. From the standard errors provided, one can construct a confidence interval based on normal probability theory. In contrast, confidence intervals based on profile likelihoods provide a nonparametric alternative. In most instances, there will be close agreement between the two methods. In cases where they do not agree, Lowther and Skalski (1996) propose that the profile likelihood confidence intervals should be used.

The theory behind profile likelihood confidence intervals and the method for calculating them can be found in Lowther and Skalski (1996).

SURPH 3 allows the user to calculate 90%, 95%, and 99% profile likelihood confidence intervals for the CJS survival parameters and the overall survival as shown in Figure 3.13. Once the estimation is complete, the corresponding **Profile Likelihood Estimates** report is available. In Figure 3.13, the 95% confidence interval has been estimated and the report is available (Figure 3.14). If you want to re-estimate estimates that are already calculated—with a different optimizer, for example—simply double-click in this case on **Estimate 95% Intervals**. The previous estimates will be deleted and replaced with the new estimates.

An unfortunate aspect of profile likelihood estimation in SURPH is that the validity of the results is somewhat optimizer-dependent—especially when estimating the confidence interval of a function of parameters (e.g., **overall s**). The estimates in Figure 3.14 were calculated using the Fletch optimizer; the estimates in Figure 3.15 were produced using the Direction Set (Powell's) optimizer. Notice that the estimates agree except in the case of the **overall s** parameter—the product of the interval-specific survival probabilities.

For this reason, SURPH provides a **CJS Profile Likelihood Diagnostic Plot**. Figure 3.16 shows the diagnostic plot for the Fletch optimizer, corresponding to Figure 3.14, and Figure 3.17 shows the corresponding plot for estimates from the Direction Set optimizer, corresponding to Figure 3.15. As can be seen in Figures 3.16 and 3.17, there are controls in the lower left to select the desired population, parameter, and confidence interval to view. The two vertical red lines indicate the upper and lower boundary of the interval.

The horizontal red line indicates the target likelihood ratio corresponding to the desired confidence interval. The blue plus signs (+) correspond to the values that the profile likelihood estimation algorithm uses to search for the boundary values. As the search algorithm gets close to the desired likelihood ratio value, the step size gets smaller—which is why the points are closer together near the boundary values.

In Figure 3.17, the points trace out a smooth curve on both sides of the boundaries.

Compare this with Figure 3.16, where there is a lot of “noise” in the points around the boundaries. This indicates that the estimates from the Direction Set optimizer should be used instead of the estimates from the Fletch optimizer. Experience seems to indicate that while Fletch is the “best” optimizer for estimating model parameters, the Direction Set optimizer seems preferable for estimating the profile likelihood confidence intervals. It is recommended that the **CJS Profile Likelihood Diagnostic Plot** be examined before using the estimated profile likelihood confidence interval.

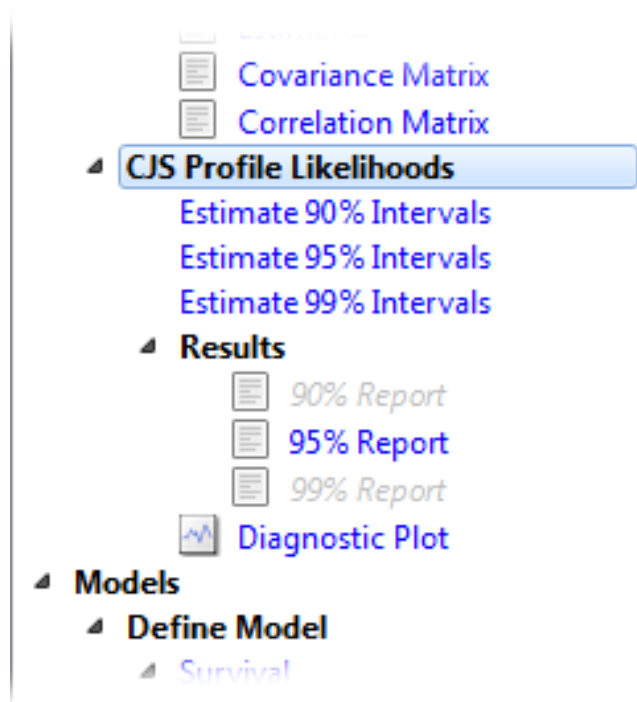


Figure 3.13. **CJS Profile Likelihoods** section of the SURPH 3 **Navigation Panel**.

95% CJS Profile Likelihood Estimates

Dataset distance
Description Yearling Chinook Releases from Snake River Hatcheries - 1993
Populations: 10
Intervals: 3
Optimizer: Fletch maxIterations=200, stepSize=1e-06, precision=1e-06

Population: 1

Parameter	Estimate	s.e.	Lower boundary	Upper boundary
s1	0.657	0.0273	0.607	0.716
s2	0.746	0.0480	0.672	0.844
overall s	0.490	0.0237	0.448	0.535

Population: 2

Parameter	Estimate	s.e.	Lower boundary	Upper boundary
s1	0.739	0.0307	0.683	0.805
s2	0.790	0.0654	0.671	0.931
overall s	0.583	0.0395	0.518	0.664

Population: 3

Parameter	Estimate	s.e.	Lower boundary	Upper boundary
s1	0.835	0.0606	0.732	0.975
s2	0.616	0.0738	0.485	0.778
overall s	0.515	0.0432	0.443	0.609

Figure 3.14. **CJS Profile Likelihood Estimates** report using the Fletch optimizer.

95% CJS Profile Likelihood Estimates

Dataset distance
Description Yearling Chinook Releases from Snake River Hatcheries - 1993
Populations: 10
Intervals: 3
Optimizer: Direction Set (Powell's), maxIterations=200, funcTolerance=1e-08

Population: 1

Parameter	Estimate	s.e.	Lower boundary	Upper boundary
s1	0.657	0.0273	0.607	0.716
s2	0.746	0.0480	0.655	0.844
overall s	0.490	0.0237	0.461	0.522

Population: 2

Parameter	Estimate	s.e.	Lower boundary	Upper boundary
s1	0.739	0.0307	0.683	0.805
s2	0.790	0.0654	0.671	0.931
overall s	0.583	0.0395	0.547	0.625

Population: 3

Parameter	Estimate	s.e.	Lower boundary	Upper boundary
s1	0.835	0.0606	0.732	0.975
s2	0.616	0.0738	0.485	0.778
overall s	0.515	0.0432	0.468	0.571

Figure 3.15. **CJS Profile Likelihood Estimates** report using the Direction Set (Powell's) optimizer.

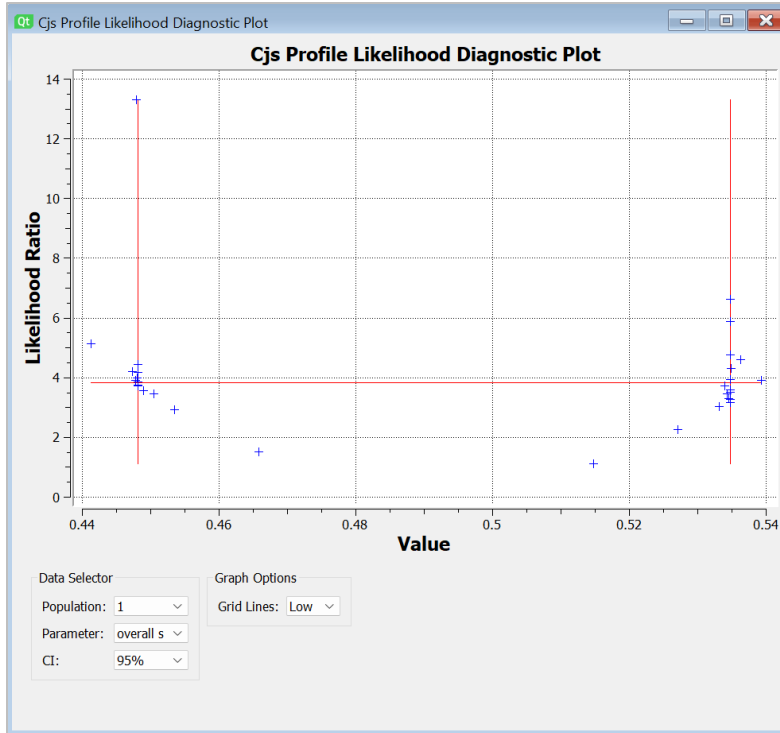


Figure 3.16. **CJS Profile Likelihood Diagnostic Plot** from the Fletcher optimizer.

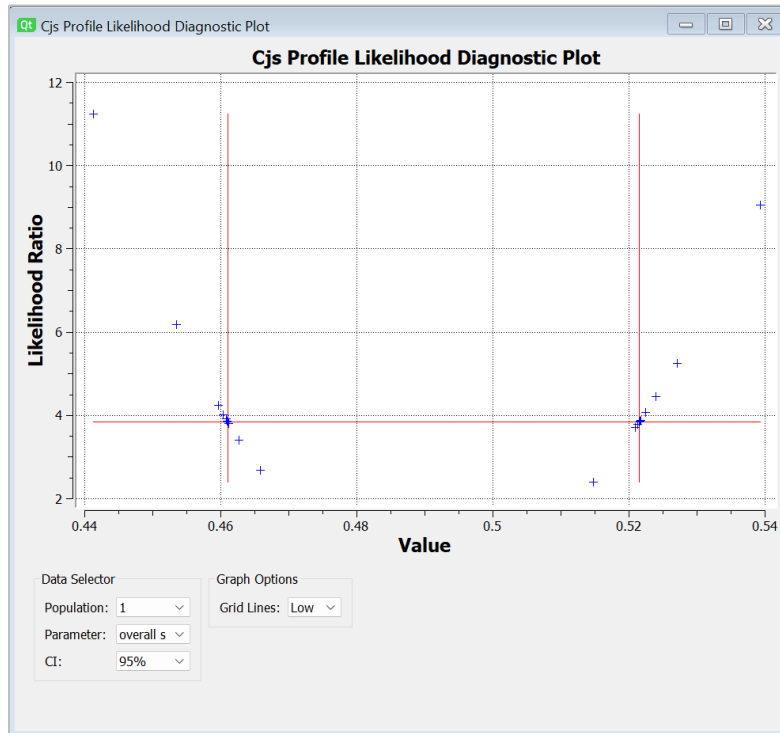


Figure 3.17. **CJS Profile Likelihood Diagnostic Plot** from the Direction Set (Powell's) optimizer.

3.5 Literature Cited

- 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* 5.
- Conover, W. J. 1980. Practical nonparametric statistics. Wiley & Sons, New York.
- Cormack, R. M. 1964. [Estimates of survival from the sighting of marked animals.](#) *Biometrika* 51:429–438.
- Jolly, G. M. 1965. [Explicit estimates from capture-recapture data with both death and immigration-stochastic model.](#) *Biometrika* 52:225–247.
- Lowther, A. B., and J. R. Skalski. 1998. [A multinomial likelihood model for estimating survival probabilities and residualization for fall chinook salmon \(*Oncorhynchus tshawytscha*\) smolt using release-recapture methods.](#) *Journal of Agricultural, Biological, and Environmental Statistics* 3:223–236.
- Manly, B. F. J., and M. J. Parr. 1968. A new method of estimating population size, survivorship, and birth rate from capture-recapture data. *Transactions of the Society for British Entomology* 18:81–89.
- Seber, G. A. F. 1965. [A note on the multiple recapture census.](#) *Biometrika* 52:249–259.

Chapter 4: Models

SURPH uses the term **Models** to encapsulate a particular parameterization of the detection, survival, and product probabilities. When the parameters are estimated, the parameter estimates are stored as **Results**. In this chapter, we look at creating the models.

SURPH supports two approaches to modeling:

1. Joint survival and capture modeling, where the survival and detection probabilities are modeled concurrently.
2. Two-stage modeling, where the detection probabilities are modeled independent of the survival process. Once an appropriate model has been selected for the detection probabilities, the user then models the survival and product probabilities.

Section 4.4 of this chapter describes joint modeling process and Section 4.5 describes the two-stage approach to modeling. First, however, we look at setting the model parameterization via the **Effects** dialogs.

4.1 Using Effects Dialog and Effects Control Buttons

Figure 4.1 shows the **Main Effects** section of the SURPH **Navigation Panel**. Under **Define Model**, there are three main sections: **Detection** probabilities, **Survival** probabilities, and the final **Product** probabilities. Each section has effects dialogs for the **Main Effects**, the **Group Effects** covariates, and the **Individual Effects** covariates. A right-pointing carat indicates which effects dialog is currently active, if any. In Figure 4.1, the **Main Effects** dialog for detection probabilities is currently active.

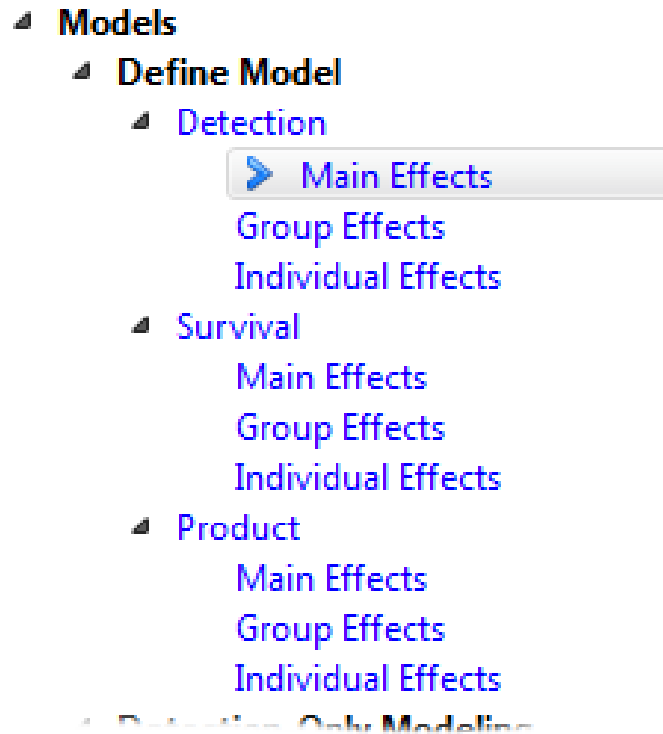


Figure 4.1. The **Main Effects** dialog section of **Navigation Panel**.

Figure 4.2 shows the **Effects Control Buttons**, providing shortcuts for setting commonly used parameterizations. The buttons affect the currently active effects dialog and are as follows:

- **Zero** sets the values in the active group covariate or individual covariate effects dialog to all zeros, indicating that the covariate is not to be included in the model. If a main effects dialog is active, the **Zero** button is grayed out (disabled).
- **Common** sets the values to all ones, indicating a common parameter across all populations and intervals.
- **Population Specific** assumes a unique parameter for each population and common parameters across intervals.
- **Interval Specific** assumes a unique parameter for each interval and common parameters across populations.
- **Population/Interval Specific** assumes a unique parameter for each population and each interval.

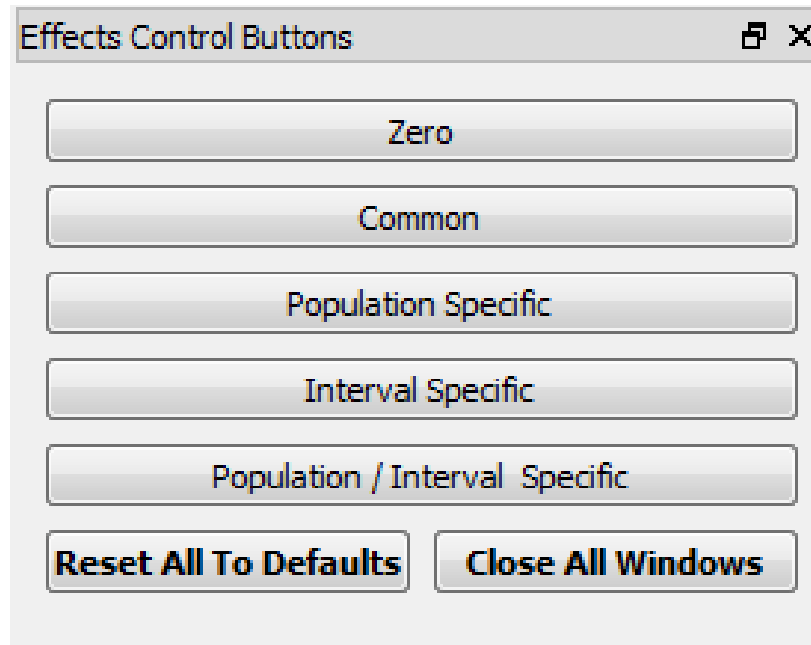


Figure 4.2. The **Effects Control Buttons**.

4.2 Modeling Main Effects

Figure 4.3 shows the **Main Effects** window for a dataset with 10 populations and 3 intervals (2 intervals for which the survival probabilities are estimable). In this case, the main effects are set for population-specific and interval-specific models. This is the population/interval-specific setting corresponding to Table B.4 in Appendix B.2.1.1. It is the default setting for the main effects window and can be set via the **Population/Interval Specific** effects control button.

	Interval 1	Interval 2
1	1	2
2	3	4
3	5	6
4	7	8
5	9	10
6	11	12
7	13	14
8	15	16
9	17	18
10	19	20

Figure 4.3. A **Survival - Main Effects** dialog assuming unique survival probabilities across populations and intervals.

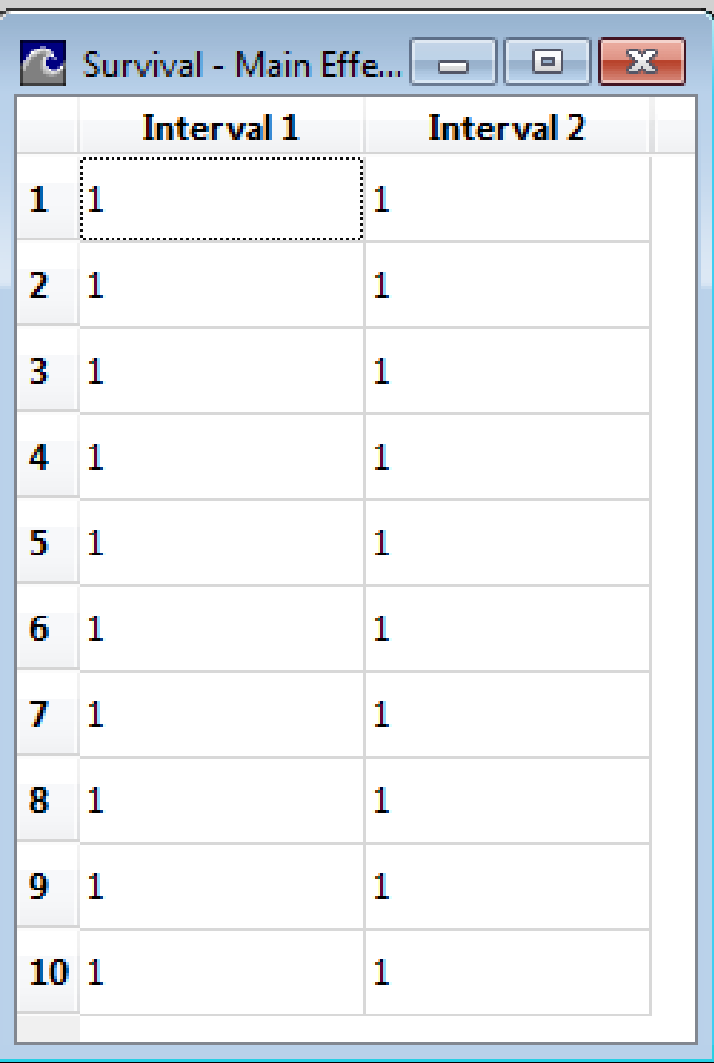
Figure 4.4 shows the **Main Effects** dialog with a common parameter for each population and interval. It can be set using the **Common** effects control button, and corresponds to Table B.5, Appendix B.2.1.2.

Figure 4.5 shows the population-specific parameterization, set via the **Population Specific** button, corresponding to Table B.6, Appendix B.2.1.3. Similarly, Figure 4.6 shows an interval-specific parameterization corresponding to Table B.7, Appendix B.2.1.4

Figure 4.7 shows a custom parameterization where unique parameters are assumed

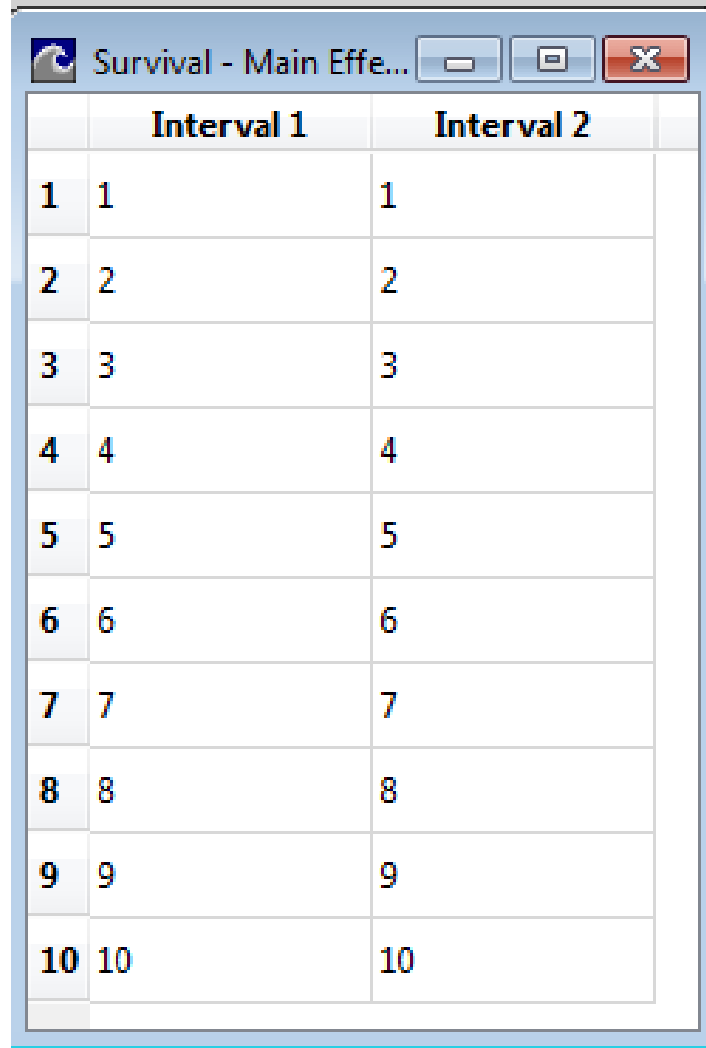
for each interval for the first 3 populations, a common parameter is assumed across intervals for populations 4 through 6, and unique parameters across intervals for populations 7 through 10. This can be achieved by, say, using the **Interval Specific** effects control button and then manually editing the value in interval one for populations 4 through 10. These settings correspond to Table B.8 of Appendix B.2.1.5.

Note that the **Main Effects** dialog for the product term has a row for each population but only one column.



	Interval 1	Interval 2
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
6	1	1
7	1	1
8	1	1
9	1	1
10	1	1

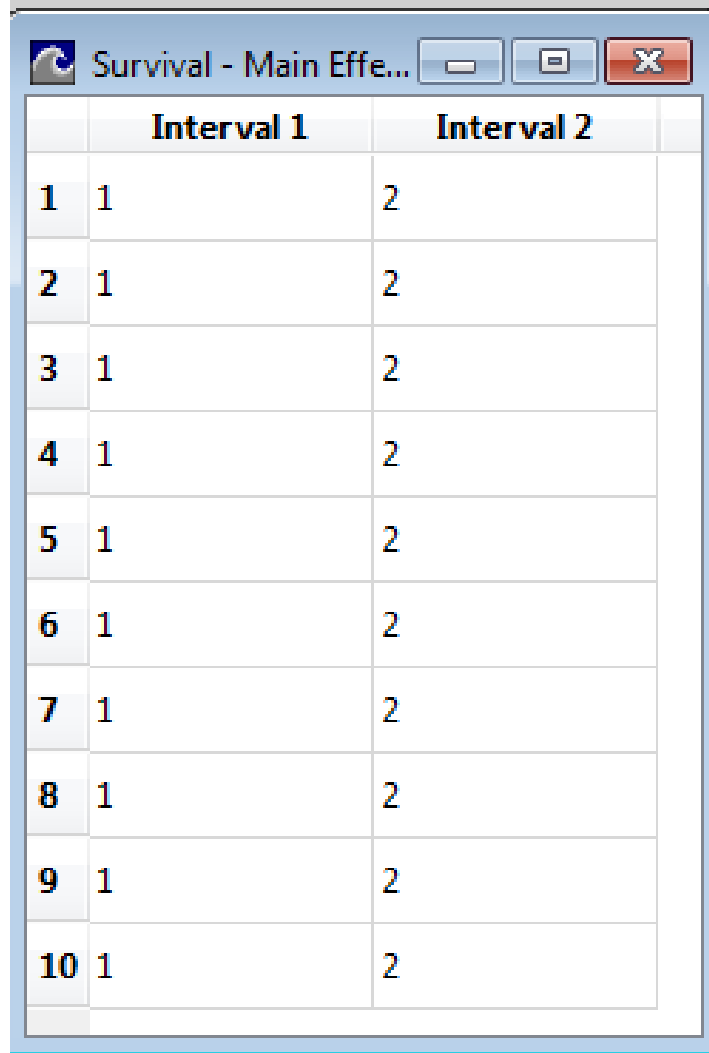
Figure 4.4. A **Survival - Main Effects** dialog assuming a common survival probability across populations and intervals.



The image shows a software dialog box titled "Survival - Main Effects". It contains a table with two columns, "Interval 1" and "Interval 2", and 10 rows. Each row contains the same number in both columns, ranging from 1 to 10. The dialog box has standard window controls (minimize, maximize, close) in the top right corner.

	Interval 1	Interval 2
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10

Figure 4.5. A **Survival - Main Effects** dialog assuming a population-specific model for the survival probabilities.



	Interval 1	Interval 2
1	1	2
2	1	2
3	1	2
4	1	2
5	1	2
6	1	2
7	1	2
8	1	2
9	1	2
10	1	2

Figure 4.6. A **Survival - Main Effects** dialog assuming an interval-specific model for the survival probabilities.

	Interval 1	Interval 2
1	1	2
2	1	2
3	1	2
4	3	3
5	3	3
6	3	3
7	4	5
8	4	5
9	4	5
10	4	5

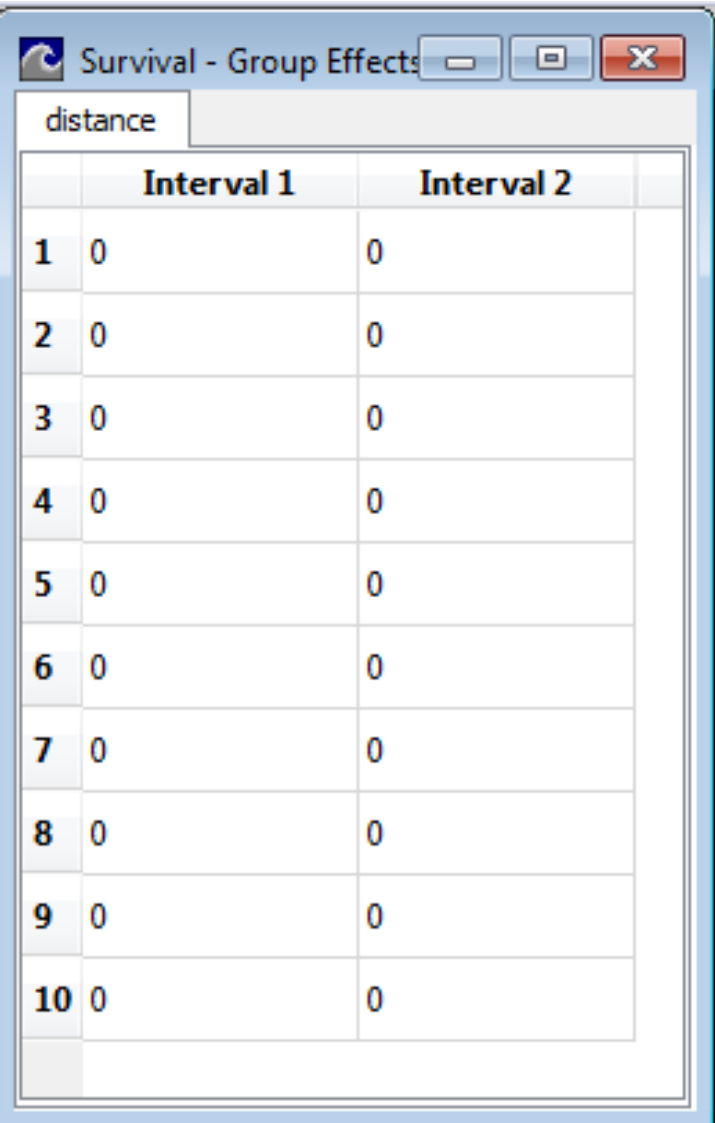
Figure 4.7. A **Survival - Main Effects** dialog for a custom parameterization.

4.3 Modeling Group and Individual Covariate Effects

Group covariate effects and individual covariate effects are modeled in much the same way as the main effects, with the difference that the number zero (0) is allowed in a cell of a group covariate effect window, indicating that the covariate effect is not included for the given population and interval.

Figure 4.8 shows the default settings for a covariate effects window, indicating that no effects for the given covariate are included in the model definition. This can also be set using the **Zero** model effects control button. Notice that the name of the covariate (**distance** in Figure 4.8) is listed in the tab at the top. If there is more than one covariate,

there will be a tab for each covariate and the user selects which covariate to work with by selecting the appropriate tab.



The screenshot shows a software dialog box titled "Survival - Group Effects" with a tab labeled "distance". Inside the dialog is a table with two columns, "Interval 1" and "Interval 2", and ten rows numbered 1 through 10. Each cell in the table contains the number "0".

	Interval 1	Interval 2
1	0	0
2	0	0
3	0	0
4	0	0
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0

Figure 4.8. A **Survival - Group Effects** covariate dialog with the default setting of no covariate effect.

Figure 4.9 shows the settings for a model definition in which the group covariate **distance** is assumed to affect survival in all populations the same way in the first interval and is assumed to not affect survival in the second interval.

	Interval 1	Interval 2
1	1	0
2	1	0
3	1	0
4	1	0
5	1	0
6	1	0
7	1	0
8	1	0
9	1	0
10	1	0

Figure 4.9. A **Survival - Group Effects** covariate dialog with effects included in the first interval but not the second interval.

4.4 Joint Survival and Capture Modeling

As mentioned at the beginning of this chapter, one approach to modeling is to model capture and survival (including product) probabilities concurrently. Joint survival and capture models are created by double-clicking **Create Model** under the **Survival Modeling** heading after all the effects dialogs have been set as desired.

SURPH will prompt the user for a unique name for the model. Once created, the model will appear under the **Available Survival Models** heading.

Figure 4.10 shows an example with two joint survival and capture models created. A green checkmark indicates the currently active model. All effects dialogs will always be set to the currently active model. The user can change the currently active model by double-clicking on the desired model. All currently open effects dialogs will be updated to reflect the currently active model. Thus, if the user wants to create a new model that is a slight variation of an existing model, the user must simply make the existing model active, make the necessary changes, and create the new model with a new name.

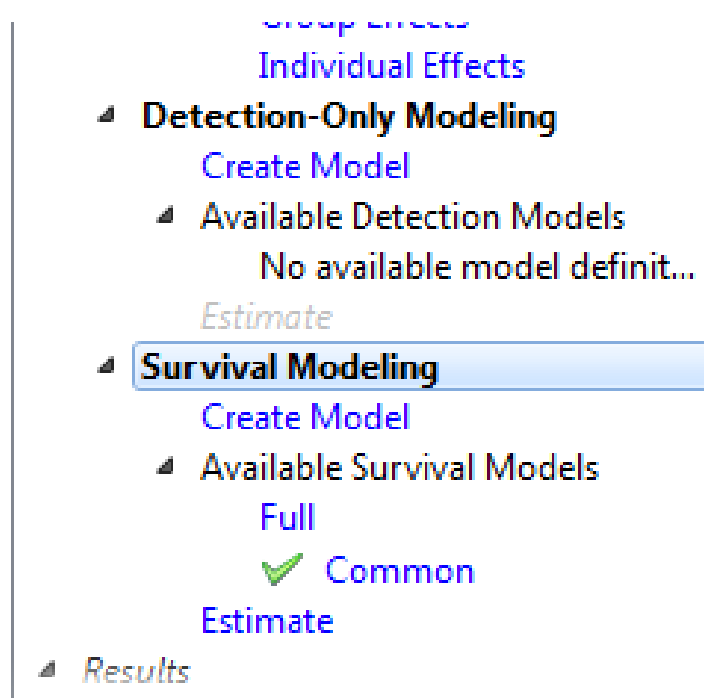


Figure 4.10. **Navigation Panel** with two survival models pending.

Notice in Figure 4.10 that the **Estimate** action is now available under **Survival Modeling**. Whenever there are pending models (models for which results have not yet been calculated), the **Estimate** action becomes available.

Double-clicking on **Estimate** will prompt SURPH to estimate the parameters for all pending models. Doing so for the example in Figure 4.10 will cause the parameters to be estimated for both **Full** and **Common** as shown in Figure 4.11. Notice that the **Estimate** action is no longer available, and the results **burnham/Full** and **burnham/Common** are

now available under the **Results** section. Results names take the form `dataset-name/model-name`.

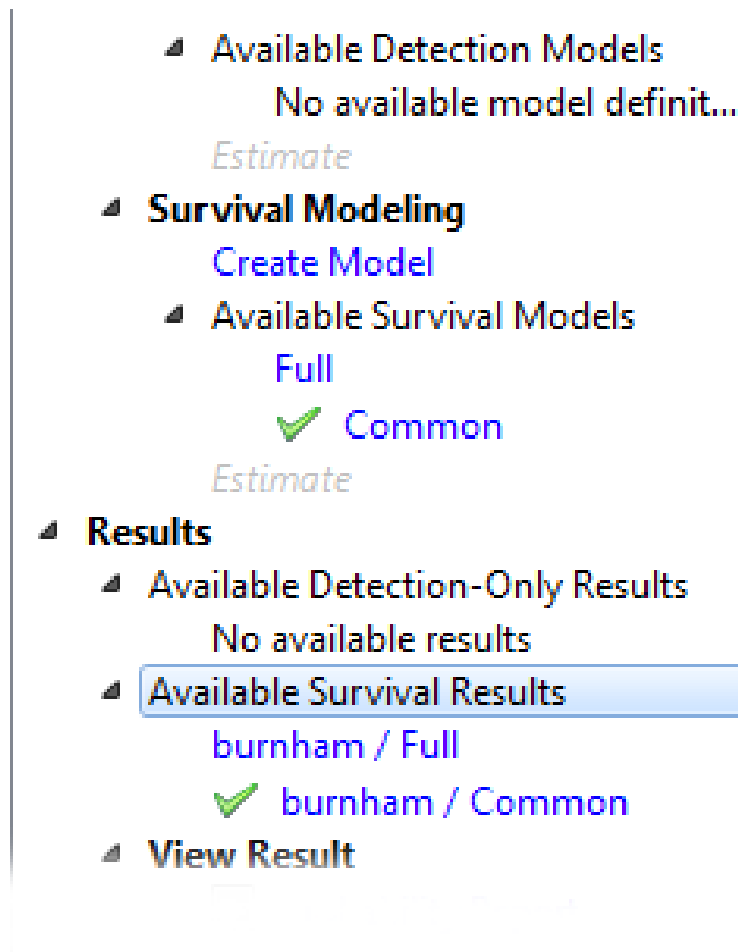


Figure 4.11. **Navigation Panel** after estimating parameters for models `Full` and `Common` for dataset `burnham`.

Note that a model may be renamed at any time by right-clicking on it and selecting **rename** as shown in Figure 4.12.

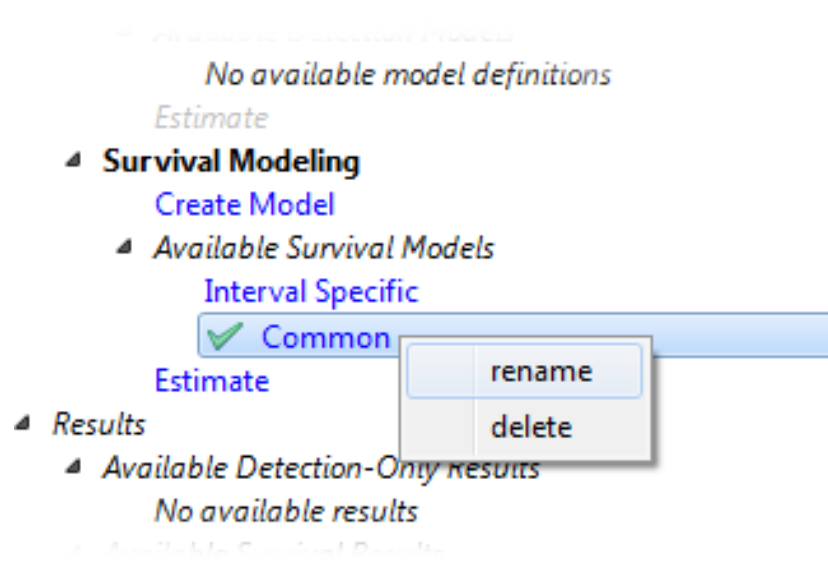


Figure 4.12. Renaming a model.

4.5 Two-stage Modeling

In the two-stage modeling approach, the detection probabilities are modeled first, independent of the survival process. Once the detection model has been selected, the user can **lock** the detection-only model and then model the survival (and product) parameters.

4.5.1 Detection-Only Modeling

When **Create Model** is clicked under **Detection-Only Modeling**, a model is created in much the same way as was described in the previous section, with the name appearing under the **Available Detection Models** heading. The difference, however, is that only the detection parameterization is saved in a detection-only model; the survival and product effects are not saved.

As with joint modeling, the **Estimate** action becomes available once there are pending detection-only models. Double-clicking on **Estimate** will cause the parameters to be estimated under using the Manly-Parr likelihood, described in Appendix B.3. The Manly-Parr likelihood estimates the detection probabilities without reference to the survival process.

4.5.1.1 Locking the Detection-Only result

Once results have been created from the detection-only modeling process, the user can lock the “best” detection-only result (see **Hypothesis Testing** for selecting the best model) by right-clicking on it under **Available Detection-Only Results** and selecting **lock** as shown in Figure 4.13. **Note that a detection result cannot be locked if there are existing results under Available Survival Results in the Results section.**

Once the detection-only result is locked, a padlock icon appears beside it in the **Available Detection Models** and **Available Detection Results** as shown in Figure 4.14. At that point, the user can no longer make changes to the **Detection-Only Modeling** section and can no longer create any further detection-only models or results. The user can look at the **Effects** dialogs for **Detection**, but their settings cannot be modified.

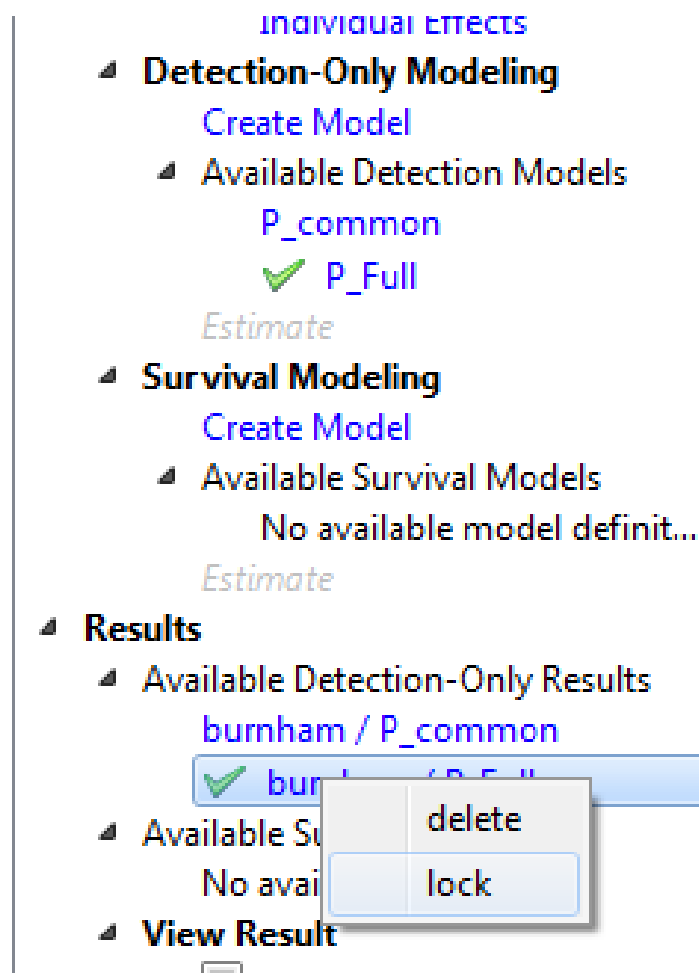


Figure 4.13. Lock the detection result by selecting it and right-clicking on it.

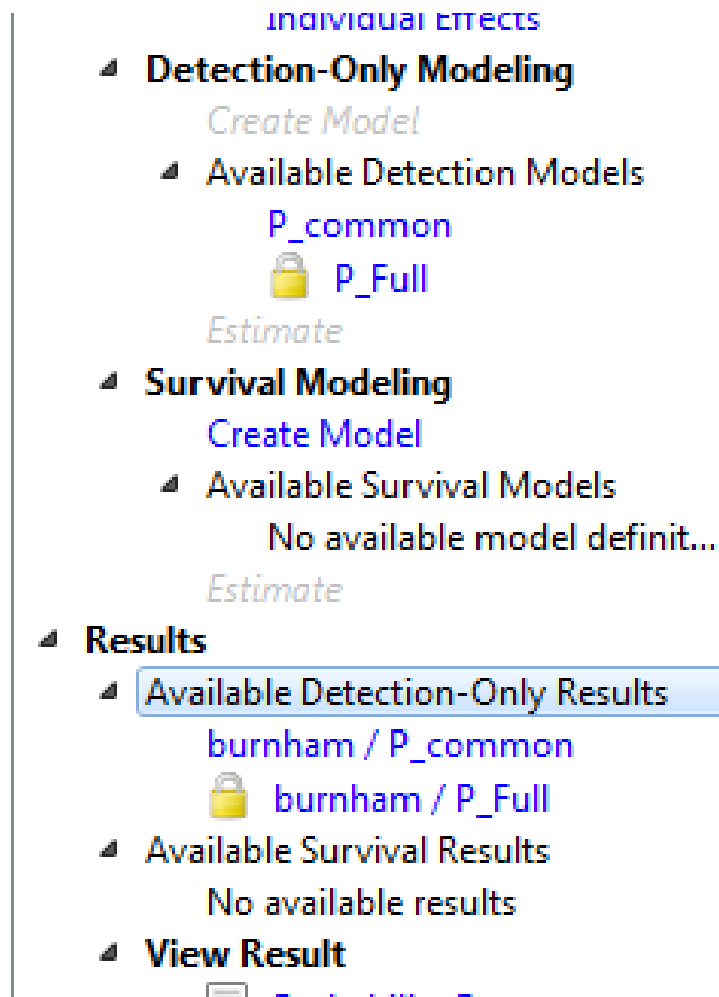


Figure 4.14. The detection result `P_Full` for dataset `burnham` is locked. A lock icon appears next to both the result and the corresponding model.

The detection-only result can be unlocked by right-clicking on it and selecting **unlock** as shown in Figure 4.15.

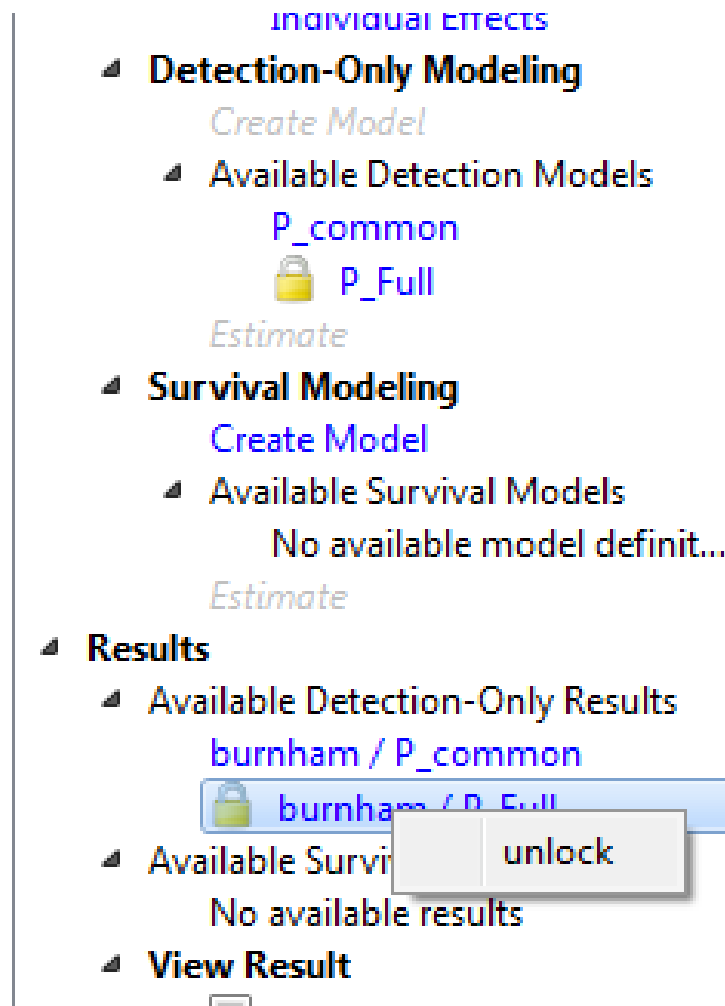


Figure 4.15. Unlocking a previously locked detection-only result.

If there are existing models under **Available Survival Models** that do not conform to the locked detection-only model—that is, their detection parameter settings are different than those of the locked detection-only result—they will be grayed out and will no longer be available in the current dataset.

4.5.2 Survival Modeling

After locking the desired detection-only result, the user can then model the survival and product parameters as described in Section 4.4.

Chapter 5: Results

As mentioned in Section 4.4, when there are pending models—models for which there are no matching results—the user can double-click on **Estimate**, which causes the selected optimizer to be used to estimate the parameter. Figure 5.1 shows an example of output from the Fletch optimizer. The output includes the name of the optimizer, the active model name, the beginning function value, and the ending function value. Once the optimization is complete, the resulting estimates are used as the new seeds and the optimization is repeated. This is repeated until no further improvement in the resulting log-likelihood is found, as defined by the ending criterion in the **Optimizer Setup** dialog (Section 2.2.2.2).

Once the parameters have been estimated successfully, the results appear under **Available Results** in the form `active-dataset/active model`.

Output

Welcome to SURPH 3

Estimating model "interval effect"

Fletch, beginning function value: 24142.7457037

Fletch, ending function value: 23048.8176341, number of iterations: 47

Re-estimating using previous estimates as seeds...

Fletch, beginning function value: 23048.8176341

Fletch, ending function value: 23048.8176341, number of iterations: 2

"interval effect" estimation complete.

Figure 5.1. **Output** from the Fletch optimizer during parameter estimation.

5.1 View Results

SURPH provides two reports under the **View Results** heading of the **Navigation Panel**: the **Probability Report** and the **Parameter Report**. There are also plots that

provide visual representation of a group covariate's or individual covariate's effect on survival (or detection).

5.1.1 Probability Report

The **Probability Report** takes the parameter estimates from the active result and translates them into survival, detection, and final product estimates, along with their standard errors, using the link function defined in the model.

Figure 5.2 shows a **Probability Report** for a study with 10 populations and 3 intervals. The report is structured like the **CJS Estimates** report (Figure 3.10), with a detection and survival estimate for each population and interval. The model that was used for this example assumed unique survival probabilities for each interval and common survival probabilities across populations. The model for the detection probabilities was kept at the default interval/population specific setting.

Probability Report - Model Estimates

Report Date: Wed Feb 1 09:23:26 2012

Data Info

Dataset: distance
Description: Yearling Chinook Releases from Snake River Hatcheries - 1993
Populations: 10
Intervals: 3

Model Info

Model Definition: interval specific
Number of Parameters: 32
Survival Link: Hazard
Detection Link: Logit
Product Link: Hazard
Log-Likelihood: -23048.818
AIC: 46161.635
Optimizer: Fletcher, maxIterations=200, stepSize=1e-08, precision=1e-08
Estimation Ending Criterion: 0.0001

Population	s1	s2	p1	p2	final product
1	0.650 (0.0086)	0.788 (0.0171)	0.445 (0.0052)	0.483 (0.0244)	0.595 (0.0274)
2	0.650 (0.0086)	0.788 (0.0171)	0.518 (0.0179)	0.367 (0.0225)	0.560 (0.0272)
3	0.650 (0.0086)	0.788 (0.0171)	0.417 (0.0177)	0.292 (0.0208)	0.453 (0.0256)
4	0.650 (0.0086)	0.788 (0.0171)	0.449 (0.0225)	0.280 (0.0238)	0.442 (0.0301)
5	0.650 (0.0086)	0.788 (0.0171)	0.487 (0.0208)	0.605 (0.0254)	0.706 (0.0277)
6	0.650 (0.0086)	0.788 (0.0171)	0.442 (0.0107)	0.403 (0.0196)	0.537 (0.0232)
7	0.650 (0.0086)	0.788 (0.0171)	0.366 (0.0179)	0.395 (0.0241)	0.533 (0.0285)
8	0.650 (0.0086)	0.788 (0.0171)	0.372 (0.0134)	0.321 (0.0161)	0.431 (0.0191)
9	0.650 (0.0086)	0.788 (0.0171)	0.503 (0.0130)	0.393 (0.0158)	0.581 (0.0186)
10	0.650 (0.0086)	0.788 (0.0171)	0.185 (0.0183)	0.134 (0.0189)	0.231 (0.0246)

Figure 5.2. Example of a **Probability Report** for a dataset with 10 populations and 3 intervals.

5.1.2 Parameter Report

The **Parameter Report** reports the actual parameters that were estimated prior to being translated into the probability estimates. Figure 5.3 shows a portion of the **Parameter Report** corresponding to the probabilities in Figure 5.2.

Estimation Ending Criterion: 0.0001

Survival parameters

Parameter #	Estimate	Std. Err.	Ratio	Definition
1	0.618	0.0378	16.3621	Baseline
2	-0.592	0.1118	-5.2933	Main effect 1

Detection parameters

Parameter #	Estimate	Std. Err.	Ratio	Definition
3	-0.223	0.0730	-3.0509	Baseline
4	0.154	0.1161	1.3288	Main effect 1
5	0.293	0.0991	2.9573	Main effect 2
6	-0.324	0.1220	-2.6532	Main effect 3
7	-0.113	0.1000	-1.1263	Main effect 4
8	-0.662	0.1249	-5.2973	Main effect 5
9	0.016	0.1071	0.1531	Main effect 6
10	-0.723	0.1390	-5.2010	Main effect 7
11	0.172	0.1079	1.5975	Main effect 8
12	0.648	0.1297	4.9971	Main effect 9
13	-0.010	0.0926	-0.1028	Main effect 10
14	-0.171	0.1108	-1.5404	Main effect 11
15	-0.329	0.1023	-3.2132	Main effect 12
16	-0.202	0.1249	-1.6205	Main effect 13
17	-0.301	0.0882	-3.4079	Main effect 14
18	-0.525	0.1041	-5.0439	Main effect 15
19	0.236	0.0858	2.7528	Main effect 16
20	-0.211	0.0996	-2.1191	Main effect 17
21	-1.258	0.1390	-9.0481	Main effect 18
22	-1.645	0.1787	-9.2039	Main effect 19

Product parameters

Parameter #	Estimate	Std. Err.	Ratio	Definition
-------------	----------	-----------	-------	------------

Figure 5.3. A portion of a **Parameter Report** for a dataset with 10 populations and 3 intervals, corresponding to Figure 5.2.

The model parameterization corresponding to the results shown in Figure 5.3 can be found in Appendix B.2.1.4, Table B.7. Table 5.1 explains the relationship of the survival parameters in Figure 5.3 to the model in Appendix B.2.1.4.

Table 5.1. Relationship between the survival parameters from the **Parameter Report** shown in Figure 5.3 and the model in Appendix B.2.1.4.

Definition	Description	Parameter
Baseline	Baseline survival	S_0
Main effect 1	Main effect for interval 2	π_2

Similarly, Table 5.2 explains how the Detection Parameters reported in the **Parameter Report** correspond to the fully parameterized model of Appendix B.2.1.1.

Table 5.2. Relationship between the detection parameters from the **Parameter Report** shown in Figure 5.3 and the model in Appendix B.2.1.1.

Definition	Description	Parameter
Baseline	Baseline survival	P_0
Main effect 1	Main effect for population 1, interval 2	π_{12}
Main effect 2	Main effect for population 2, interval 1	π_{21}
Main effect 3	Main effect for population 2, interval 2	π_{22}
⋮	⋮	⋮
Main effect 19	Main effect for population 10, interval 2	$\pi_{10,2}$

5.2 Survival Plots

5.2.1 Individual Covariate Survival

Figure 5.4 shows an individual covariate survival plot for the individual covariate

length. The smooth red curve is the modeled survival as a function of the selected covariate value. The jagged black line is a nonparametric estimate of survival as a function of the covariate. Figure 5.4 also shows a histogram of the individual covariate data superimposed on the plot.

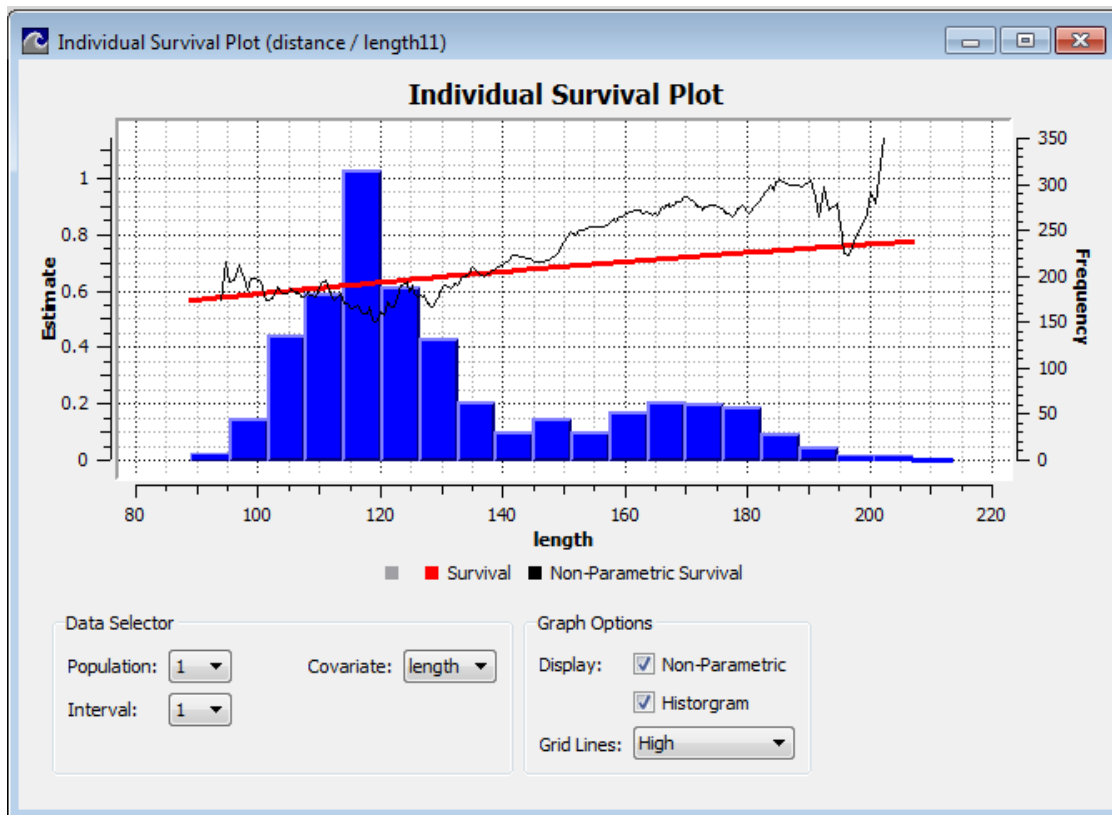


Figure 5.4. **Individual Survival Plot** with one individual covariate **length**.

The X-axis for the **Individual Survival Plot** is the covariate value. The Y-Axis on the left is the survival estimate, pertaining to both the modeled survival function and the nonparametric survival function. The Y-axis on the right side is the frequency, referring to the histogram.

Note that the controls at the bottom of the **Individual Survival Plot** allow the user to select the population and interval of interest. If there is more than one individual covariate in the data, the user can select the individual covariate of interest. The user can also selectively display or hide the nonparametric survival function or the histogram.

Because the nonparametric survival function does not depend on any model assumptions, the degree to which the modeled survival function “fits” the nonparametric function can be used to measure the degree of how well the chosen model fits the data.

For details of how both the parametric survival function and the nonparametric function are calculated, see Appendix B.4.1.

5.2.2 Group Survival Plot

Figure 5.5 shows a group covariate survival plot for group covariate **distance**. As with the **Individual Survival Plot**, the red line shows the modeled survival estimate as a function of the group covariate. The points (indicated by +) are the CJS estimates for the selected interval.

In the example in Figure 5.5, there are ten populations; hence, there are ten CJS estimates plotted—one for each population in interval 1. As with the nonparametric survival function on the **Individual Survival Plot**, the CJS estimates do not rely on the covariate assumptions and can be used to judge how well the selected model fits the data.

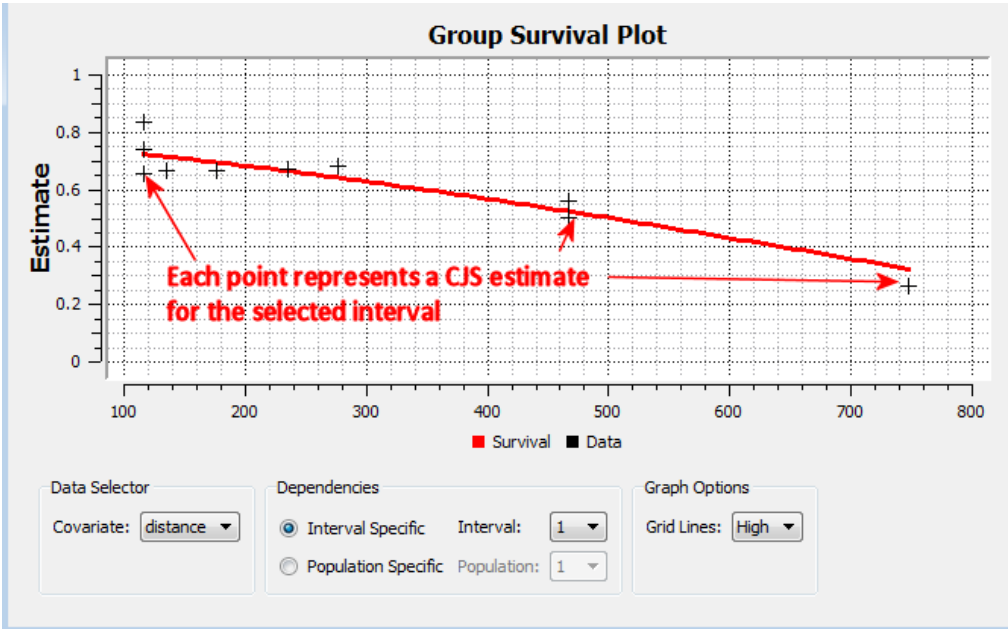


Figure 5.5. Interval-specific **Group Survival Plot** for group covariate **distance**.

Note that the plot in Figure 5.5 uses the default setting of **Interval Specific**. If the user selects **Population Specific** in the controls at the bottom of the plot, the population-specific survival function will be plotted and there will be one CJS estimate for each interval for the selected population.

For details on calculating both the interval-specific and population-specific group covariate parametric survival function, see Appendix B.4.2.

5.3 Diagnostic Plots

SURPH provides several diagnostic plots for assessing how well the current model fits the data. These are (1) **Model Estimates vs. CJS Estimates**, (2) **Residuals vs. Capture Histories**, and (3) **Residual - Normal QQ**. These plots can be found under **View Result Diagnostic Plots** in the **Results** section of the **Navigation Panel**.

5.3.1 *Model Estimates vs. CJS Estimates*

Figure 5.6 shows an example of the **Model Estimates vs. CJS Estimates** plot that is generated after clicking on **Modeled Estimates vs. CJS Estimates** on the **Navigation Panel**. **CJS Estimates** are on the X-axis and **Model Estimates** on the Y-axis for the currently selected results. The **CJS Estimates** are the maximum likelihood estimates of the survival and capture probabilities and thus are the “best” parameter estimates.

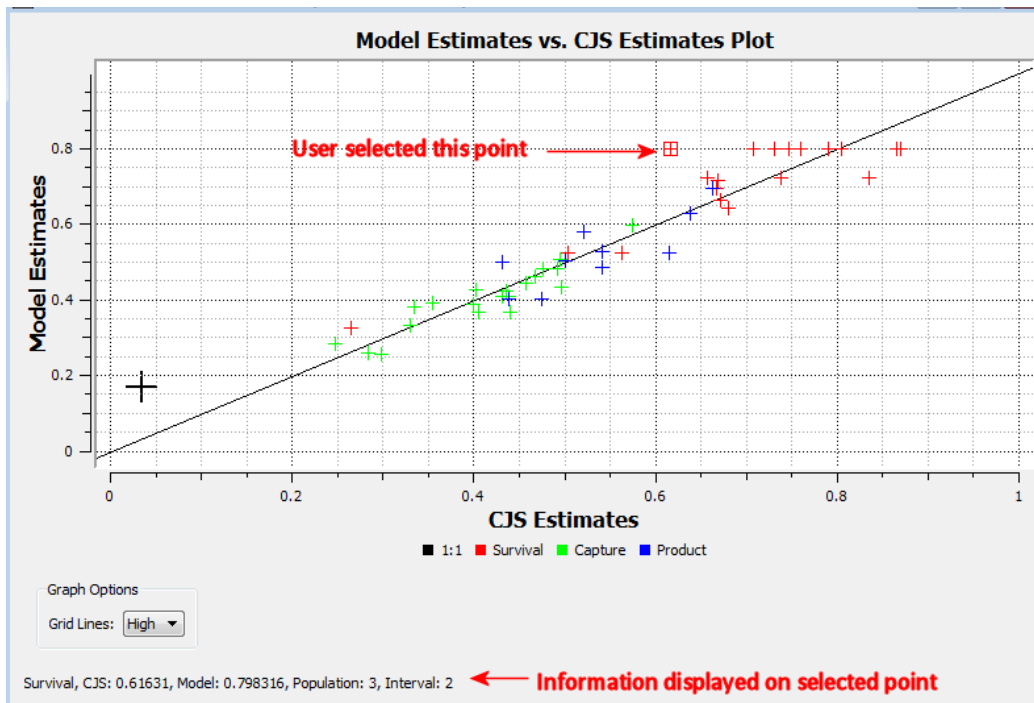


Figure 5.6. **Model Estimates vs. CJS Estimates** plot.

If the currently selected model adequately fits the data, then the resulting estimates should be close to the CJS estimates and the points should be on or close to the identity line drawn on the plot. Large deviations from the identity line may indicate that the model does not adequately fit the data.

As shown in Figure 5.6, the user may click on any point, and information about that point will appear on the lower left of the plot.

5.3.2 Residuals vs. Capture Histories

An example of the **Residuals vs. Capture Histories** plot is shown in Figure 5.7. The X-axis is simply an index of the capture histories. The Y-axis is the corresponding Anscombe Residuals (McCullagh and Nelder 1989:37–38).

The Anscombe Residuals refer to transformations of the residuals that, if the model fits the data, follow a standard normal distribution, meaning that the mean of the residuals will be zero and approximately 95% of the residuals will fall in the interval between -2 and 2.

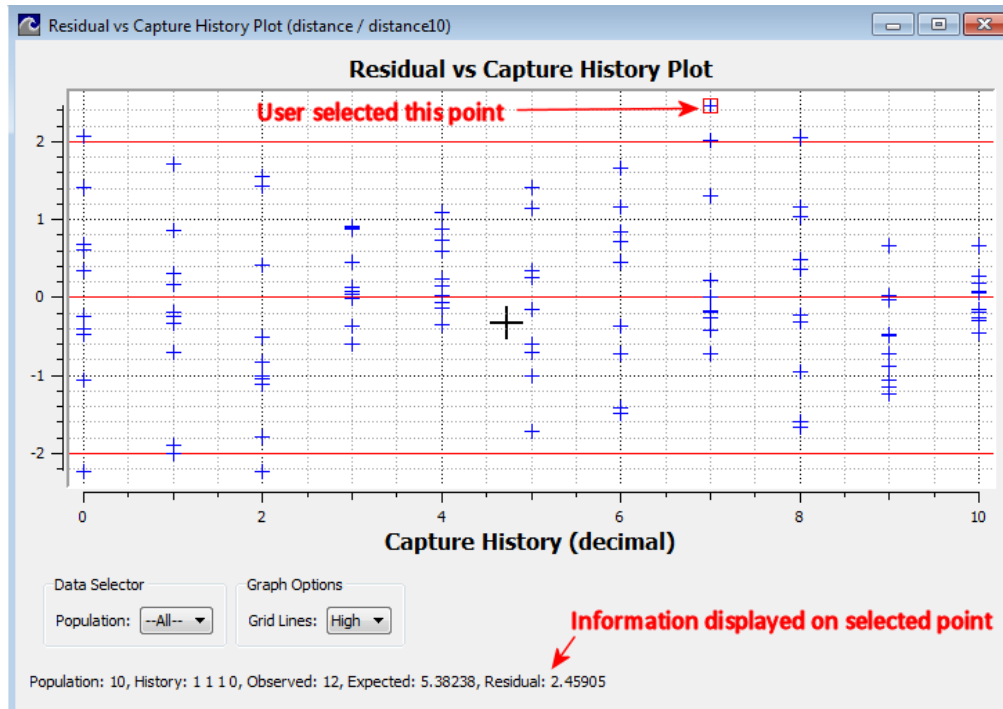


Figure 5.7. **Residuals vs. Capture Histories** plot.

As shown in Figure 5.7, red horizontal lines are drawn at -2 and 2, and about 95% of the residuals should fall between the lines. As with the **Model Estimates vs. CJS Estimates** plot, the user can select a point to get information about the point displayed in the lower left.

Figure 5.8 shows a contrasting example of the **Residuals vs. Capture Histories** plot. Notice that there are many residuals outside the -2 to 2 interval and residuals with values outside of the -4 to 4 interval. The probability of observing values this extreme with a standard normal distribution is about 1 in 30,000—indicating that the selected model in Figure 5.8 may not be adequate.

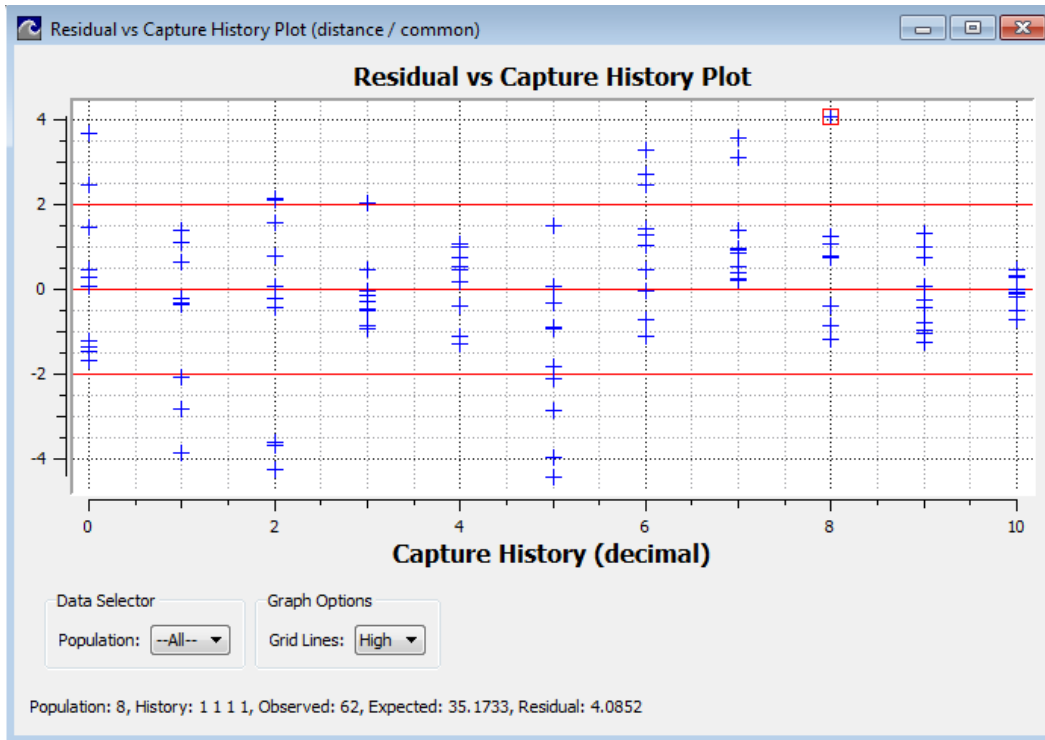


Figure 5.8. **Residuals vs. Capture Histories** plot for a model that does not adequately fit the data.

5.3.3 Residual - Normal QQ Plot

Clicking on **Normal QQ** under Results brings up the **Residual - Normal QQ** plot. This plot is another tool to assess whether or not the residuals from the model are normally distributed, indicating an adequate fit of the model to the data. An example is shown in Figure 5.9.

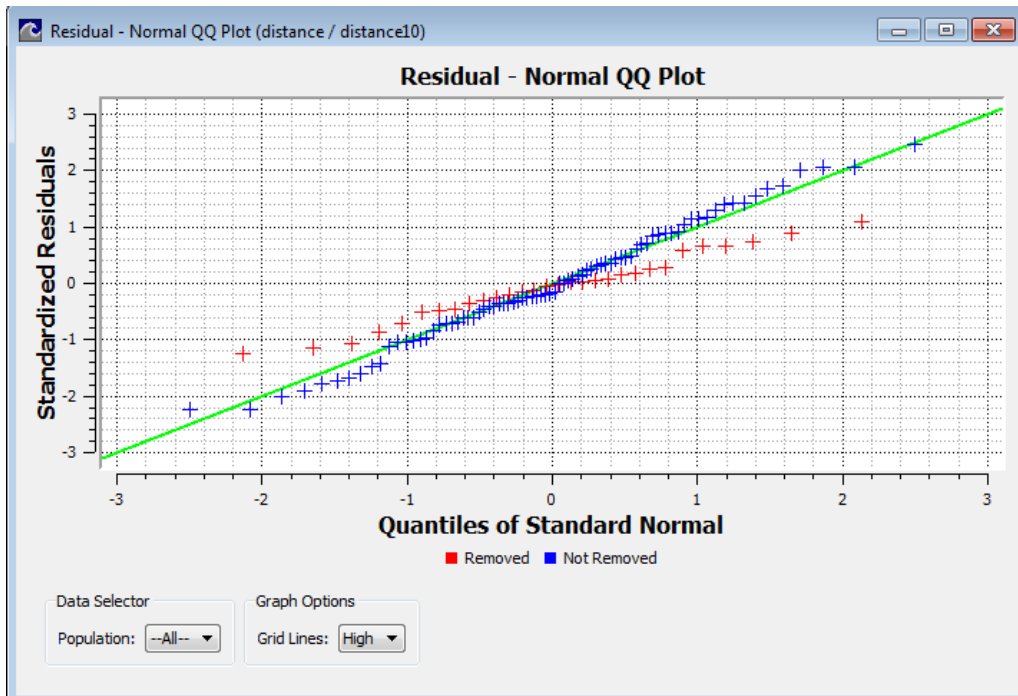


Figure 5.9. **Residual - Normal QQ Plot.**

The X-axis is the quantiles of a standard normal distribution and the Y-axis is the standardized residuals. If the residuals are normally distributed, the points should fall on or near the identity line drawn on the plot. The **Residual - Normal QQ** plot shows the capture histories for individuals removed (censored) in red and those not removed in blue.

Details of the **Residual - Normal QQ** plot can be found in Sokal and Rohlf (1995:88–93).

5.4 Advanced Export Functions

Under the **Advanced Export Functions** heading on the **Navigation Panel**, SURPH provides the capability of exporting both the covariance matrix and the correlation matrix for the current model. The rows and columns are numbered from 0 to $N-1$ where N is the number of parameters in the current model. The parameter order matches the **Parameter Report** in Section 5.1.2.

5.5 Literature Cited

McCullagh, P., and J. A. Nelder. 1989. *Generalized linear models*. London: Chapman & Hall.

Sokal, R. R., and F. J. Rohlf. 1995. *Biometry, the principles and practice of statistics in biological research*. New York: W. H. Freeman.

Chapter 6: Hypothesis Testing

Testing

SURPH provides two methods of hypothesis testing: (1) the Likelihood Ratio Test and (2) Analysis of Deviance. The Likelihood Ratio Test is a standard method to test one model against another. Studies by Smith (1991), however, show that for models with group covariate effects, Analysis of Deviance is the preferred method (see explanation in Appendix B.5.1).

For nonhierarchical models, the relative Akaike Information Criteria (AIC) is provided to compare models. The relative AIC (ΔAIC) is given both in the **Master Model List** of the **Likelihood Ratio Test** dialog and the models list of the **Analysis of Deviance** (also referred to in program as **ANODEV**) dialog. The model with the smallest AIC (indicating the “best” model) is assigned the value of 0.0, with the relative AIC for all other models being the difference from the “best” model.

6.1 Likelihood Ratio Test

Figure 6.1 shows the **Likelihood Ratio Test** dialog that appears when the user double-clicks on **LRT and AIC** on the **Navigation Panel**. The top portion is the **Master Model List** and shows all current models for which there are results.

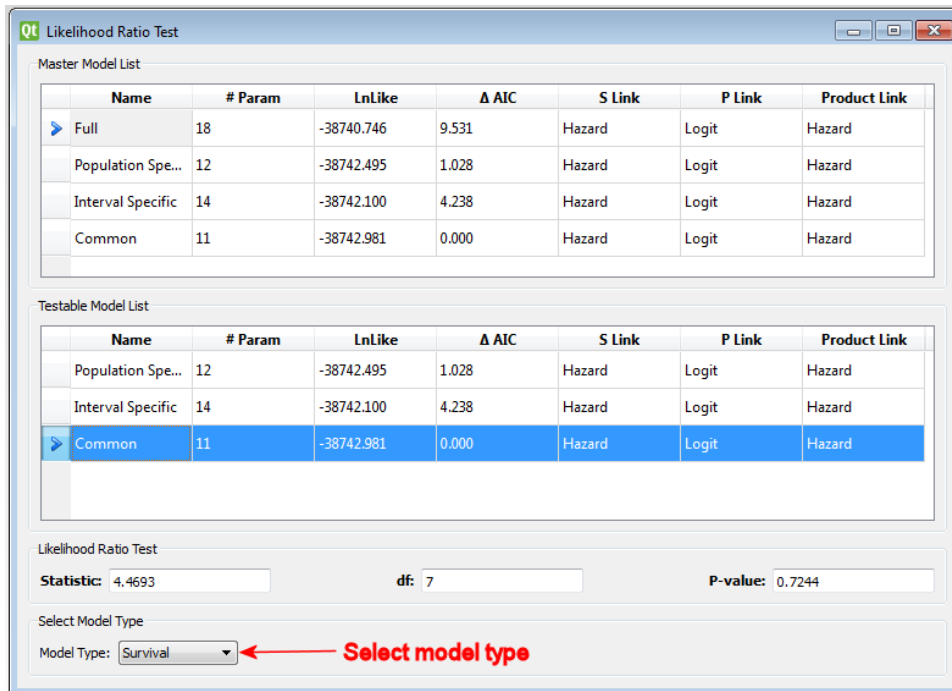


Figure 6.1. Example of a **Likelihood Ratio Test** dialog.

At the bottom, the user must select either **Survival** or **Detection Only** for the model type. In Figure 6.1, the results were created using joint detection and survival modeling; thus, we select the **Survival** model type at the bottom of the dialog.

The user selects a model of interest on the **Master Model List** by clicking on it. A right-pointing blue arrow immediately appears to the left of it, indicating that it is the currently selected model. All testable models appear below in the **Testable Model List**.

A testable model is one that is “nested” in the master model. A nested model is one that can be derived from the master model by equating one or more parameters or setting one or more parameters to zero.

In the example in Figure 6.1, the user has selected the model named **Full** in the master model list. The models that are nested in **Full** then appear in the **Testable Model List**. In this case, the user has selected **Common**. The Likelihood Ratio Test statistic comparing model **Full** with model **Common** is 4.4693, with 7 degrees of freedom and a *P*-value of 0.7244.

6.2 Analysis of Deviance

The ANODEV method of hypothesis testing is an alternative to the Likelihood Ratio Test in SURPH. This section describes how to perform an ANODEV in SURPH. Appendix B.5 explains the calculations behind the ANODEV.

The ANODEV method is preferred for hypothesis testing when testing the significance of group covariate effects on survival, as explained in Appendix B.5.1. The **ANODEV** action under **Hypothesis Testing** on the **Navigation Panel** is available only if the current dataset has at least one group covariate and there is more than one population.

Figure 6.2 shows the **ANODEV Setup** dialog that appears when you first initiate the Analysis of Deviance. As explained in Appendix B.5, ANODEV requires three base models at the start: (1) the **Full** model, (2) the **Null** model, and (3) the **Full Covariate** model. The **ANODEV Setup** dialog allows the user to use the default base models, or to select already existing results.

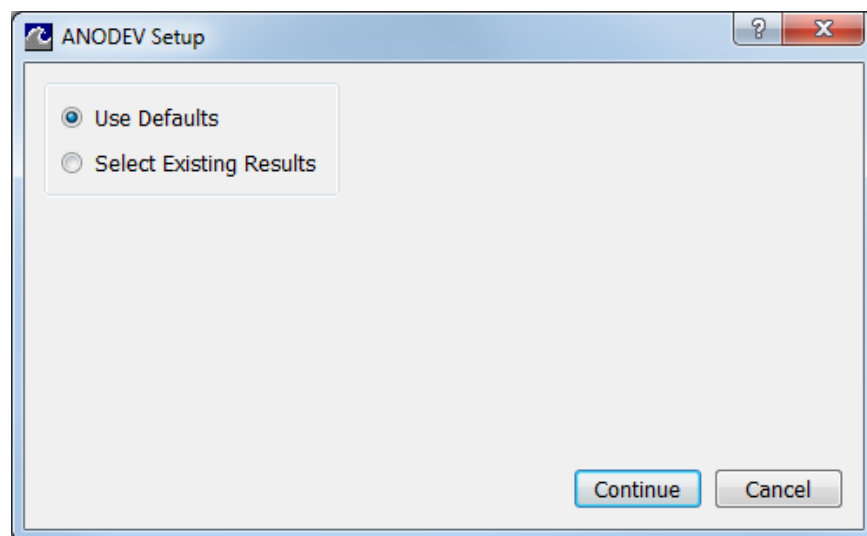


Figure 6.2. The **ANODEV Setup** dialog.

6.2.1 Using the Defaults

When the default option is chosen in the **ANODEV setup** dialog, SURPH creates

the following models and corresponding results in the **Analysis of Deviance** dialog window if they do not already exist.

1. The **Full** model and result are given the name **Full**. The survival main effects are set to population/interval-specific, and all group covariate effects are set to zero.
2. The **Grand Mean** model and result, given the name **Grand Mean**, takes on the role of the **Null** model. The survival main effects are set to interval specific, and the group covariate effects for all group covariates are set to zero.
3. The **Full Covariate** model and result are given the name **Full Covariate**. The survival main effects are set to interval specific, as are the group covariate effects for all group covariates.

The underline (“ ”) in front of the names allows SURPH to identify the models as ANODEV models. Models with names starting with the underline cannot be created any other way by the user.

The detection probabilities and product probabilities in the default ANODEV models will be modeled as defined by the currently selected model upon initiating Analysis of Deviance. If the user wants to model the detection and product probabilities, it needs to be done prior to generating the results in the **Analysis of Deviance** window.

6.2.2 Select Existing Results

If the user selects the **Select Existing Results** option, as shown in Figure 6.3, the user must select the existing results to fill the role of the **Null** model and the **Full Covariate** model. The **Full** model, however, is not selectable by the user; instead, the model **Full** will be created if it does not already exist, as described in the previous section.

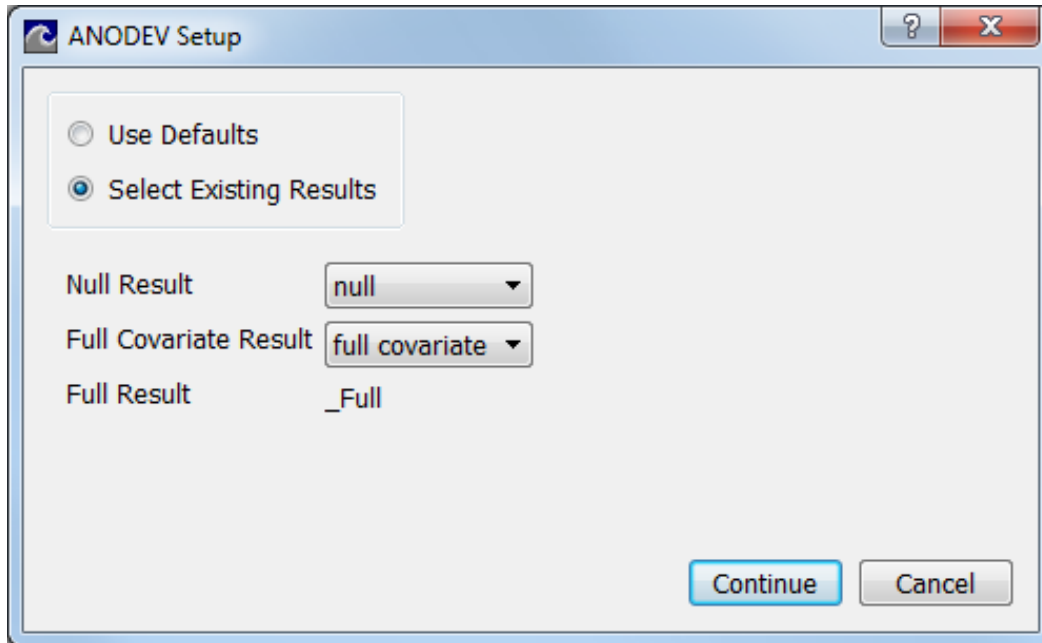


Figure 6.3. The **ANODEV Setup** display for selecting existing results.

The models selected by the user must meet the nesting requirements as described in Appendix B.5

6.2.3 Analysis of Deviance Dialog

Once the models and results have been created, the **Analysis of Deviance** dialog appears. An example is shown in Figure 6.4 The **Role** column, which is the first column of the **Models** table at the top, labels the role of the three required ANODEV results. When using the defaults, the **_Full** result takes on the role of the **Full** result, the **_Grand Mean** result takes on the role of the **Null** result, and the **_Full Covariate** result takes on the **Full Covariate** role.

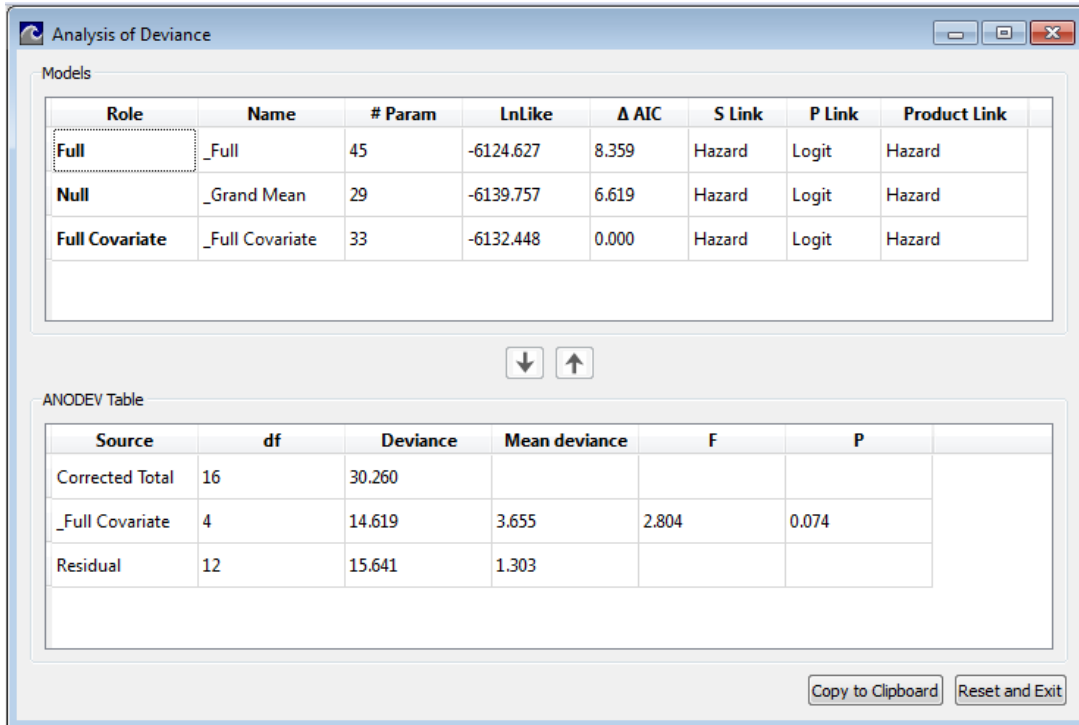


Figure 6.4. The **Analysis of Deviance** dialog with minimal ANODEV table.

The current ANODEV table is shown at the bottom of Figure 6.4. The **Copy to Clipboard** button allows the user to copy the current table and paste it into another application such as a spreadsheet or a report.

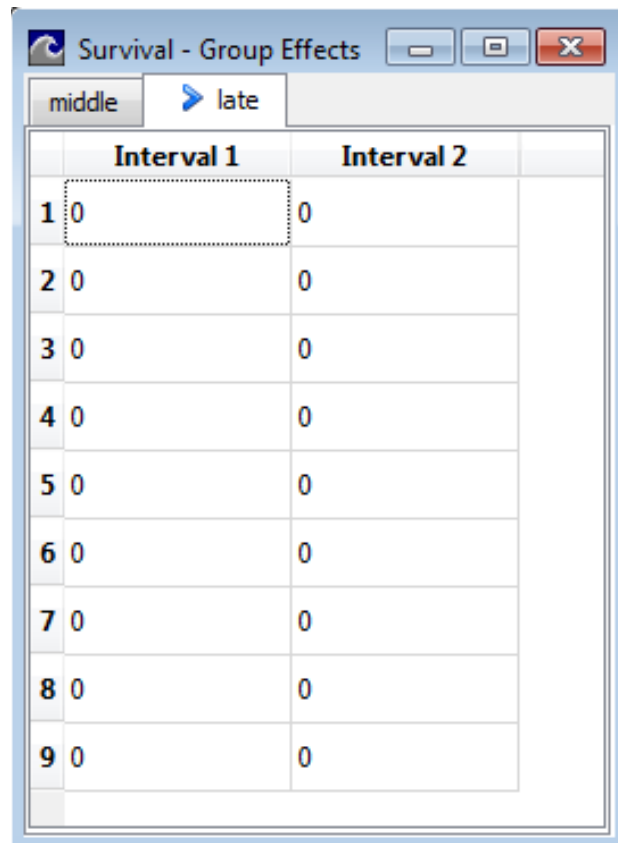
The **Reset and Exit** button will close the **Analysis of Deviance** dialog, and when the **ANODEV** action is selected again, the **ANODEV setup** dialog (Figure 6.2) will reappear asking the user to select either the default models and results or user defined models and results. If the **Analysis of Deviance** dialog is simply closed, selecting the **ANODEV** action will cause the existing **Analysis of Deviance Dialog** (Figure 6.4) to reappear.

6.2.3 Adding and Removing Models from the ANODEV Table

In the example in Figure 6.4, there are two group covariates in the dataset (**middle** and **late**). In the default **Full Covariate** model, both covariates are included. In order to test the effect of only one of the covariates, the user must create the appropriate models and results outside of the **Analysis of Deviance** dialog. The user can minimize or close

the **Analysis of Deviance** dialog and create the desired models and results.

To create a model that looks at only the effect of group covariate **middle**, make sure that the **_Full Covariate** model is active and then set the **late** group covariate effects to the zero setting, as shown in Figure 6.5. Then select **Create Model** on the **Navigation Panel** and call the model **Full Covariate Middle**.



	Interval 1	Interval 2
1	0	0
2	0	0
3	0	0
4	0	0
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0

Figure 6.5 Survival group covariate effects for **late** for modeling only the effects of covariate **middle**.

Similarly, to look at the effects of the covariate **middle** in the first interval only, make **Full Covariate Middle** the active model and set the **middle** group covariate effects as shown in Figure 6.6. Now double-click on **Estimate** to estimate the results for the two new models.

	Interval 1	Interval 2
early1	1	0
early2	1	0
early3	1	0
middle1	1	0
middle2	1	0
middle3	1	0
late1	1	0
late2	1	0
late3	1	0

Figure 6.6. Survival group covariate effects for `middle` for modeling the effects of the covariate in the first interval only.

Now reopen the **Analysis of Deviance** dialog. It should appear as shown in Figure 6.7 with the two new results shown in the **Models** table at the top.

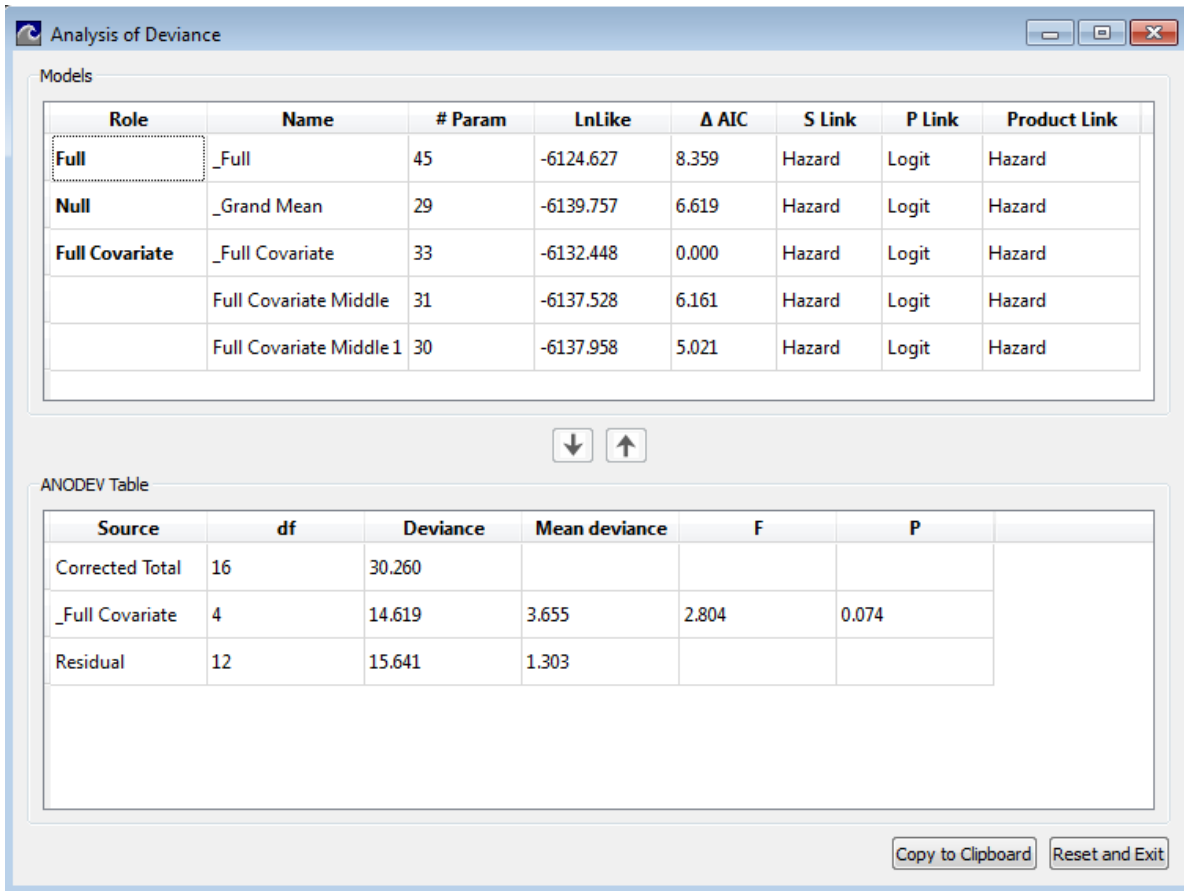


Figure 6.7. **Analysis of Deviance** dialog with user-defined results included.

In order to add the **Full Middle Covariate** result to the ANODEV table, select it in the **Models** table and click the down arrow as shown in Figure 6.8. Note that if the selected result does not meet the nesting requirements as described in Appendix B.5, the result cannot be added to the ANODEV table.

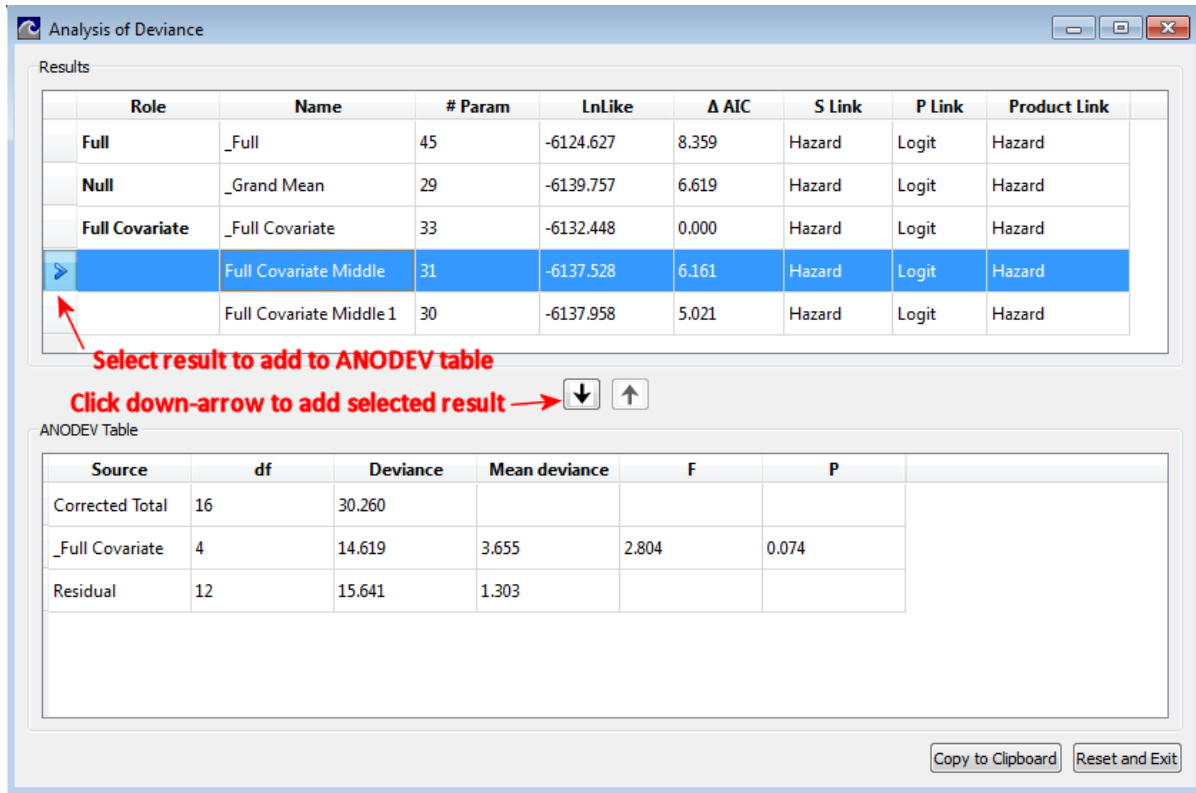


Figure 6.8. **Analysis of Deviance** dialog before adding the **Full Covariate Middle** result.

Once the new model is added, the **Analysis of Deviance** dialog will appear as shown in Figure 6.9, with the new result now included in the ANODEV table and no longer showing in the Results table.

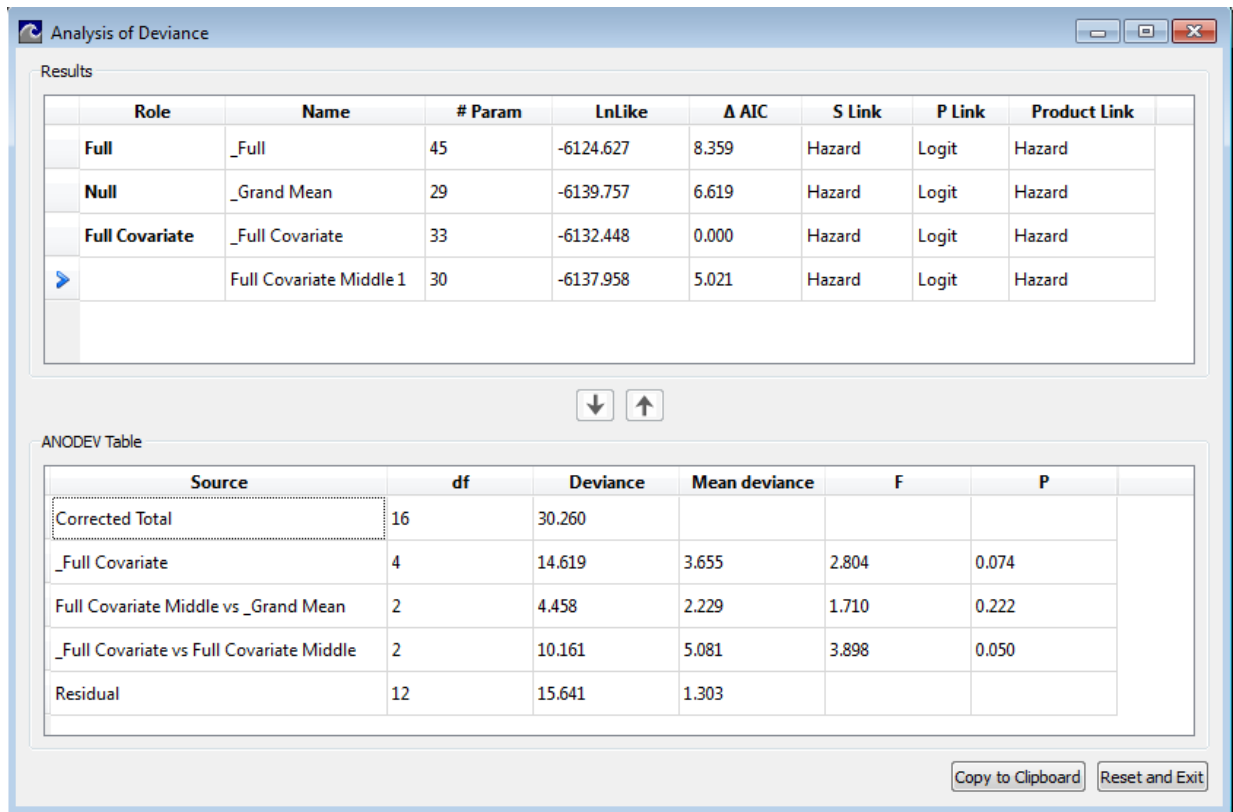


Figure 6.9. **Analysis of Deviance** dialog with **Full Covariate Middle** result added to the ANODEV table.

The model **Full Covariate Middle 1** can be added to the ANODEV table in the same way, as shown in Figure 6.10.

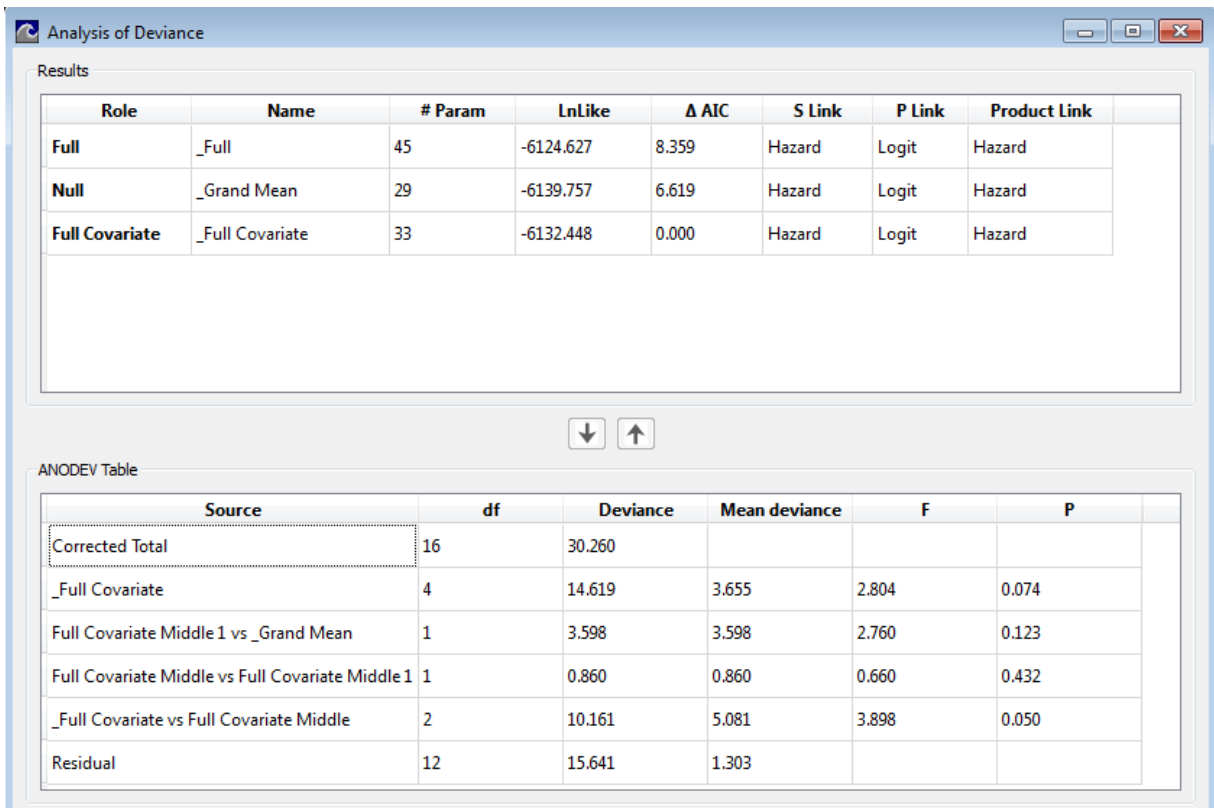


Figure 6.10. **Analysis of Deviance** dialog with **Full Covariate Middle** and **Full Covariate Middle 1** results added to the ANODEV table.

In order to remove a result from the ANODEV table, select it in the ANODEV table and press the up arrow as shown in Figure 6.11. The removed result will now reappear in the Results table. The three required results cannot be removed from the ANODEV table.

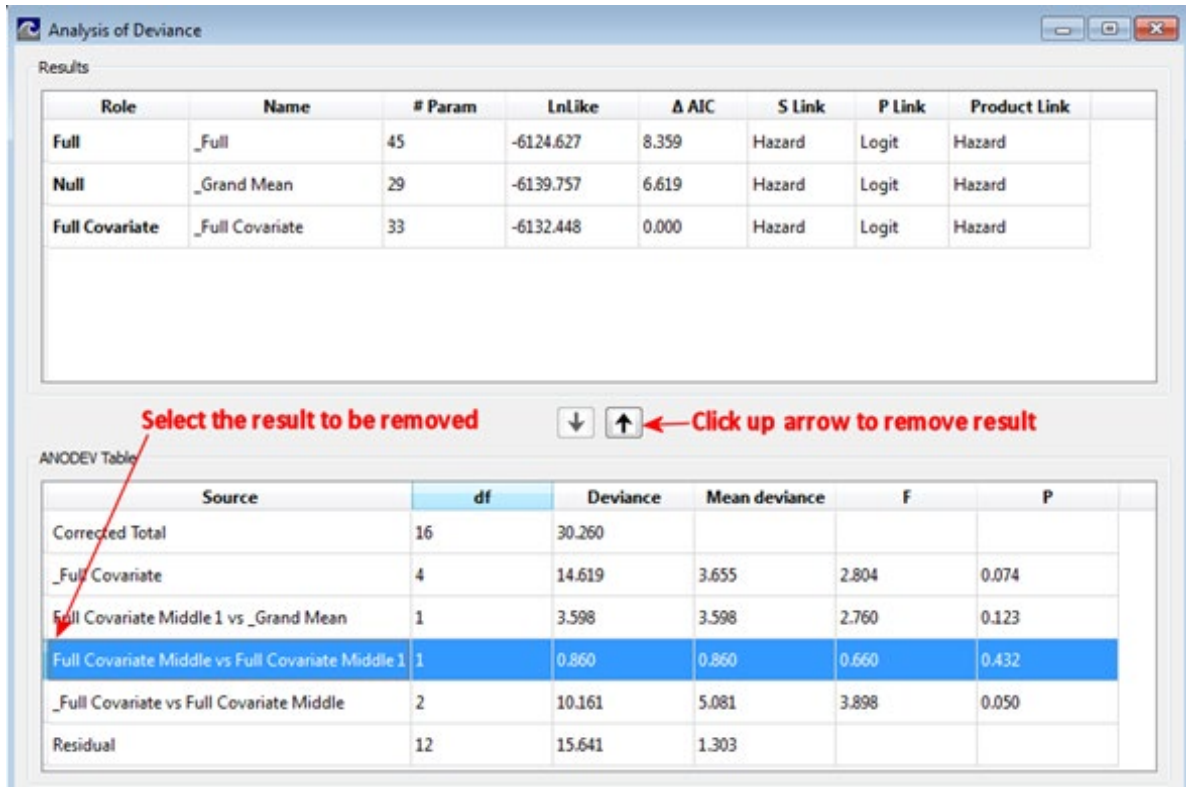


Figure 6.11. **Analysis of Deviance** dialog before removing a model from the ANODEV table.

6.3 Literature Cited

Smith, S. G. 1991. Assessing hazards in wild populations using auxiliary variables in tag-release models (Dissertation). University of Washington, Seattle, Washington.

Chapter 7: Models Across Multiple Datasets

Program SURPH distinguishes between a **Model**, which defines the parameterization for estimating the parameters, and a **Result**, which contains the actual parameter estimates from applying the data to the **Model**. This allows **Models** to be used across compatible datasets. Two datasets are defined as compatible if they have:

1. the same number of intervals,
2. the same number of populations,
3. the same number of group covariates, and the group covariates have the same name, and
4. the same number of individual covariates, and the individual covariates have the same name.

The best way to illustrate how models can be used across compatible datasets is by example. Figure 7.1 shows the top portion of an input data file for year 1 of a hypothetical study. Figure 7.2 shows the data for year 2 of the same study. Because they both have three populations and two intervals, the same three individual covariates, and no group covariates, they are compatible datasets.

```
Surph3
Data from year 1

npop 3
nint 2

icov length
icov weight
icov condix

population one
1 1 1 100 10.8 0.0108
1 1 0 116 19.4 0.0124287588667022
1 1 1 97 10.1 0.01106639508345267
1 1 0 115 18.8 0.01236130516972138
1 1 0 98 11.4 0.01211230014704757
1 0 1 114 16.4 0.01106953286571307
```

Figure 7.1. Top portion of an input data file for year one of a hypothetical study.

```
Surph3
Data from year 2

npop 3
nint 2

icov length
icov weight
icov condix

population one
1 1 0 106 13.2 0.01108297453602638
1 1 0 104 12.3 0.01093465521165225
1 1 0 95 10 0.01166350779997084
1 0 1 115 16.7 0.0109805210816142
1 0 1 103 13 0.01189684157159107
1 0 1 105 13.5 0.01166180758017493
```

Figure 7.2. Top portion of an input data file for year two of a hypothetical study

Figure 7.3 shows the **Navigation Panel** with year 1 data loaded. In this case, the user has used two-stage modeling, locking the detection-only model **P_Full** and subsequently creating two survival models based on it, with corresponding results.

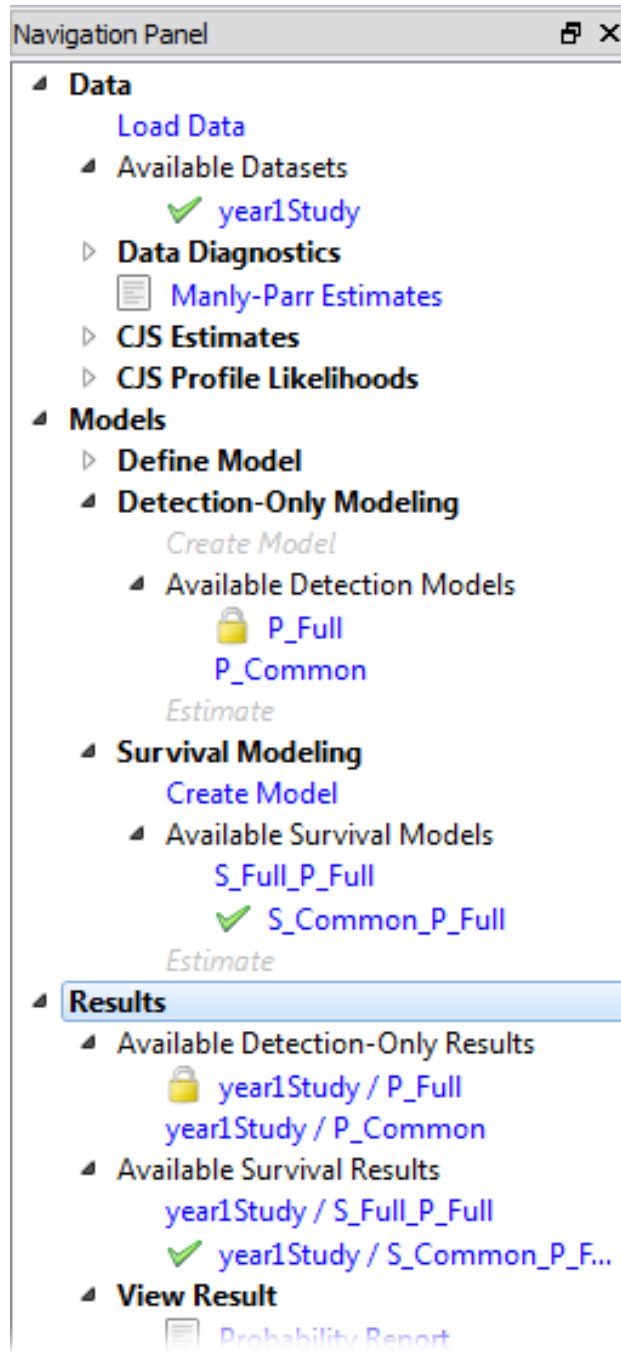


Figure 7.3. **Navigation Panel** after loading year 1 data and creating models and results.

Next, the user loads the data for year two. Figure 7.4 shows the **Navigation Panel** after doing so. Because the two datasets are compatible, the models that were created for the year one data are still available and do not need to be recreated.

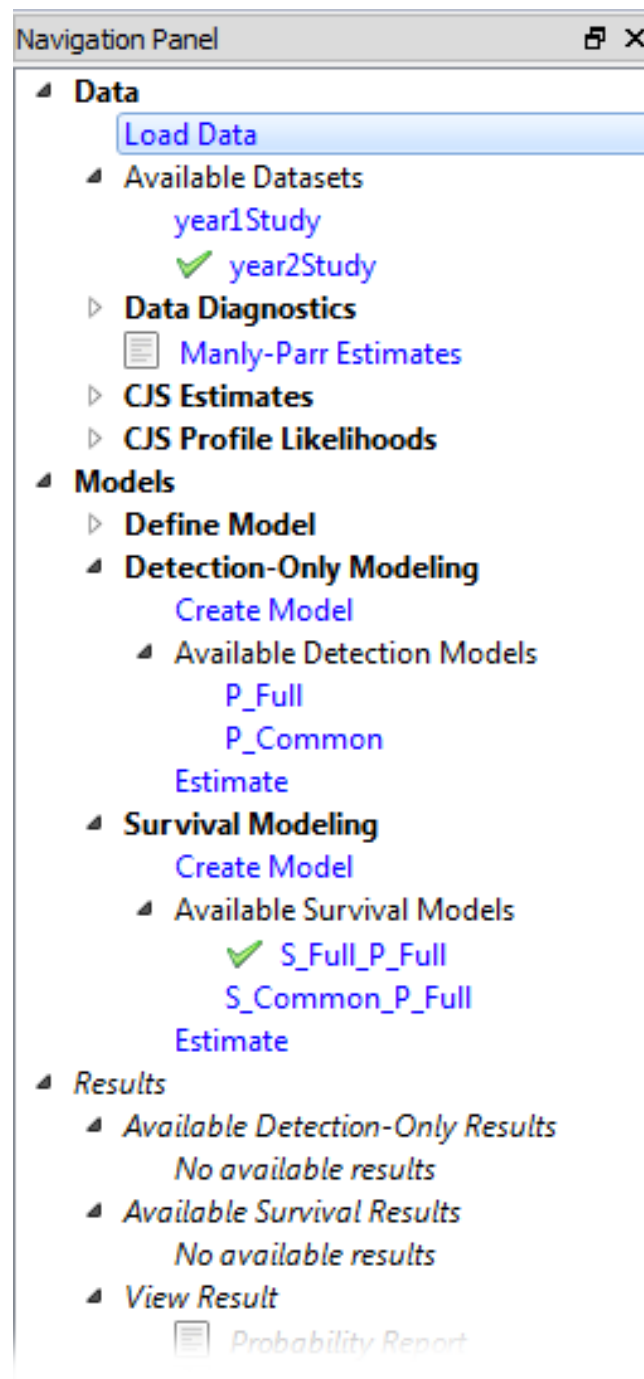


Figure 7.4. **Navigation Panel** after loading year 2 data.

Following the two-stage modeling approach that was used in year 1, the user double-clicks on **Estimate** under **Detection-Only Modeling** to estimate the parameters for all the detection-only models. This time, however, the user finds that the **P_Common** model best fits these data; therefore, the user locked the **year2Study/P_Common** result. The resulting **Navigation Panel** is shown in Figure 7.5.

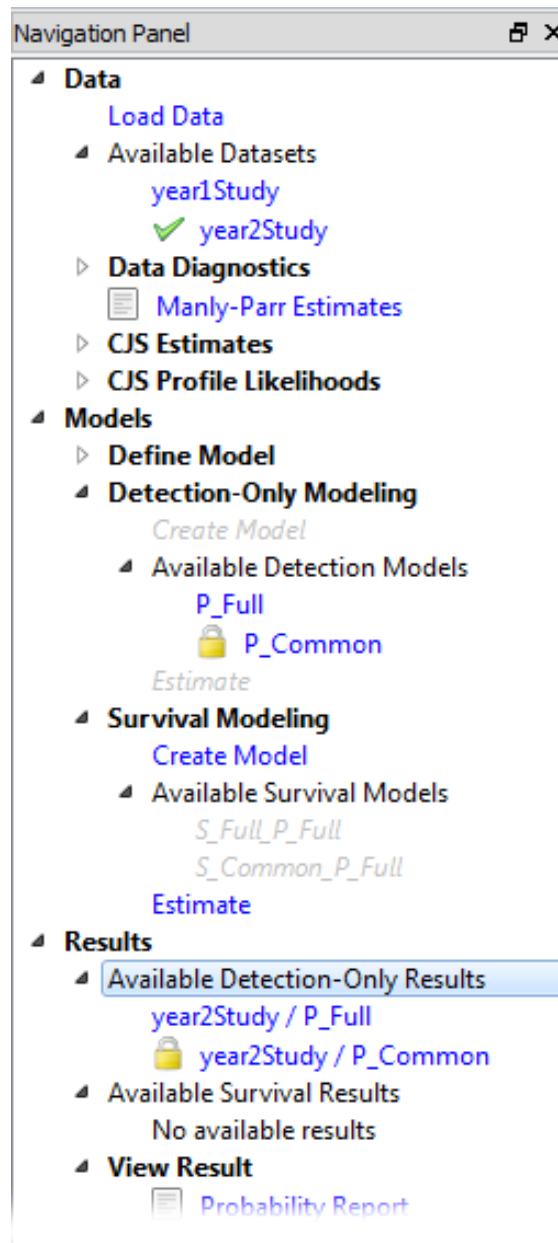


Figure 7.5. **Navigation Panel** after locking the **year2Study/P_Common** result for year 2 data.

Notice that the survival models `S_Full_P_Full` and `S_Common_P_Full` are grayed out. Because they are based on the detection-only model `P_Full`, and the user has locked `P_Common` for the dataset `year2Study`, they are not available to the user for the year 2 dataset. The user can now create new survival models based on the `P_Common` detection model.

Chapter 8: Analysis of Examples

In this chapter, we analyze sample datasets, all from PIT-tag studies of salmonid smolts in the Columbia River system, to illustrate how to use SURPH to analyze release-recapture data. The sample data files for this chapter can be found on [the Columbia Basin Research SURPH page](#) after installation of SURPH 3.5.

8.1 Paired Release Example

In this example, we look at joint detection and survival modeling and hypothesis testing using the Likelihood Ratio Test. The data for this example comes from the American Fisheries Society monograph by Burnham et al. (1987) and is a simulated paired release of PIT-tagged salmon smolts in the Columbia River system. The data is in the file “burnham.dat.” The treatment population is assumed to be released directly into the turbine of a hydroelectric project. The control population is released directly into the tailrace below the dam. The fish are then detected at four downstream hydroelectric projects, as shown in Figure 8.1.

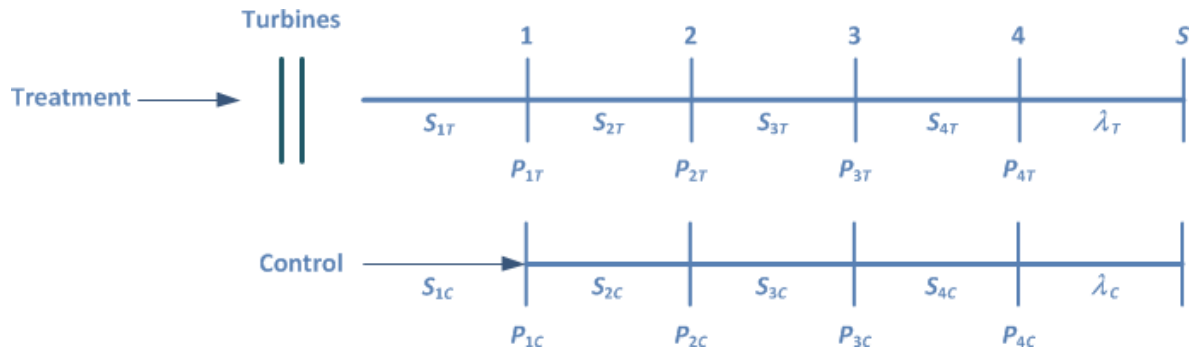


Figure 8.1. Simulated paired releases of salmon smolts in the Columbia River system with four downstream dams.

The assumption of a paired-release study is that both the treatment and control mix downstream of the turbine passage experienced by the treatment fish, thus having common detection and survival probabilities. Burnham et al. (1987) recommends a series of models to use to test for downstream mixing, as illustrated in Table 8.1. The hypothesis testing begins at the most downstream location and moves upstream.

Table 8.1. Recommended series of models to test for downstream mixing from Burnham et al. (1987).

Name	H	H _A
H5S	$\lambda_T = \lambda_C$	$\lambda_T \neq \lambda_C$
H4P	$p_{4T} = p_{4C}$	$p_{4T} \neq p_{4C}$
H4S	$S_{4T} = S_{4C}$	$S_{4T} \neq S_{4C}$
H3P	$p_{3T} = p_{3C}$	$p_{3T} \neq p_{3C}$
H3S	$S_{3T} = S_{3C}$	$S_{3T} \neq S_{3C}$
H2P	$p_{2T} = p_{2C}$	$p_{2T} \neq p_{2C}$
H2S	$S_{2T} = S_{2C}$	$S_{2T} \neq S_{2C}$
H1P	$p_{1T} = p_{1C}$	$p_{1T} \neq p_{1C}$
H0	$S_{1T} = S_{1C}$	$S_{1T} \neq S_{1C}$

8.1.1 Create the Models

There are no covariate data, so only main effects are modeled in this example.

Figure 8.2 shows the **Main Effect** dialogs for detection, survival, and product parameters in their default settings for this example.

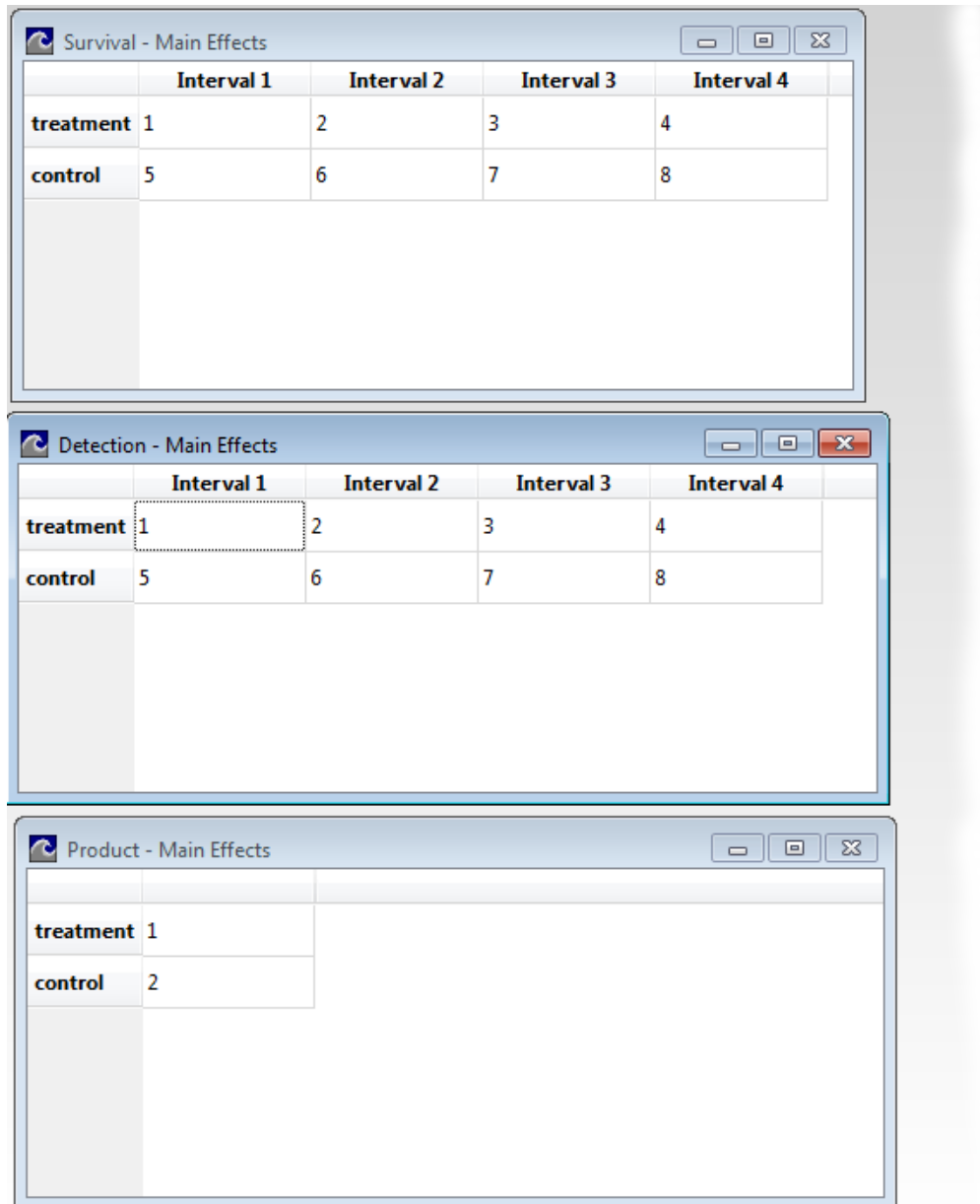


Figure 8.2. **Main Effects** dialogs for the Paired Release example, full model.

This is the full parameterization and represents the alternate hypothesis for the model named **H5S** in Table 8.1. So, double-click on **Create Model** under **Survival Modeling** and name the model **Fu11**. To create the model for the null hypothesis, click

on the **Product - Main Effects** dialog. The carat on the **Navigation Panel** should now show it to be the active window as shown in Figure 8.3.

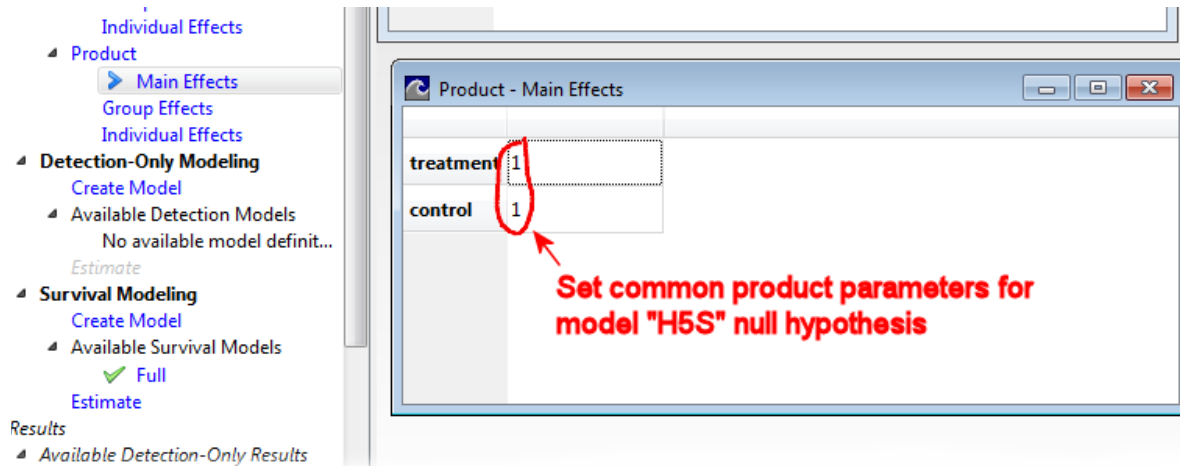


Figure 8.3. **Main Effects** setting for H5S model null hypothesis.

Now click on the **Common** button on the **Effects Control Buttons**, equating the product parameters as shown in Figure 8.3. Double-click on **Create Model** and name the model **H5S**.

We proceed to create the remaining models. For the **H4P** model, ensure that the **H5S** model is active, and change the **Detection - Main Effects** as shown in Figure 8.4. Create the model and name it **H4P**.

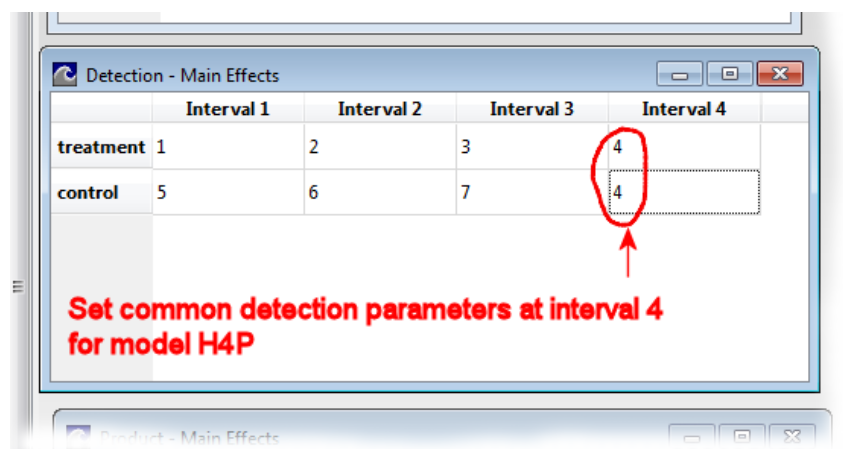


Figure 8.4. Detection main effects for the **H4P** model for the Paired Release example.

Similarly, for the H_{4S} model, ensure that the H_{4P} model is active, and set the **Survival - Main Effects** as shown in Figure 8.5. Continue in a similar fashion for all models in Table 8.1.

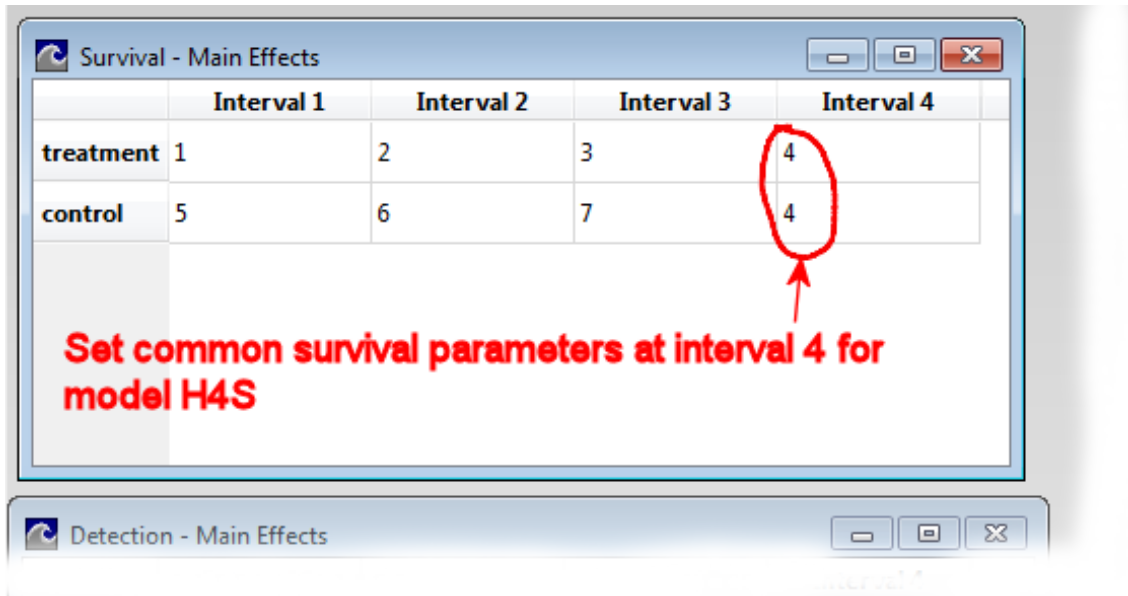


Figure 8.5. Survival main effects for the H_{4S} model for the Paired Release example.

Figure 8.6 shows the **Detection - Main Effects** settings for model H_{3P} .

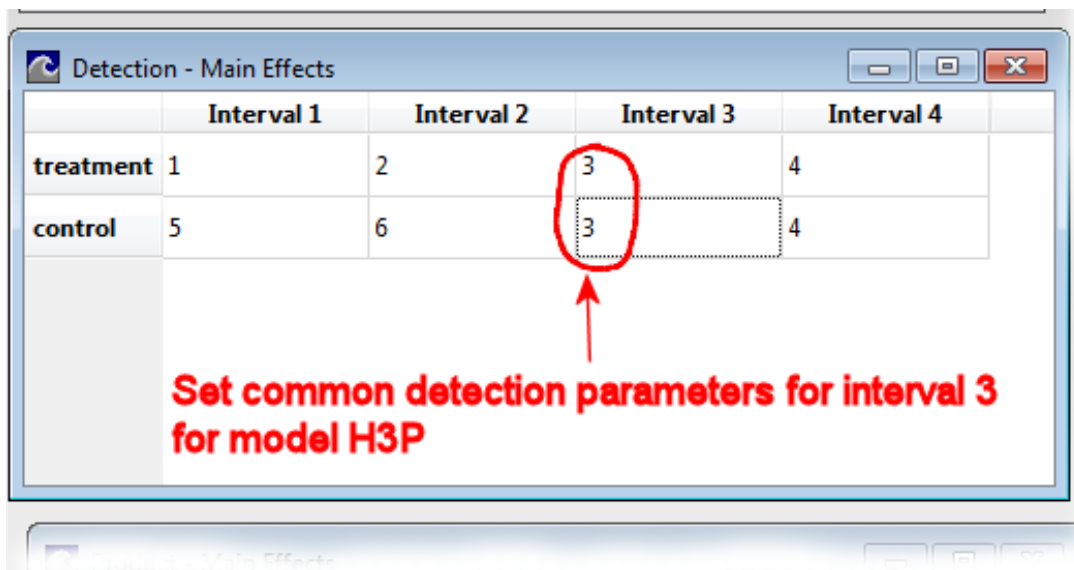


Figure 8.6. Detection main effects for the H_{3P} model for the Paired Release example.

The settings for the final model, H_0 , should look as shown in Figure 8.7.

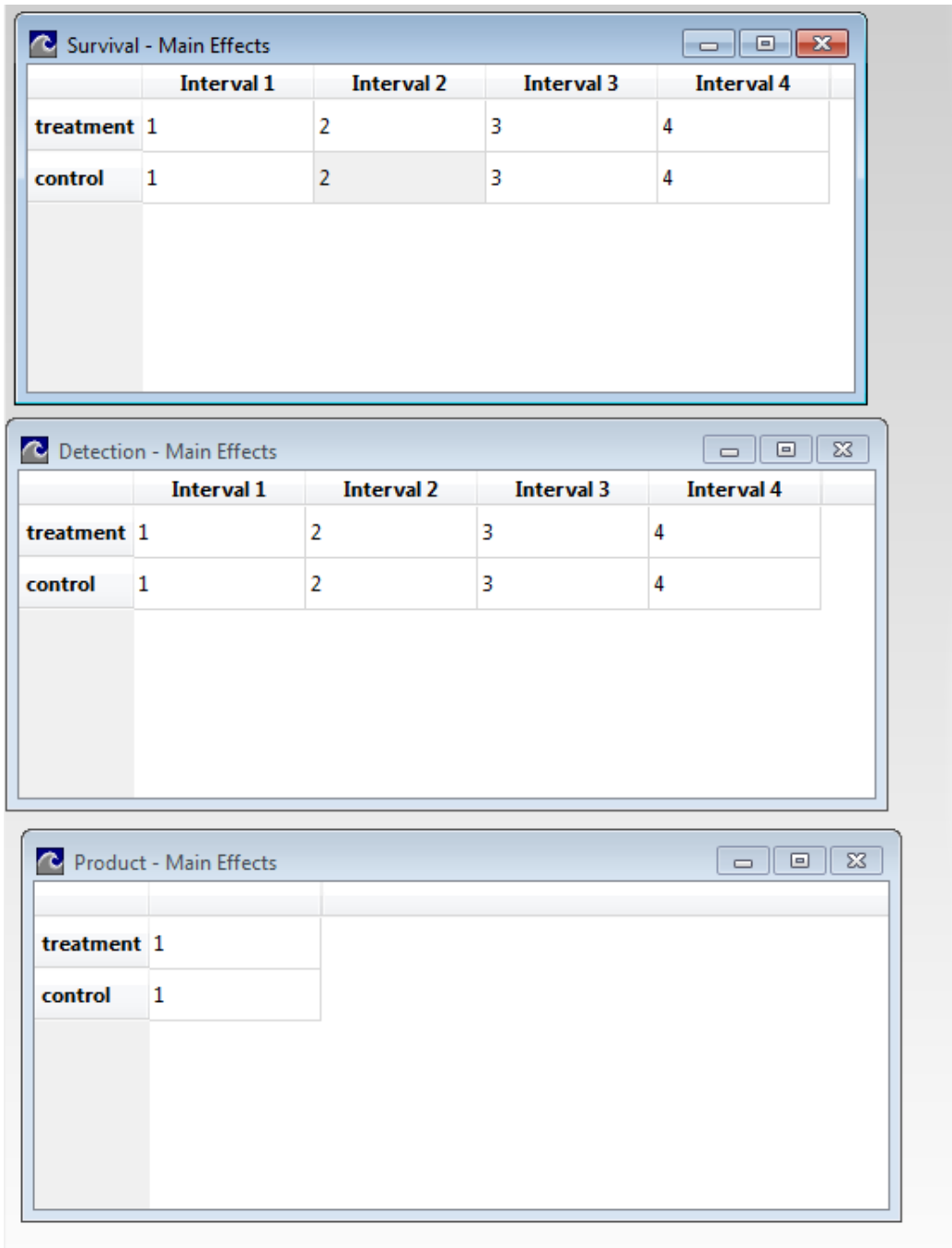


Figure 8.7. Main effects for the H_0 model for the Paired Release example.

The **Navigation Panel** should now look as shown in Figure 8.8.

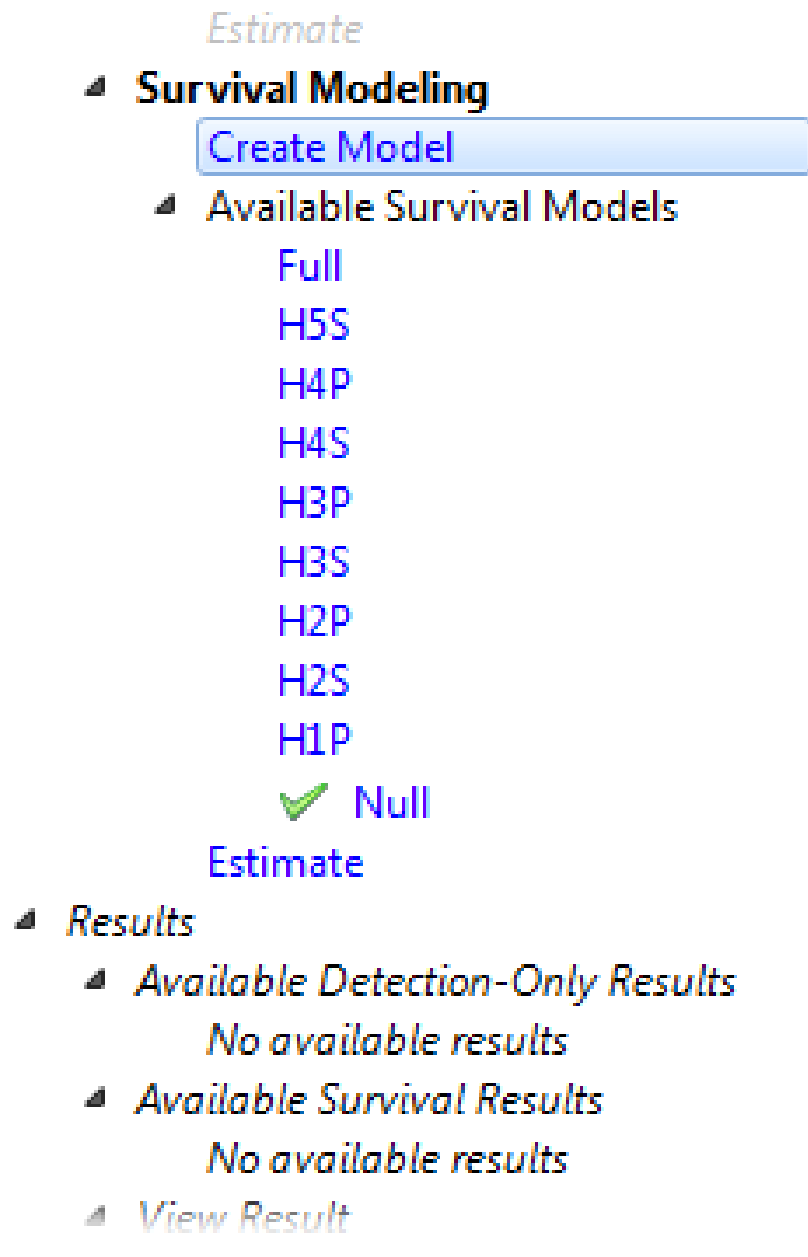


Figure 8.8. All defined models for the Paired Release example.

Double-click on **Estimate** under **Survival Modeling** to create the results.

8.1.2 Test the Hypotheses

To test the hypotheses, double-click on **LRT and AIC** under **Hypothesis Testing** on the **Navigation Panel** and select **Survival** for the **Model Type** at the bottom of the

dialog. The **Likelihood Ratio Test** dialog appears as shown in Figure 8.9. To test the hypothesis H5S, simply click on the **Full** result on the **Master Model List** and **H5S** on the **Testable Model List**.

The corresponding Likelihood Ratio Test statistic of 0.8347 is shown at the bottom, with one degree of freedom and a *P*-value of 0.36, indicating that the **H5S** model fits the data as well as the **Full** model.

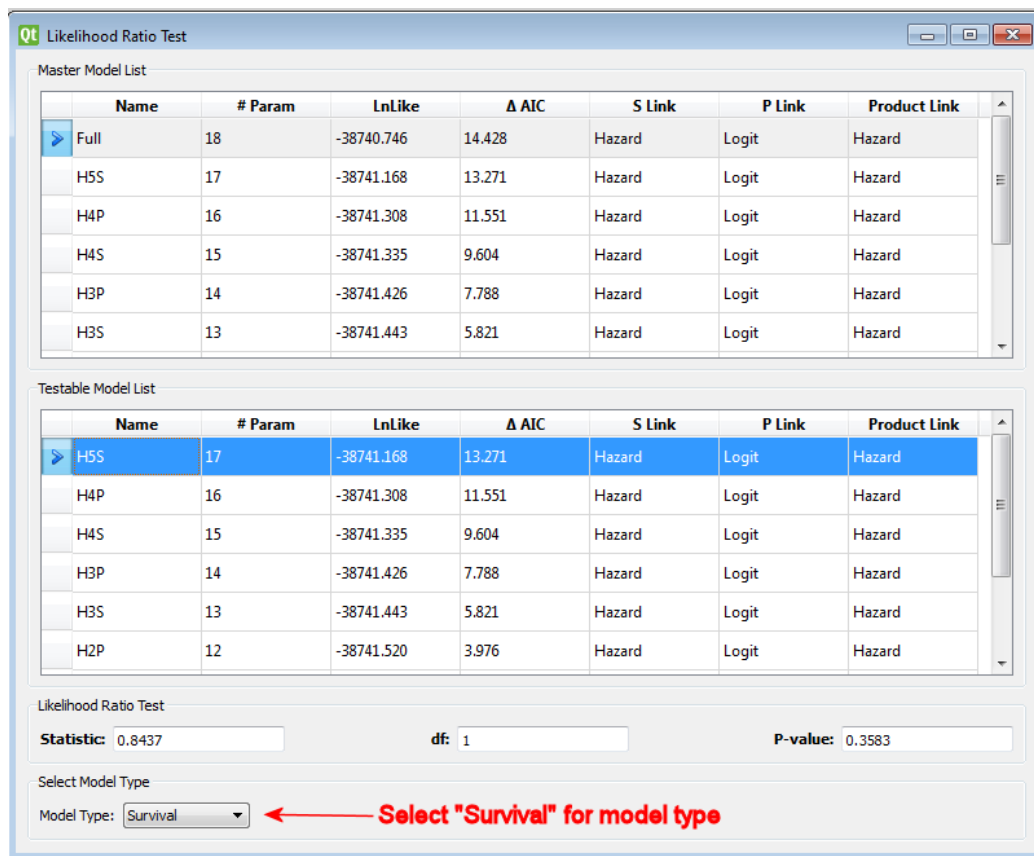


Figure 8.9. **Likelihood Ratio Test** dialog for the Paired Release example.

Next, we can test H5S against H4P as shown in Figure 8.10. Again, the result is insignificant with a *P*-value of 0.60. We can continue the testing in a similar manner. In this example, all the tests have an insignificant *P*-value until we test H0 vs H1P, which has a significant *P*-value. This is what we would expect if there is a significant effect on mortality due to turbine passage. The insignificant *P*-values for the other Likelihood Ratio Tests show that the two releases are mixing downstream.

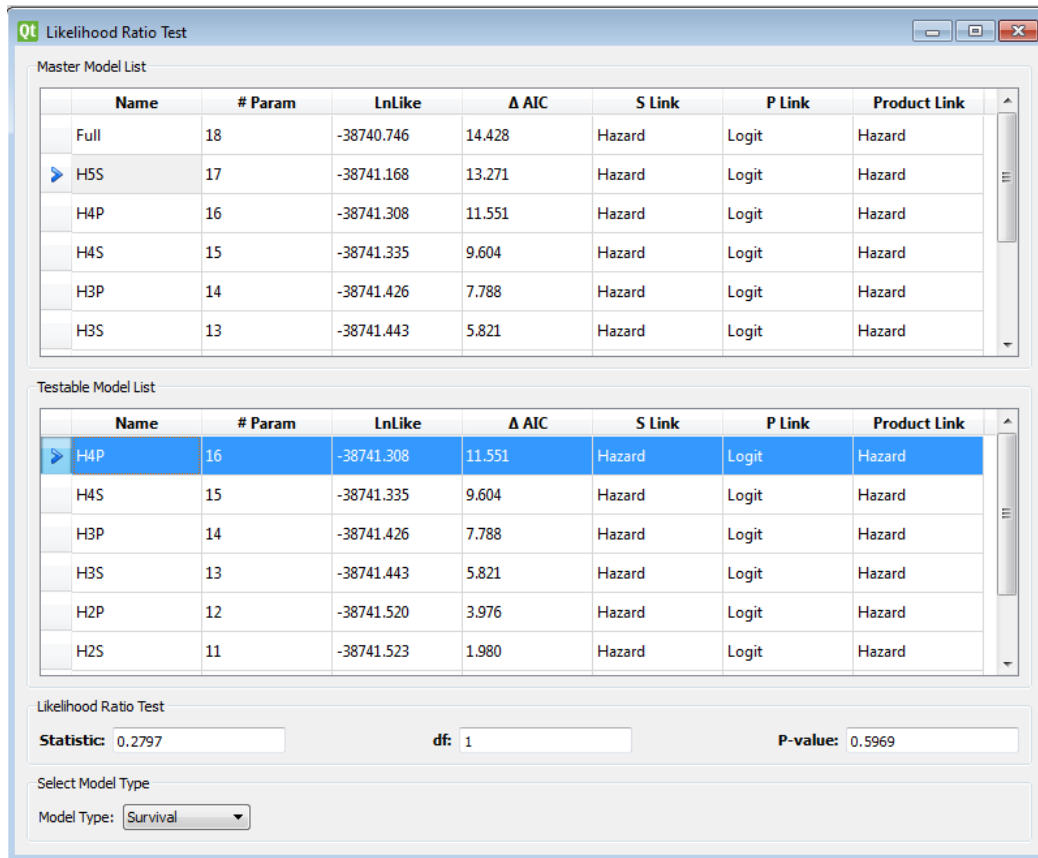


Figure 8.10. H4P vs. H4S for the Paired Release example.

8.1.3 Increase Precision by Reducing the Number of Parameters

The parameter estimates can be examined by double-clicking on **Parameter Report** under **View Results**. Figure 8.11 shows the top portion of the **Parameter Report** for the final H1P model. Notice the large standard error for survival parameter #3.

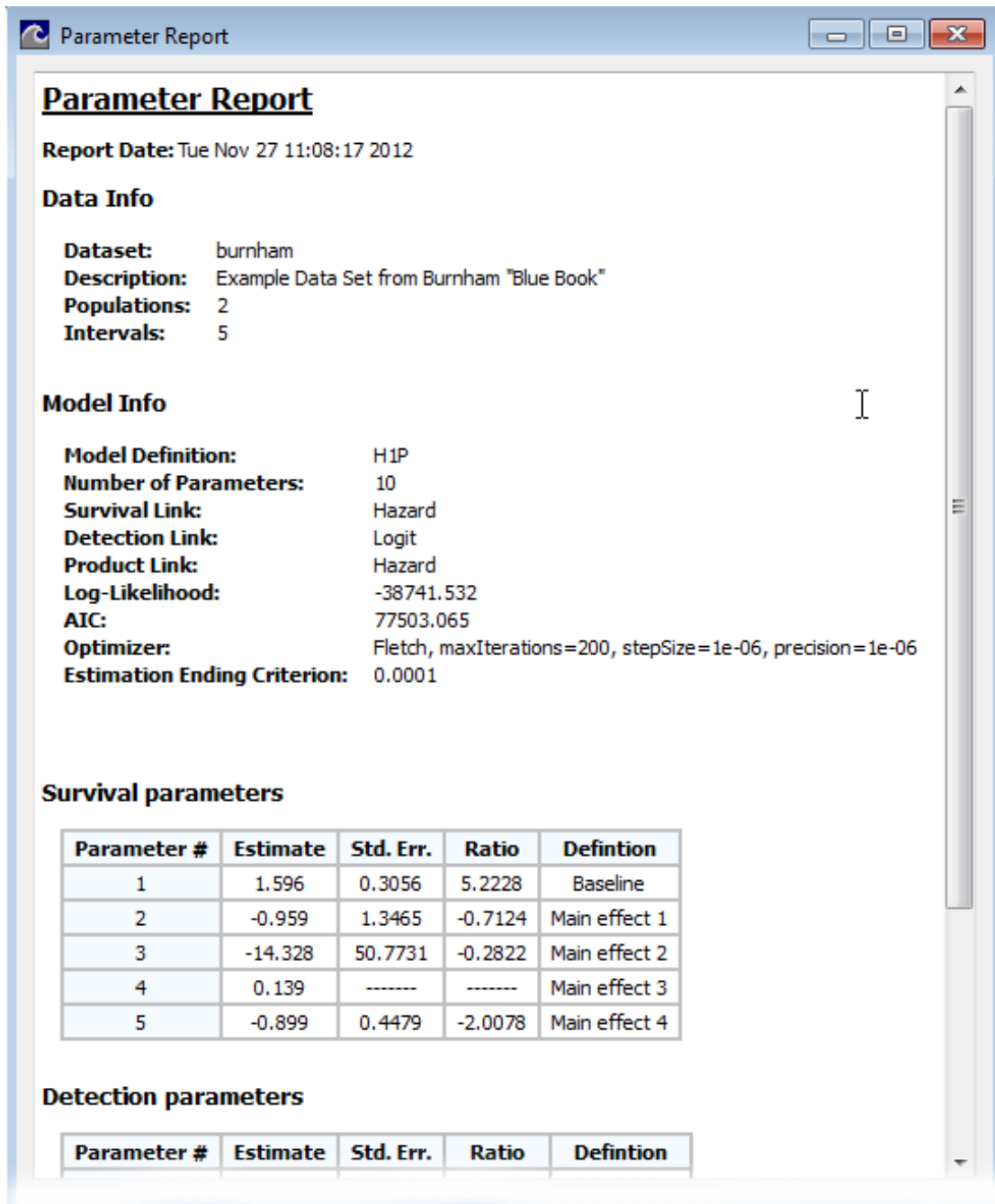


Figure 8.11. Top portion of the **Parameter Report** for the H1P results for the Paired Release example

Double-click on **Probability Report** to obtain the resulting probability estimates (Figure 8.12). Notice the large standard errors for s_1 and s_2 , with s_3 pegged at 1.0 with 0.0 variance. The evidence suggests that it may be possible to model the downstream

survival probabilities with a common survival parameter at each detection site and thus increase the resulting precision of the survival probabilities of interest.

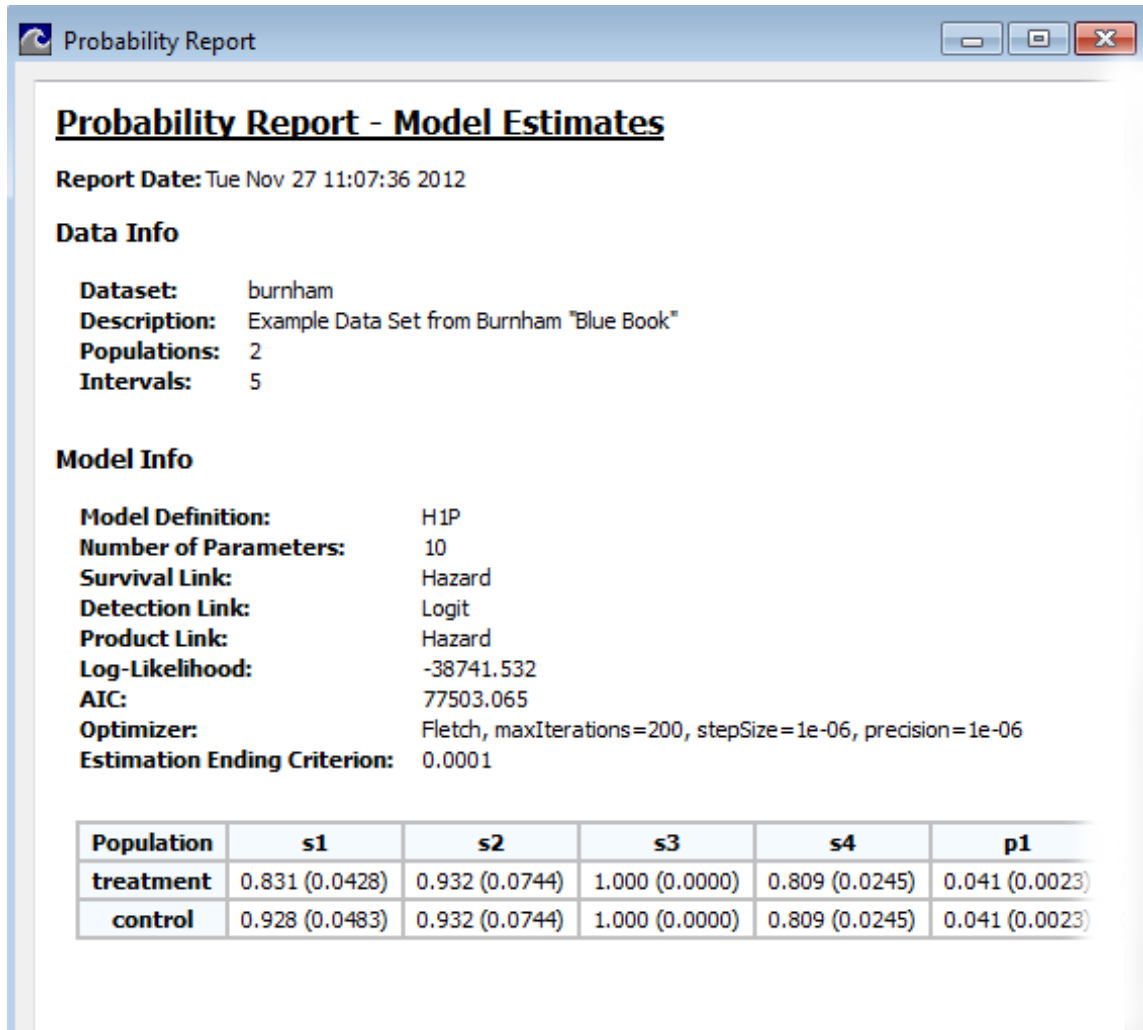


Figure 8.12. **Probability Report** for the **H1P** results for the Paired Release example.

Set the survival main effects as shown in Figure 8.13 and create the model **H1P_common**. Note that all survival probabilities share a common parameter except for the treatment survival (survival through the turbine passage). After estimating the parameters, the Likelihood Ratio Test comparing **H1P_common** with **H1P** yields a *P*-value of 0.76 as shown in Figure 8.14, indicating that the new model fits the data as well as the model with unique survival probabilities.

	Interval 1	Interval 2	Interval 3	Interval 4
treatment	1	2	2	2
control	2	2	2	2

Figure 8.13. Survival main effects with one common survival parameter downstream of the treatment effect for the Paired Release example.

Master Model List

Name	# Param	LnLike	Δ AIC	S Link	P Link	Product Link
H3S	13	-38741.443	10.648	Hazard	Logit	Hazard
H2P	12	-38741.520	8.802	Hazard	Logit	Hazard
H2S	11	-38741.523	6.807	Hazard	Logit	Hazard
H1P	10	-38741.532	4.827	Hazard	Logit	Hazard
Null	9	-38756.347	32.456	Hazard	Logit	Hazard
H1P_common	7	-38742.119	0.000	Hazard	Logit	Hazard

Testable Model List

Name	# Param	LnLike	Δ AIC	S Link	P Link	Product Link
Null	9	-38756.347	32.456	Hazard	Logit	Hazard
H1P_common	7	-38742.119	0.000	Hazard	Logit	Hazard

Likelihood Ratio Test

Statistic: 1.1733 df: 3 P-value: 0.7594

Select Model Type

Model Type: Survival

Figure 8.14. Likelihood Ratio Test dialog for the H1P_common result for the Paired Release example.

Figure 8.15 shows the probability estimates of the new model. Note the improved precision of the survival probabilities for survival in the first reach for both the treatment and the control.

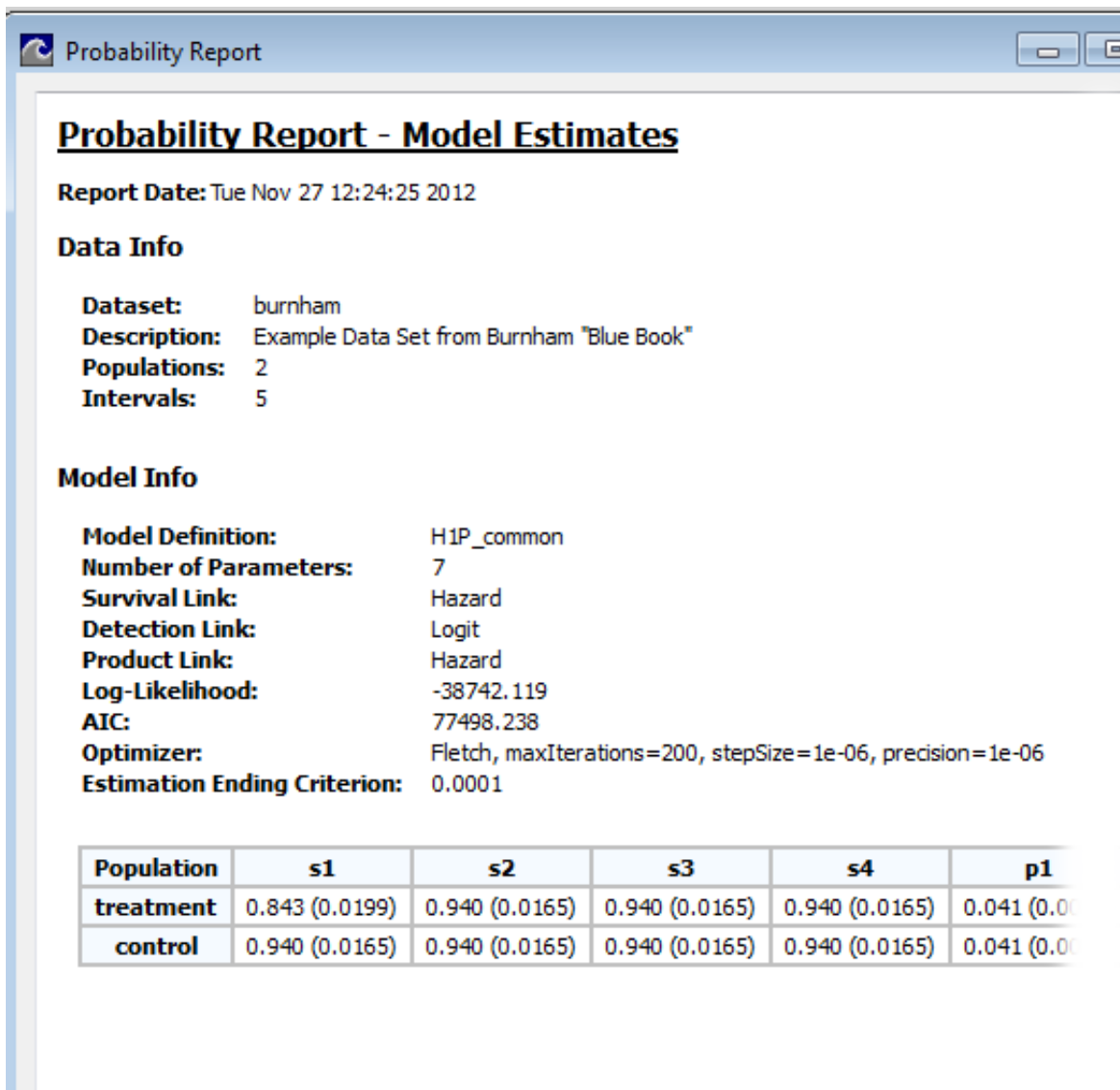


Figure 8.14. **Probability Report** for the final model for the Paired Release example

8.1.4 Results Diagnostics

Under the **View Result Diagnostic Plots** heading on the **Navigation Panel**, double-click on **Modeled Estimates vs. CJS Estimates** to bring up the plot as shown in Figure 8.15. The CJS estimates are the "best" albeit uninformative estimates of the

parameters. If the selected model fits the data well, there should be little discrepancy between the two and all points on the graph should lie near the 45° line. In this case, the detection estimates line up quite well, as do four of the survival estimates. The other three survival estimates do not match up with the CJS estimates well, reflecting the heterogeneity in the CJS point estimates for downstream reaches. However, there is good agreement between the modeled estimate and the CJS estimate for the treatment survival, which is the point with the lower modeled estimate in Figure 8.15. It may be worthwhile investigating an additional model that leaves survival in reach 4 unique compared to the previous three reaches, in particular if survival in that reach is of interest.

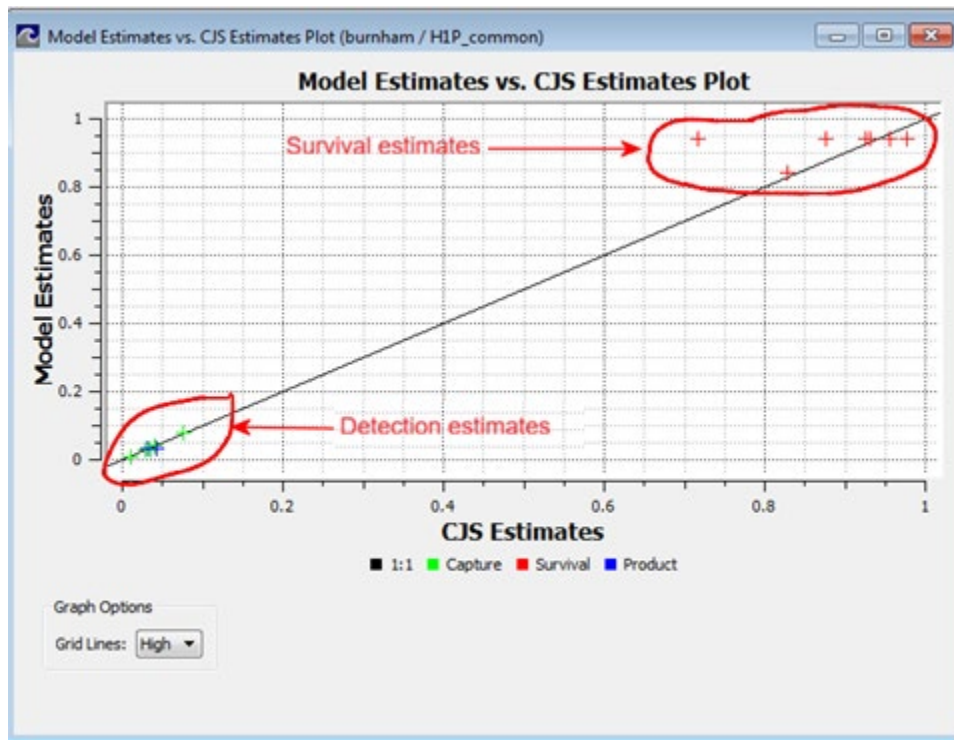


Figure 8.15. **Model Estimates vs. CJS Estimates** plot for the Paired Release example.

Figure 8.16 shows the **Residuals vs Capture Histories** plot. If the model fits the data adequately, one should not observe any trend in the residuals, and 95% of them should be within the two red horizontal lines (plus or minus 1.96).

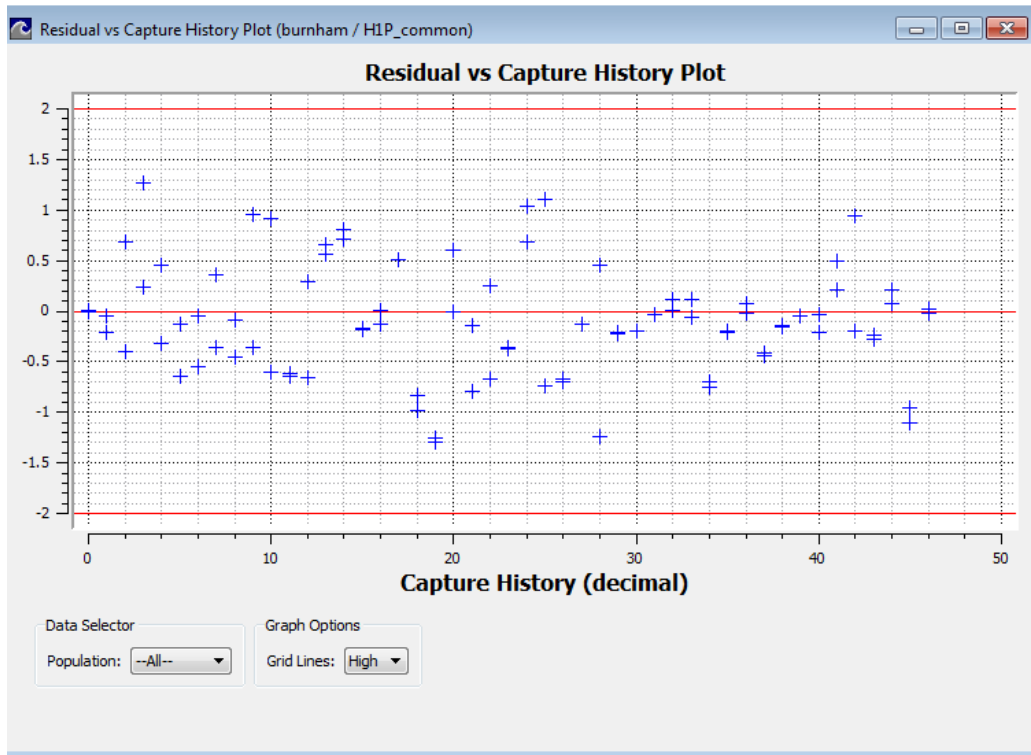


Figure 8.16. **Residuals vs Capture History** plot for the selected model for the Paired Release example.

8.2 Effect of Timing of Releases from a Hatchery Example

This example illustrates two-stage model development and hypothesis testing with group covariate data using Analysis of Deviance.

A study of the effect on timing of release on in-river salmon smolts was conducted in the Columbia Basin River system in 1993. The data are in the file "dworshak.dat." The top portion of the input data file is shown in Figure 8.17.

```

Surph3
Dworshak Hatchery 4/8/93, 4/22/93, and 5/6/93 Releases
npop 9
nint 3

gcov middle
0 0 0 1 1 1 0 0 0

gcov late
0 0 0 0 0 0 1 1 1

icov length

population early1
7F7B011B19 1 0 0 0 118
7F7B01341A 1 0 0 0 138
7F7B013E69 1 0 0 0 115
7F7B013F49 1 0 0 0 119
7F7B014069 1 0 0 0 126
7F7B014A3D 1 0 0 0 139
7F7B014C14 1 0 0 0 109
7F7B016876 1 0 0 0 104
7F7B016E7A 1 0 0 0 118
7F7B017014 1 0 0 0 125

```

Figure 8.17. Input file shows the use of an indicator group covariate to indicate release timing.

The group covariates `middle` and `late` are indicator variables with “1,” indicating that the group covariate applies to the population. The first three populations were released early in the season, populations 4 through 6 were released in the middle of the season, and populations 7 through 9 late in the season. There is no group covariate for early release—this parameterization defines early release as the baseline survival.

Note that the flexibility of SURPH would allow the modeling to be done without the use of the indicator group covariates. However, as noted in Chapter 6, ANODEV is the preferred method of hypothesis testing with group covariates; release timing can be considered a group covariate, so indicator group covariates are used to allow ANODEV to be used.

8.2.1 Detection-only Modeling

In this example we will create four models encapsulating the four basic detection models: (1) population/interval specific, (2) population-specific, (3) interval-specific, and (4) common. First, open the **Detection Main Effects** dialog, and press the **Population / Interval Specific** button on the **Effects Control Buttons**. Create the detection-only model, and name it **Full** for full parameterization. Then press the **Population Specific** button and create a detection-only model named **Population specific**. Do the same for naming an **Interval specific** model and a **Common** model. Once all four models have been created, double-click on **Estimate** under **Detection-Only Modeling** to estimate the parameters for all four models. The **Navigation Panel** should now appear as shown in Figure 8.18.

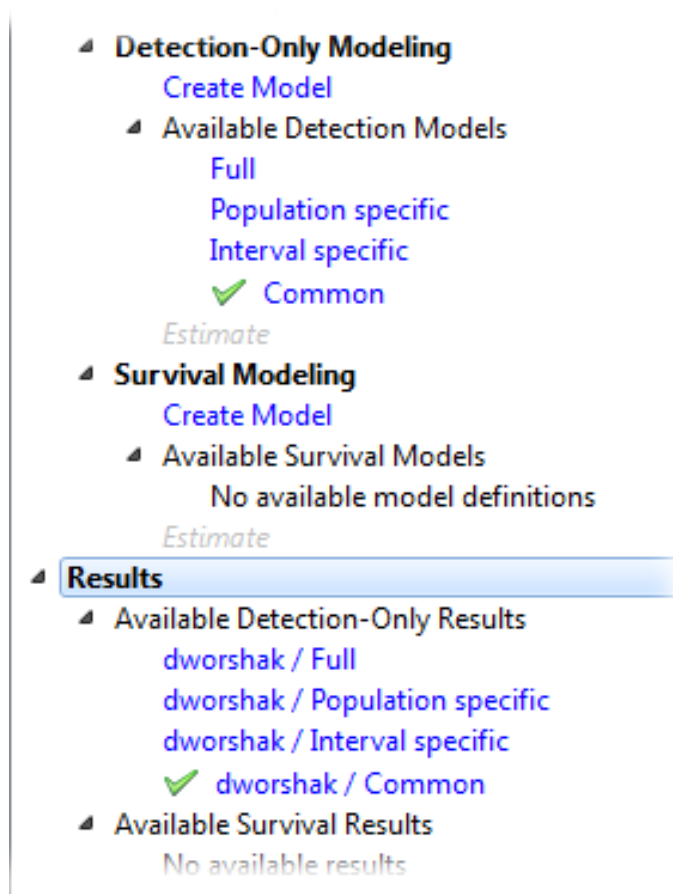


Figure 8.18. Four basic detection-only results created for the Timing of Release example.

Next double-click on **LRT and AIC** under **Hypothesis Testing**. Ensure that the **Model Type** at the bottom of the dialog is set to **Detection-Only**. When we click on the **Full** model in the **Master Model List** and click on each model in the **Testable Model List**, the *P*-value is significant for each Likelihood Ratio Test, indicating that the **Full** model is the best model for the detection probabilities. However, we have not yet taken into account the structure of the study design: the first three populations represent the early release, populations 4 through 6 represent the middle- timed release, and the final three populations represent the late release. Double-click on **Manly-Parr Estimates** under **Data Diagnostics** to bring up the closed-form estimates of detection probability as shown in Figure 8.19. They suggest that there may be a difference in detection probabilities among the releases. If we can reduce the number of parameters for the detection probabilities, we can increase the precision of the survival estimates.

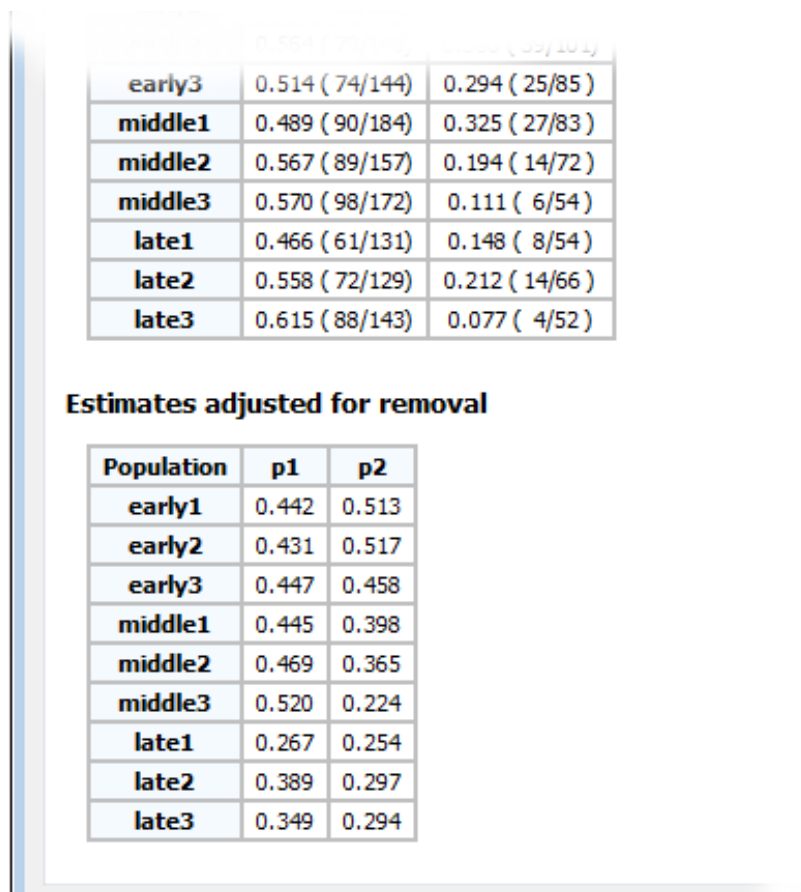
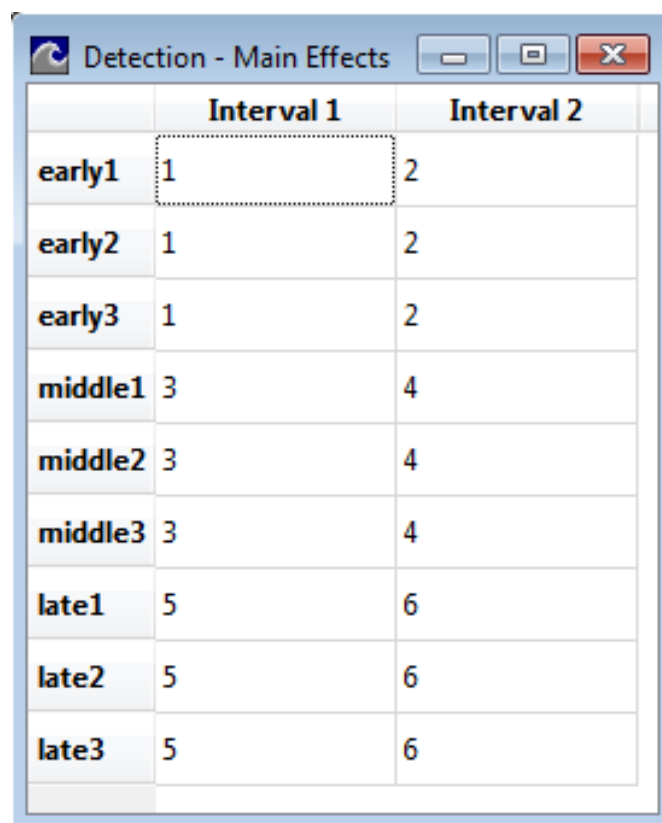


Figure 8.19. Bottom portion of the **Manly-Parr Estimates** report for the Timing of Release example.

Set the **Detection - Main Effects** dialog as shown in Figure 8.20 and create the detection-only model **grouped**. In this parameterization, we assume common detection probabilities within a release group but unique detection probabilities across releases groups. Double-click on **Estimate** to estimate the parameters.



	Interval 1	Interval 2
early1	1	2
early2	1	2
early3	1	2
middle1	3	4
middle2	3	4
middle3	3	4
late1	5	6
late2	5	6
late3	5	6

Figure 8.20. **Detection main effects** settings for custom parameterization for the Timing of Release example.

Returning to the **Likelihood Ratio Test** dialog, the new model should appear under the **Testable Model List**, as shown in Figure 8.21. When we test the **Full** model against the **grouped** model, the *P*-value is not significant, indicating that the grouped model fits the data as well as the **Full** model and reduces the number of parameters from 18 to 6. This will increase the precision of the resulting survival probability estimates.

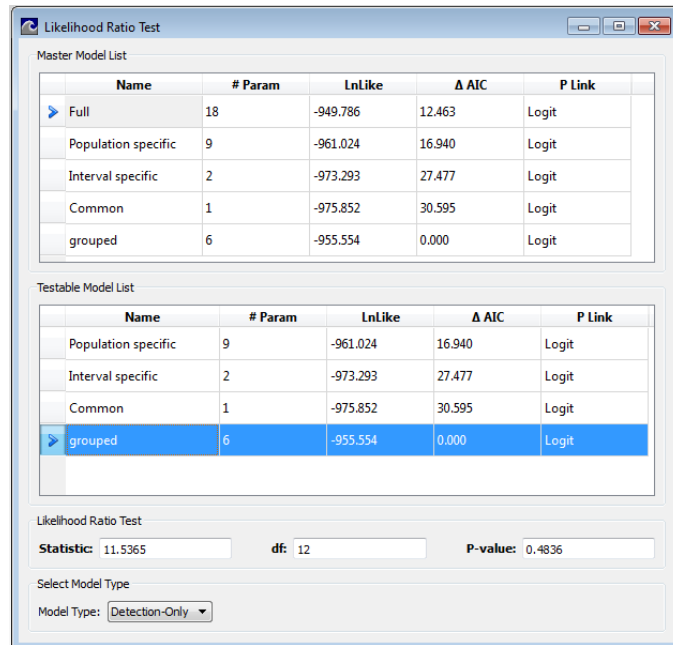


Figure 8.21. **Likelihood Ratio Test** dialog for detection-only models for the Timing of Release example.

We can now lock the detection parameterization and move on to modeling the survival probabilities. Right-click on the **dworshak/grouped** result and select lock as shown in Figure 8.22.

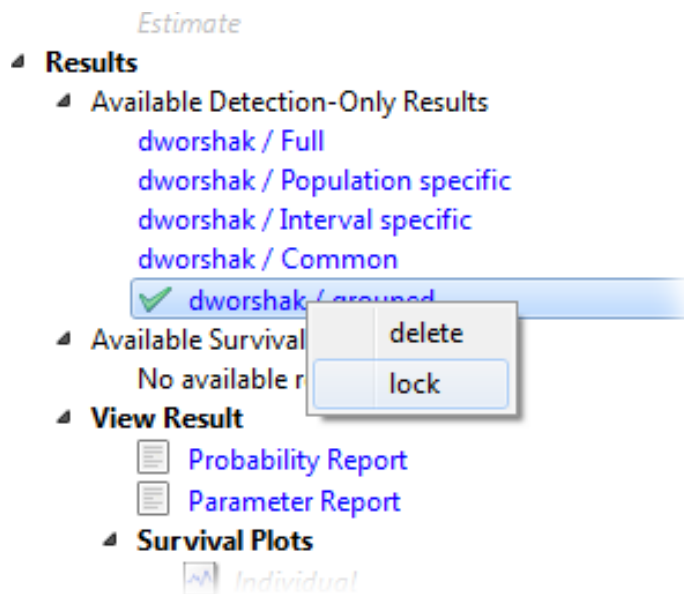


Figure 8.22. Lock the **dworshak/grouped** result for the Timing of Release example.

The **Navigation Panel** should now appear as in Figure 8.23. The detection parameterization can no longer be changed without first unlocking the locked result.

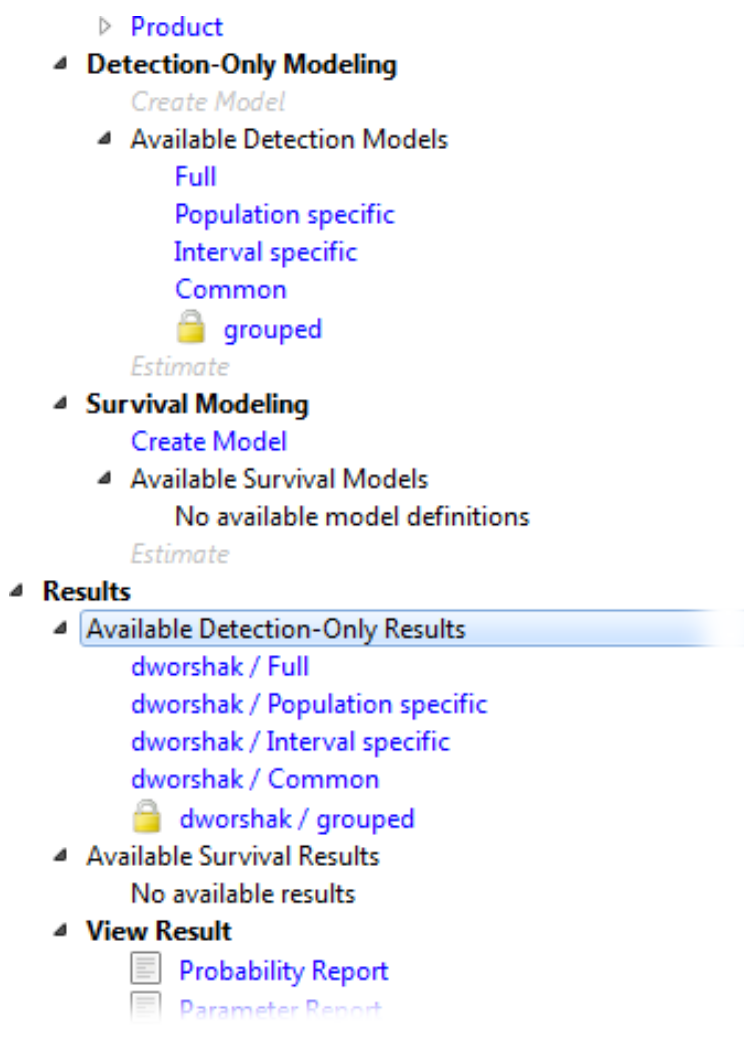
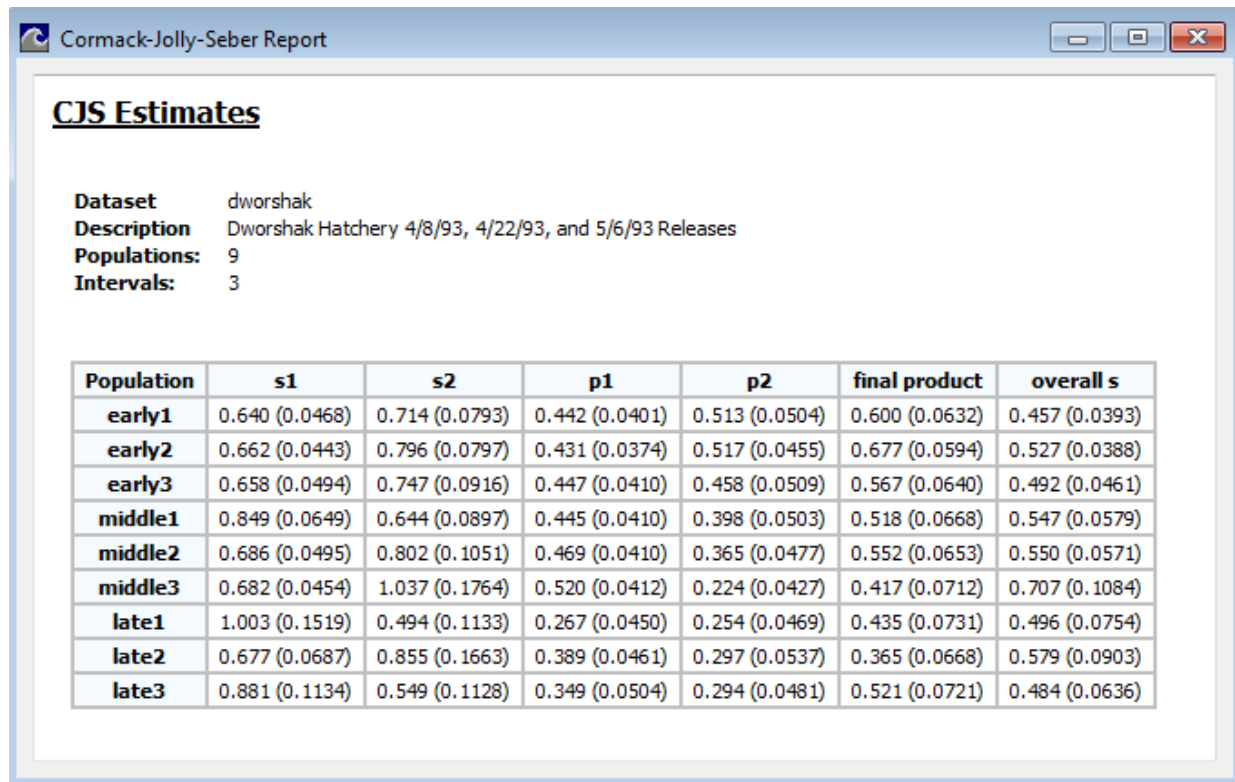


Figure 8.23. The **dworshak/grouped** detection-only result is now locked for the Timing of Release example.

8.2.2 Survival Modeling with Analysis of Deviance

Figure 8.24 shows the CJS estimates (double-click on **Estimates** under **CJS Estimates** in the **Data** section of the **Navigation Panel**). As with the detection probabilities, there is evidence that there may be a difference in survival probabilities across release groups. We will look at the effect of release timing on survival using Analysis of Deviance.



CJS Estimates

Dataset: dworshak
Description: Dworshak Hatchery 4/8/93, 4/22/93, and 5/6/93 Releases
Populations: 9
Intervals: 3

Population	s1	s2	p1	p2	final product	overall s
early1	0.640 (0.0468)	0.714 (0.0793)	0.442 (0.0401)	0.513 (0.0504)	0.600 (0.0632)	0.457 (0.0393)
early2	0.662 (0.0443)	0.796 (0.0797)	0.431 (0.0374)	0.517 (0.0455)	0.677 (0.0594)	0.527 (0.0388)
early3	0.658 (0.0494)	0.747 (0.0916)	0.447 (0.0410)	0.458 (0.0509)	0.567 (0.0640)	0.492 (0.0461)
middle1	0.849 (0.0649)	0.644 (0.0897)	0.445 (0.0410)	0.398 (0.0503)	0.518 (0.0668)	0.547 (0.0579)
middle2	0.686 (0.0495)	0.802 (0.1051)	0.469 (0.0410)	0.365 (0.0477)	0.552 (0.0653)	0.550 (0.0571)
middle3	0.682 (0.0454)	1.037 (0.1764)	0.520 (0.0412)	0.224 (0.0427)	0.417 (0.0712)	0.707 (0.1084)
late1	1.003 (0.1519)	0.494 (0.1133)	0.267 (0.0450)	0.254 (0.0469)	0.435 (0.0731)	0.496 (0.0754)
late2	0.677 (0.0687)	0.855 (0.1663)	0.389 (0.0461)	0.297 (0.0537)	0.365 (0.0668)	0.579 (0.0903)
late3	0.881 (0.1134)	0.549 (0.1128)	0.349 (0.0504)	0.294 (0.0481)	0.521 (0.0721)	0.484 (0.0636)

Figure 8.24. CJS estimates for the Timing of Release example.

Double-click on **ANODEV** under **Hypothesis Testing**. On the **ANODEV Setup** dialog, press **Continue** to accept the default choice of **Use Defaults**. SURPH will then create the three basic ANODEV models as described in Chapter 6, and the **Analysis of Deviance** dialog will appear as shown in Figure 8.25.

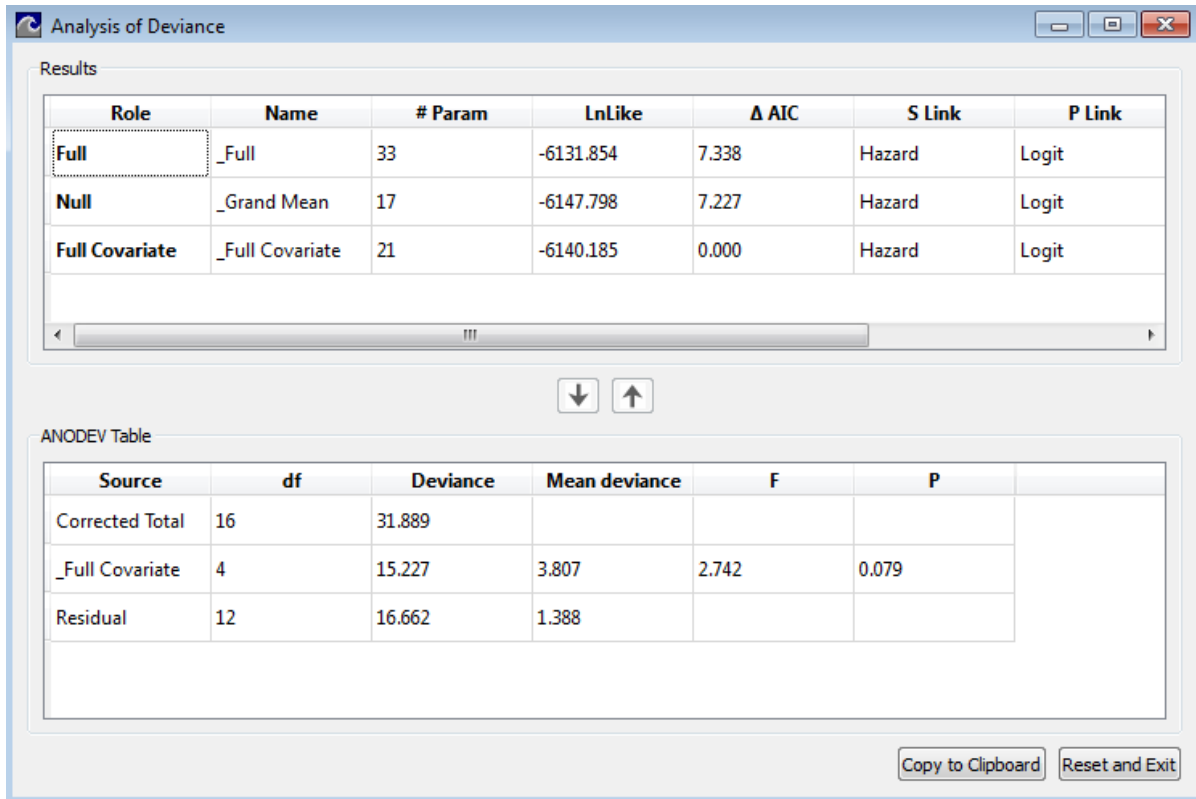


Figure 8.25. Analysis of Deviance dialog for the Timing of Release example.

A *P*-value of 0.079 for the **Full Covariate** model indicates that the release timing is marginally significant. However, the **Full Covariate** model assumes that the covariate effect applies to both intervals. It may be that release timing primarily affects survival in the first interval.

To test this, we need to create a model with this parameterization. Make sure that the **_Full Covariate** model is the active model, open the **Group Effects** dialog for survival, and set the parameterization for both covariates `middle` and `late` as shown in Figure 8.26.

	Interval 1	Interval 2
early1	1	0
early2	1	0
early3	1	0
middle1	1	0
middle2	1	0
middle3	1	0
late1	1	0
late2	1	0
late3	1	0

Figure 8.26. Effect settings for testing the effect of covariates on survival in Interval 1 only for the Timing of Release example.

The “1s” in the **Interval 1** column indicate that the covariate is to be included for that interval. The “0s” for **Interval 2** indicate that the interval 2 survival is not modeled as a function of the covariate. Double-click on **Create Model** under **Survival Model** and name the model `interva11`. Now return to the **Analysis of Deviance** dialog.

The **Analysis of Deviance** dialog should now appear as in Figure 8.27.

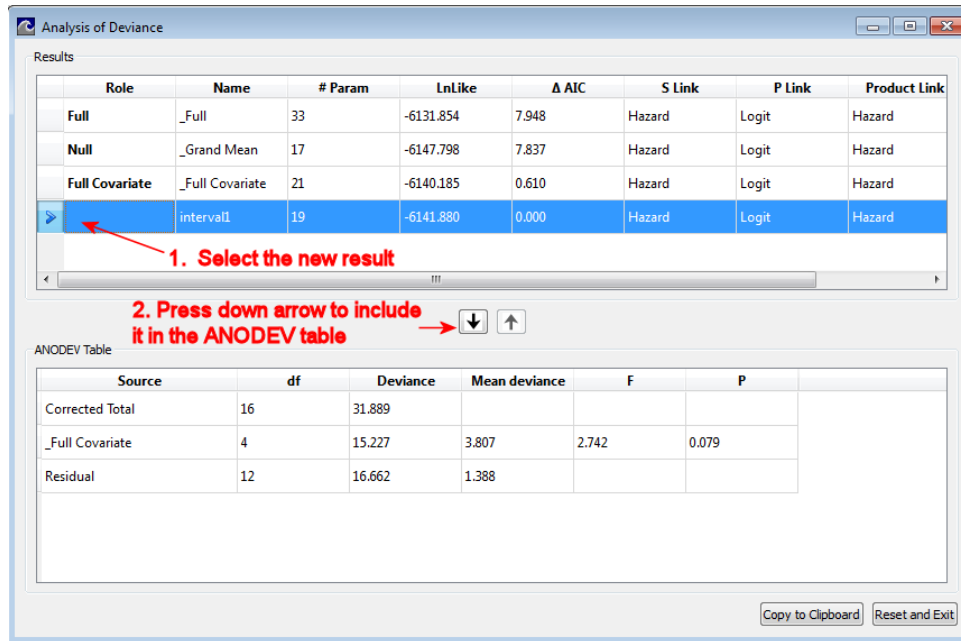


Figure 8.27. **Analysis of Deviance** dialog with the `interval1` result ready to be included for the Timing of Release example.

Select the new model and press the down arrow to include it in the ANODEV model.

The **Analysis of Deviance** dialog should now appear as in Figure 8.28. The P -value 0.04 for the line `interval1 vs _Grand Mean` indicates that the model that included an effect for release timing in the first interval survival is a significantly better fit than the null model (`_Grand Mean`). However, comparing the **Full Covariate** model with the `interval1` model produces an insignificant P -value (0.33), indicating that release timing does not have a significant effect on survival in the second reach.

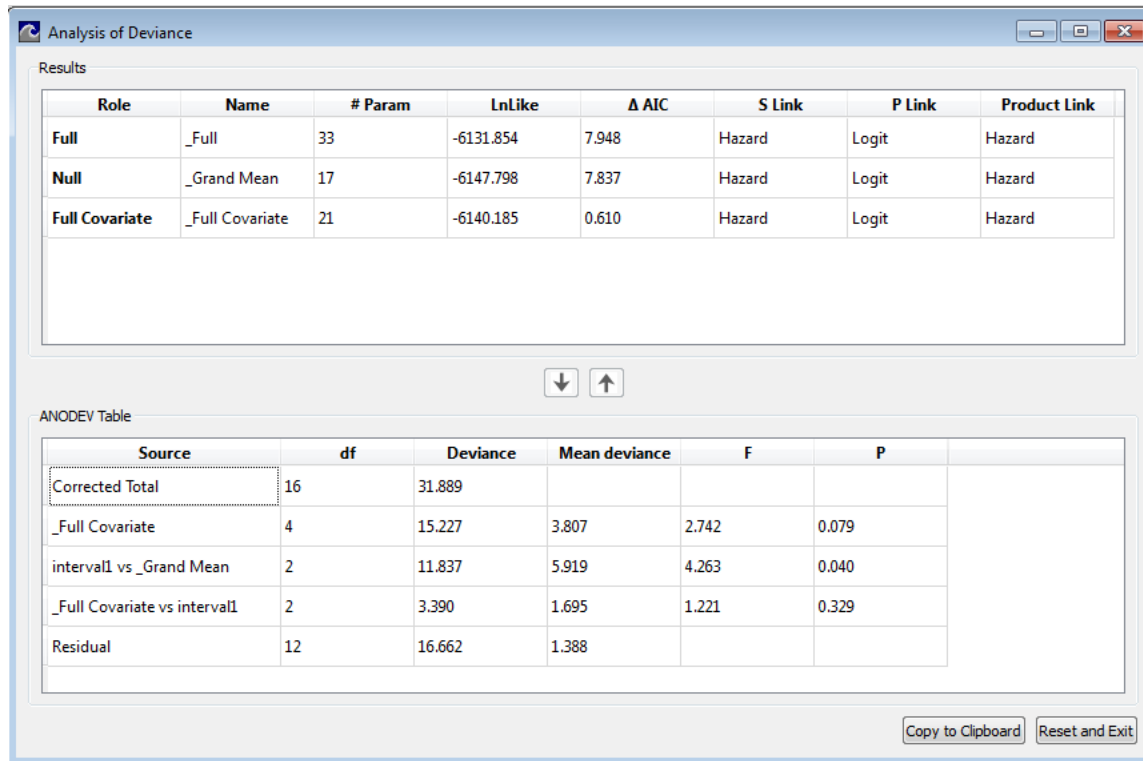


Figure 8.28. **Analysis of Deviance** dialog with new model `interval1` included for the Timing of Release example.

8.3 Relationship of Migration Distance to Survival

This example illustrates modeling survival as a function of a group covariate and using the **Group Survival Plot** to visualize the modeled survival probabilities as a function of a group covariate.

The data for this example is in the file `distance.dat`. Figure 8.29 shows that the first three populations traveled 116 kilometers to the first detection point (dam), the fourth 176 kilometers, and so on.

```

Surph3
Yearling Chinook Releases from Snake River Hatcheries - 1993
npop 10
nint 3
icov length

gcov distance
116 116 116 176 235 135 467 467 277 748

population 1
7F7B011B19 1 0 0 0 118
7F7B01341A 1 0 0 0 138
7F7B013E69 1 0 0 0 115
7F7B013F49 1 0 0 0 119
7F7B014069 1 0 0 0 126
7F7B014A3D 1 0 0 0 139
7F7B014C14 1 0 0 0 109
7F7B016876 1 0 0 0 104
7F7B016E7A 1 0 0 0 118
7F7B017014 1 0 0 0 125

```

Figure 8.29. Data file for the Migration Distance example, showing the covariate data.

We can perform two-stage modeling and begin with modeling the detection probability as was done in the previous example (Section 8.2). The result of this will be to select the **Full** model with population- and interval-specific effects for the detection model. There is no reason to believe that migration distance will have any effect on detection probabilities.

8.3.1 Analysis of Deviance

Once the **Full** detection model is selected, double-click on **ANODEV** and select the **Use Defaults** option. The **Analysis of Deviance** dialog will appear as shown in Figure 8.30. A P -value 0. 0.000 for the **Full Covariate** model indicates that migration distance has a very significant effect on survival. It might be a reasonable assumption that it would affect survival for the first interval but not for the second. We can test this by creating a new model **distance1** that models only the first interval survival as a function of distance.

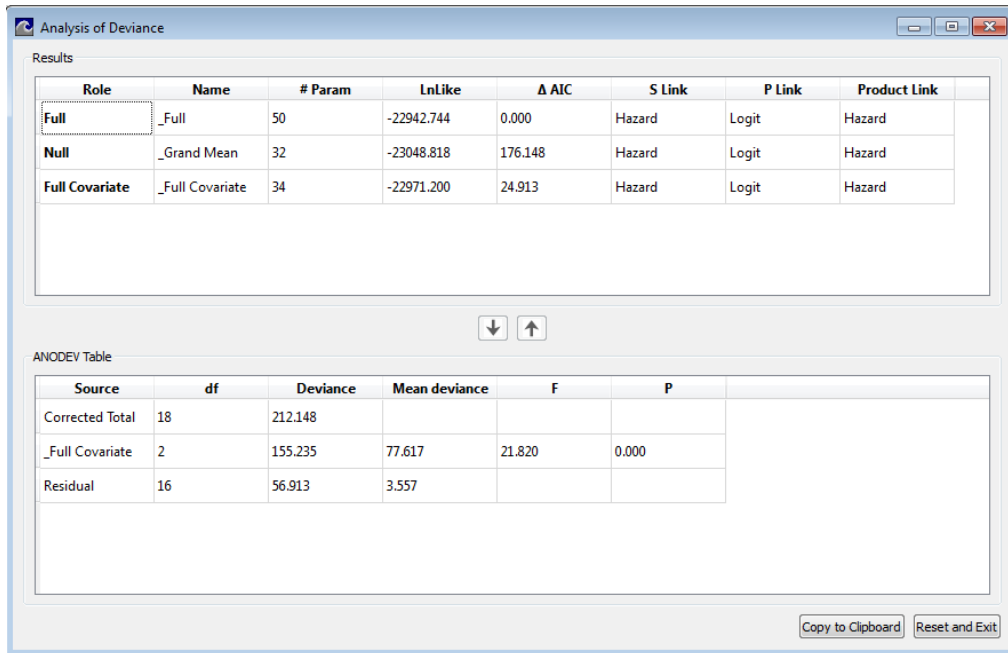


Figure 8.30. **Analysis of Deviance** dialog for the Migration Distance example.

Make sure that the model **_Full Covariate** is the active model and open the **Group Effects** dialog for **Survival**. Set it to the configuration shown in Figure 8.31.

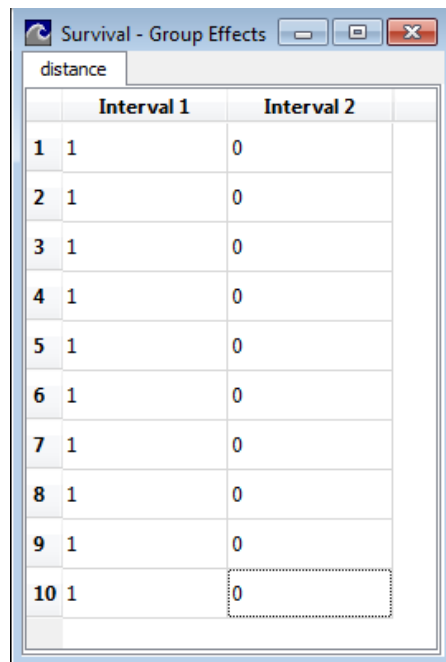


Figure 8.31. Survival effects for modeling survival in the first interval as a function of distance for the Migration Distance example.

Create the survival model, naming it `distance1`. After double-clicking **Estimate**, the new model will show up in the **Analysis of Deviance** dialog. Add it to the ANODEV table by selecting it and pressing the down arrow.

The result is shown in Figure 8.32.

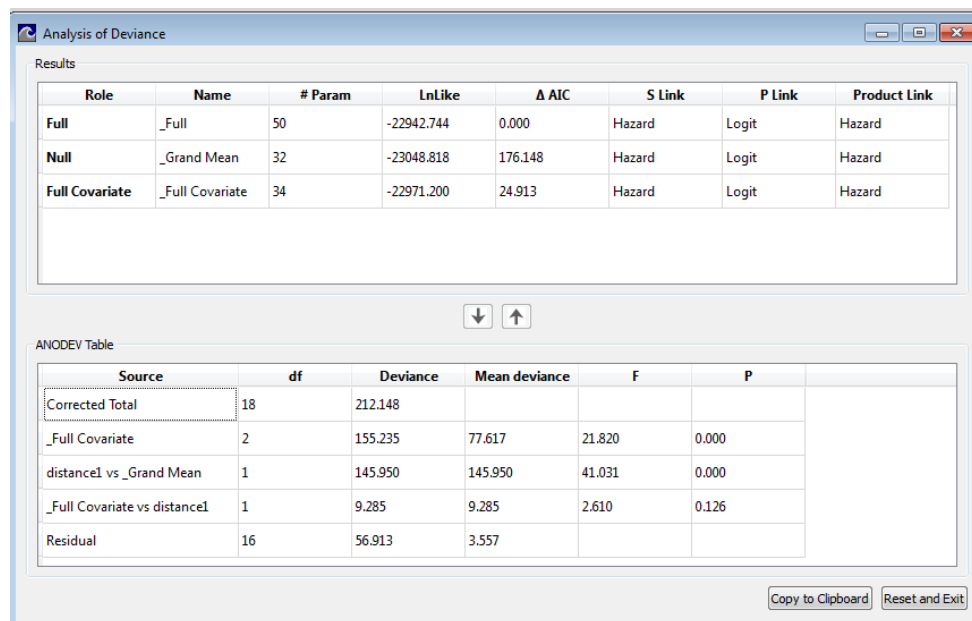


Figure 8.32. Final Analysis of Deviance results for the Migration Distance example.

A *P*-value of 0.000 for `distance1 vs _Grand Mean` shows that migration distance has a very significant effect on survival in the first reach. A *P*-value of 0.126 for `_Full Covariate vs distance1` shows that the migration distance does not have a significant effect on survival in the second reach.

8.3.2 Group Survival Plots

We can use the **Group Survival Plot** to visualize the effect distance has on survival. Make sure that the preferred result, `distance/distance1` is the active result, and double-click on **Group** under **Survival Plots** under **View Results** in the **Results** section. It brings up the plot shown in Figure 8.33.

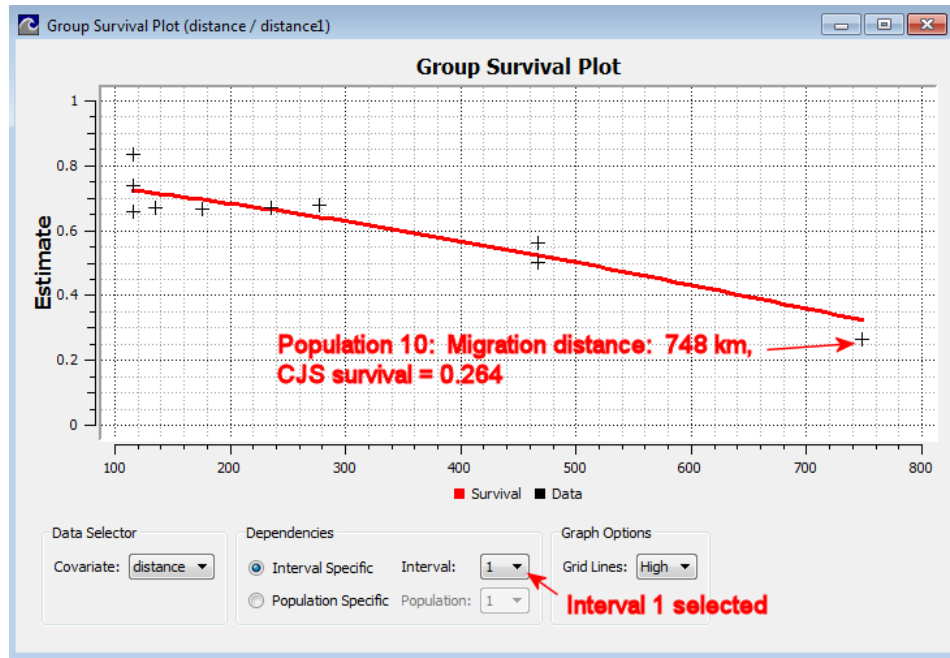


Figure 8.33. Survival as a function of migration distance in Interval 1 for the Migration Distance example.

The X-axis is migration distance and the Y-axis is the survival probability. The +s on the graph are the survival estimates for each of the 10 populations, using the CJS model. In Figure 8.33, the data point for population 10 is pointed out as an example.

The red curve is the modeled survival probability as a function of covariate. The agreement between the modeled survival curve and the CJS estimates indicates that the model fits the data adequately.

8.4 Relationship of Length to Survival

This example illustrates the process of modeling survival as a function of an individual covariate.

In 1993, a study was conducted in the Snake River to study the relationship of body length to survival. A total of 1,405 fish were captured, their lengths were recorded, and they were PIT-tagged and released above the Lower Granite Dam. The file “reservoir.dat” contains the data for this study.

8.4.1 Initial Data Analysis

When a dataset with individual covariate data is loaded into SURPH, an **Individual Covariate Graph** is provided for exploring the individual covariate data, shown in Figure 8.34.

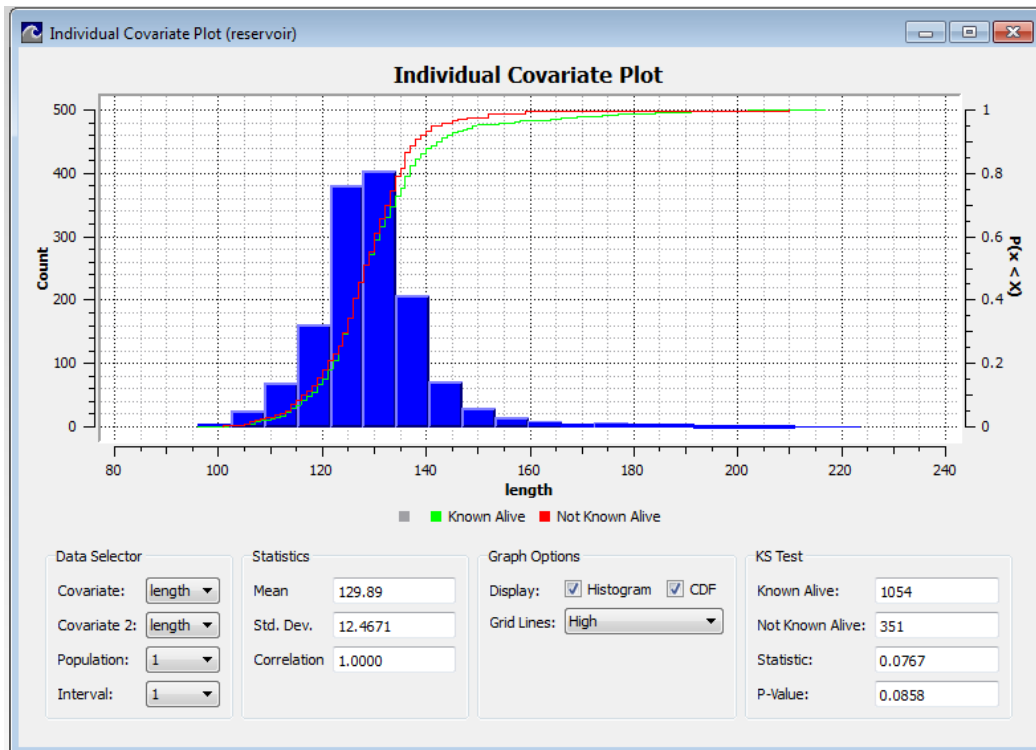


Figure 8.34. Individual Covariate Plot for fish length for the Relationship-of-Length-to-Survival example

The Kolmogorov-Smirnov (KS) test shows a P -value of 0.09, which is marginally significant, showing that individuals known to be alive in interval 1 may be larger than those not known to be alive. This may be because length affects the detection process and/or the survival process.

Using the two-stage modeling approach will allow us to first model detection as a function of length independent of its effect on survival. Once the detection modeling is done, we can then model survival as a function of length. Modeling the detection probabilities first will prevent the effects of length on detection from being mistakenly attributed to the effects of survival.

8.4.2 Modeling the Detection Probabilities

Because there is only one population in this study, there are only two parameterizations for the detection main effects: a unique detection probability for all intervals or a common detection probability. The models are created by using the **Main Effects** dialog to define the parameterization as described in previous examples and then creating the models using **Create Model** under **Detection-Only Modeling**. The interval-specific model is named **Full detection**, and the model with common detection probabilities is named **Common Detection**.

After estimating the parameters, we can bring up the **Likelihood Ratio Test** dialog as shown in Figure 8.35.

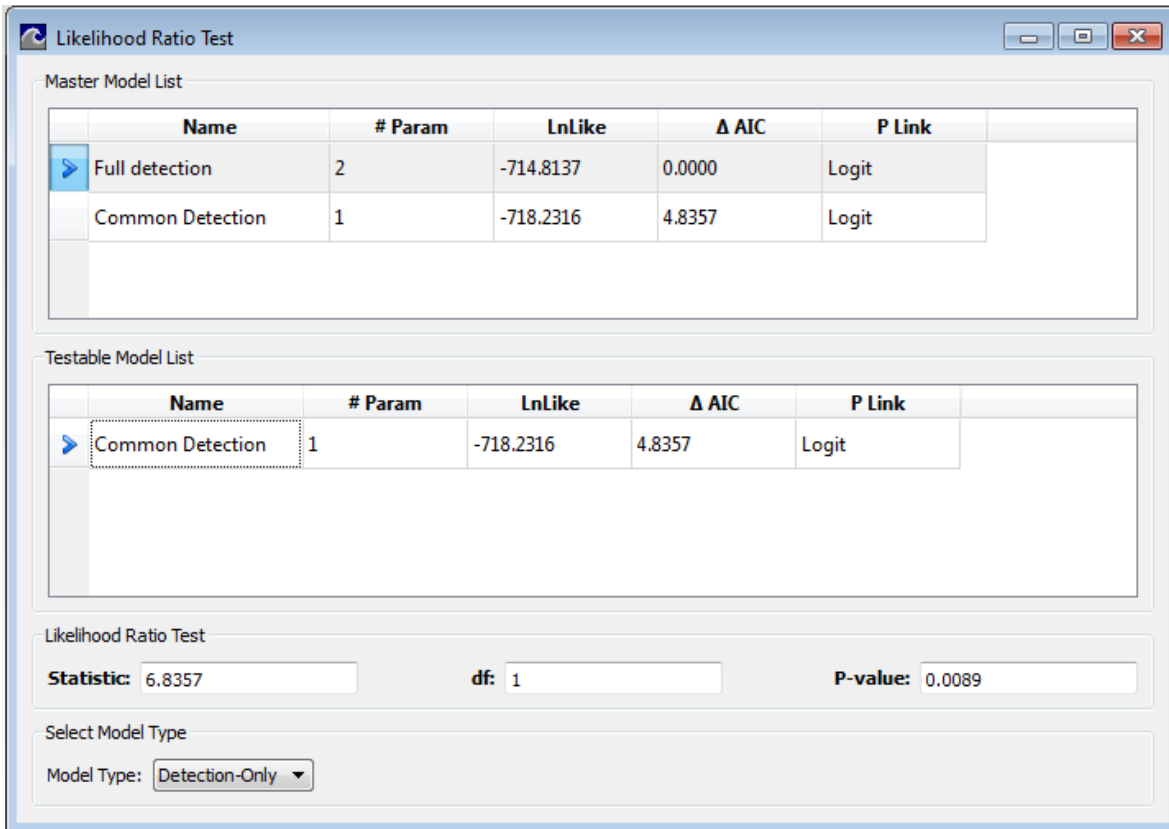


Figure 8.35. **Likelihood Ratio Test** dialog for comparing the detection models for the Relationship-of-Length-to-Survival example.

When comparing the full model with the common model, the P -value is less than 0.01, indicating that the full model is a much better fit.

We next explore whether `length` has a significant effect on detection, by opening the **Individual Effects** dialog for detection. Ensuring that the model named `Full detection` is the active model before proceeding, we create two models: (1) the model parameterized in Figure 8.36 that assumes `length` affects detection in the same way in both intervals and (2) the model parameterized in Figure 8.37 that assumes `length` affects detection differently in each interval.

Call the first model `Common length det.` and the second `Unique length det.`

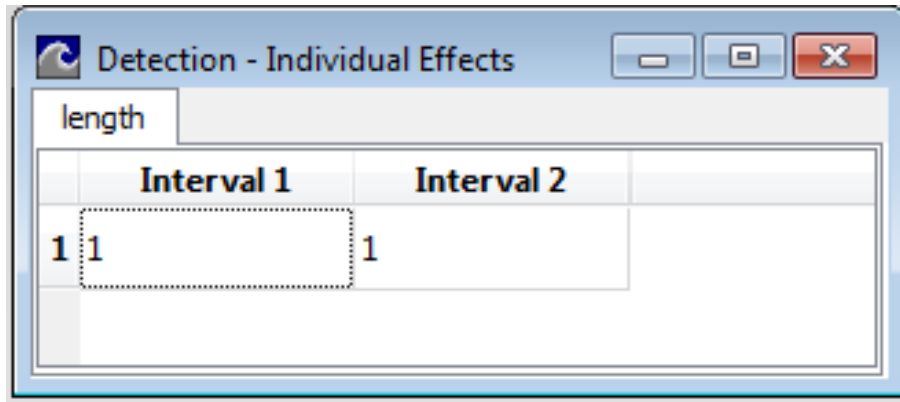


Figure 8.36. Parameterization assuming that `length` affects detection the same way in both intervals for the Relationship-of-Length-to-Survival example.

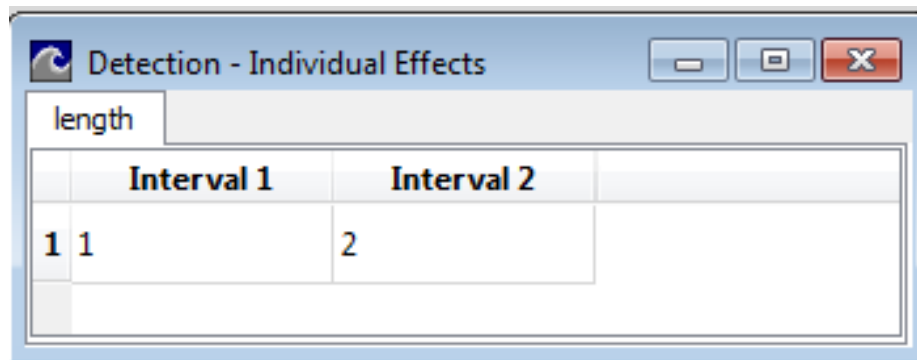


Figure 8.37. Parameterization assuming that `length` affects detection in a unique way in each interval for the Relationship-of-Length-to-Survival example.

We then estimate the parameters and return to the **Likelihood Ratio Test** dialog. Testing either model against the **Full** detection model fails to show a significant P -value. We thus conclude that `length` has no significant effect on detection and conclude that the **Full** detection model is the best model for the detection probabilities. We thus lock that model and proceed to the modeling of the survival probabilities.

The **Navigation Panel** should now appear as shown in Figure 8.38.

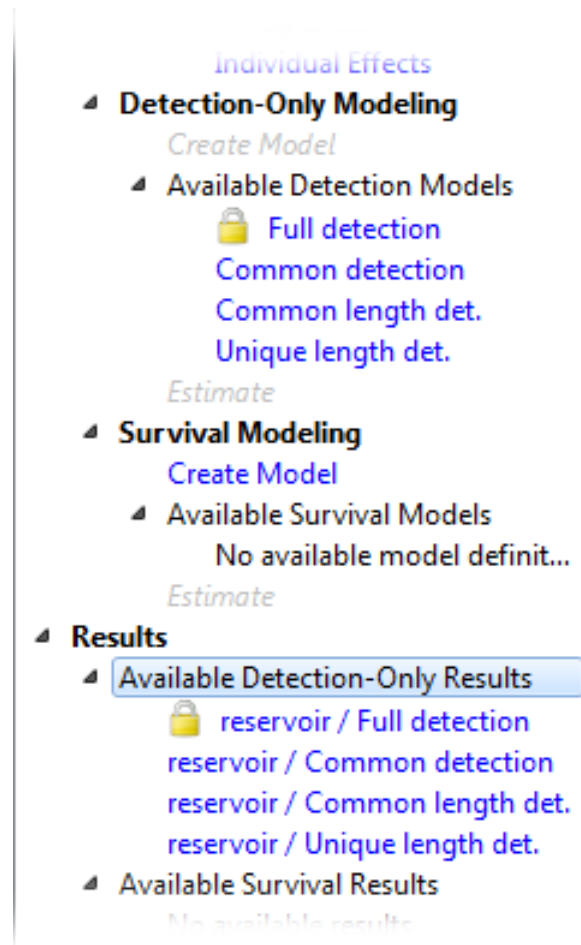


Figure 8.38. **Navigation Panel** after locking the **Fu11** detection model for the Relationship-of-Length-to-Survival example.

8.4.3 Modeling the Survival Probabilities

As with the detection probabilities, having one population allows for only two possibilities for modeling the main effects. First, set the survival main affects to **Interval Specific** and create the survival model named **Fu11**. Then set the main effects to **Common** and create the survival model named **Common**. We can estimate the parameters and then use the Likelihood Ratio Test to compare the models as shown in Figure 8.39.

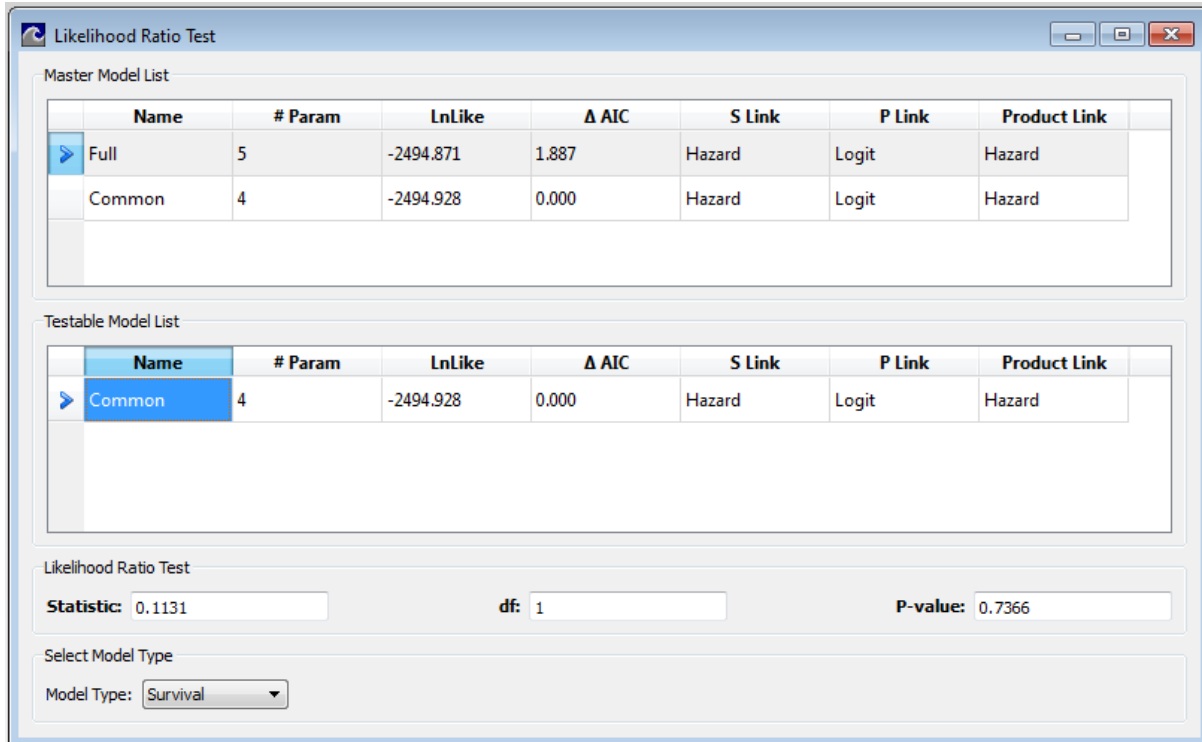


Figure 8.39. **Likelihood Ratio Test** dialog for comparing the survival main effects parameterization for the Relationship-of-Length-to-Survival example.

The P -value indicates that the **Full** model is no better than the **Common** model. We can thus use the **Common** model as the base for modeling the individual covariate effects.

Ensuring that the **Common** model is active, we can model the survival individual covariate effects as we did with the detection probabilities. We create a model named **Common length** that assumes that **length** affects survival the same way in each interval, and another model called **Unique length** that assumes a unique **length** parameter for each interval. We estimate the parameters and return to the **Likelihood Ratio Test** dialog.

The **Likelihood Ratio Test** dialog with the new results is shown in Figure 8.40.

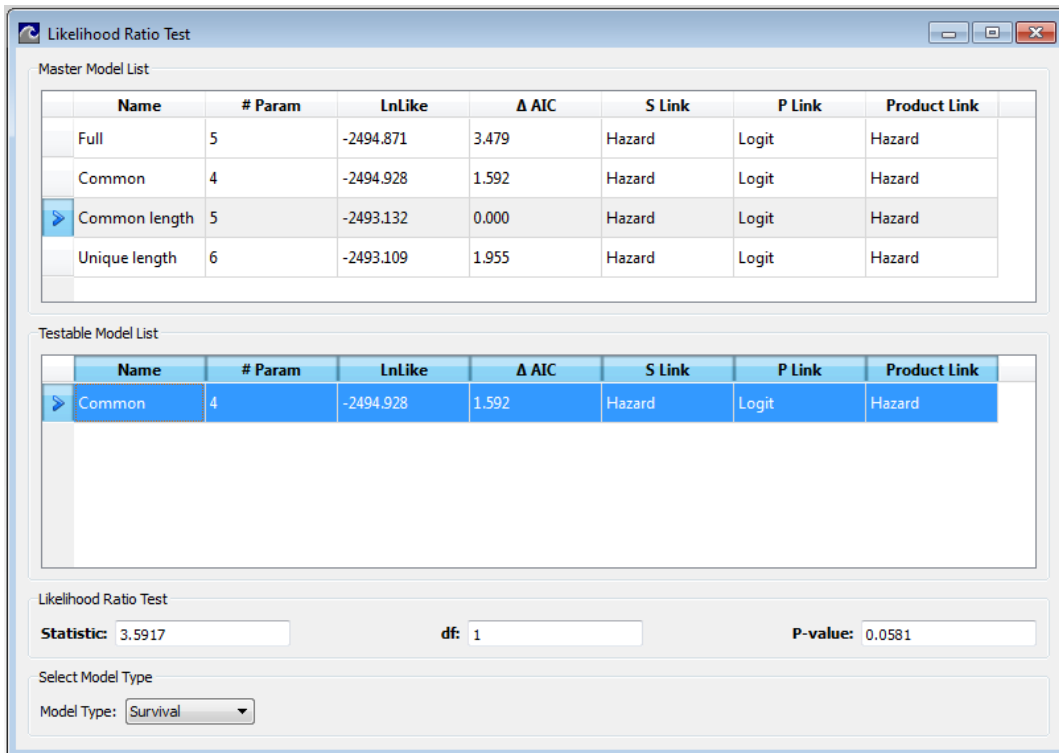


Figure 8.40. **Likelihood Ratio Test** dialog for results that model survival as a function of **length** for the Relationship-of-Length-to-Survival example.

Comparing the **Common length** result with the **Common** result yields a P -value of <0.06 , indicating a significant result at the 0.10 level. In comparing the **Unique length** result to either the **Common** result or the **Common length** result yields an insignificant P -value, indicating that the model that uses a common effect of **length** on both intervals is the best model. Making sure that the result named **reservoir/Common length det.** is the active result, we can look at the **Individual Covariate Survival Plot** for a visual representation of survival as a function of **length**.

8.4.4. Individual Covariate Plot

Double-clicking on **Individual** under **Survival Plots** under **View Results** brings up the individual covariate plot as shown in Figure 8.41.

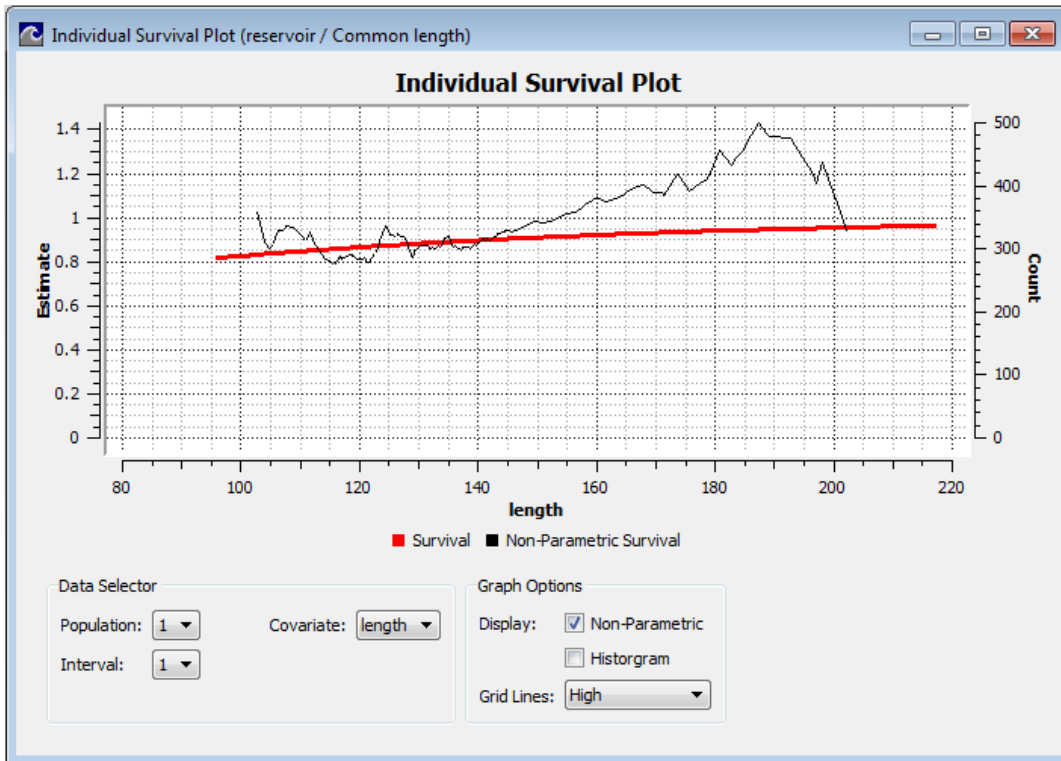


Figure 8.41. **Individual Covariate Plot** showing survival as a function of **length** in Interval 1 for the Relationship-of-Length-to-Survival example.

The smooth line represents survival modeled as a function of **length** for Interval 1, as defined by the currently selected model, and the more jagged line is a nonparametric estimate of survival as a function of **length**.

The parametric estimate appears to conform well with the nonparametric estimate for lengths less than about 150 mm but does not conform as well for larger individuals.

If we select **Display Histogram**, a histogram of the individuals at risk in Interval 1 is superimposed on the plot as shown in Figure 8.42.

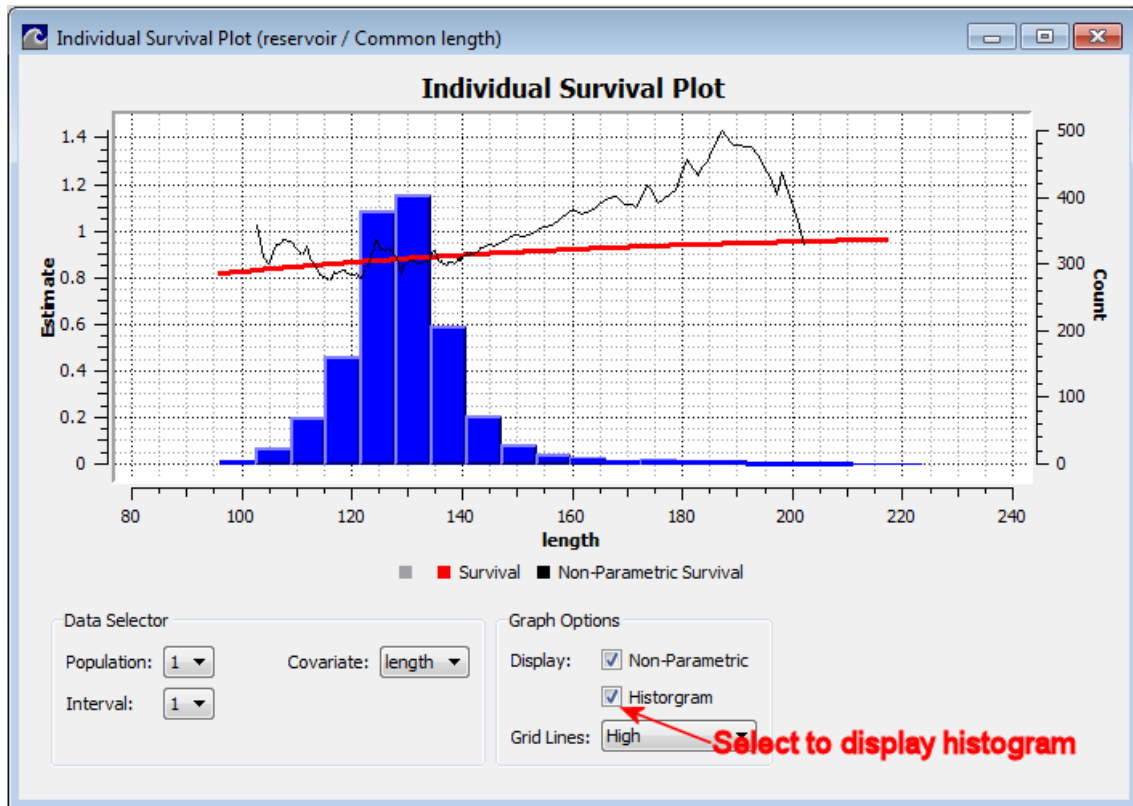


Figure 8.42. **Individual Covariate Plot** of survival as a function of **length** with a histogram of the observed lengths for the Relationship-of-Length-to-Survival example.

Figure 8.42 shows that where most of the individuals' lengths are, the parametric function conforms well to the nonparametric function. However, where the parametric functions conforms less well, there are very few individuals.

Figure 8.43 shows the same plot for the survival in Interval 2. In this case, the fit of the parametric function to the nonparametric is not as good.

More modeling is therefore indicated.

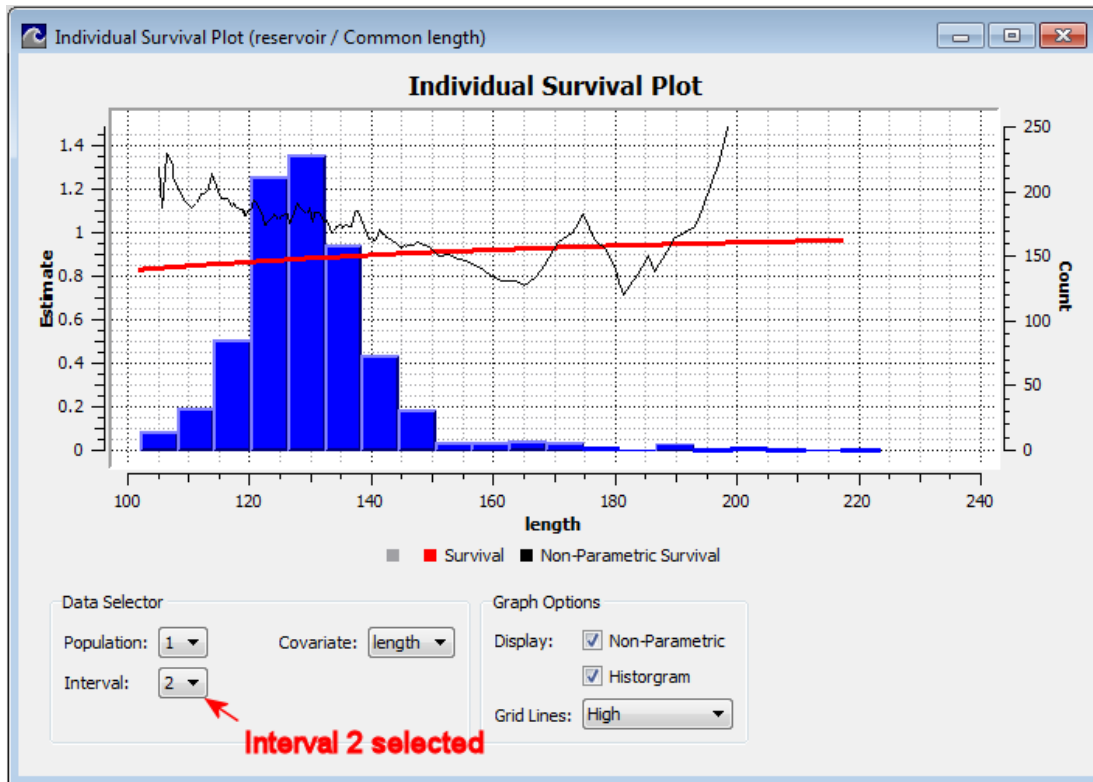


Figure 8.43. Survival as a function of `length` for Interval 2 for the Relationship-of-Length-to-Survival example.

8.5 Literature Cited

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* 5.

Appendix A: SURPH Input Data File

This Appendix describes the input data file for SURPH version 3.5. This is a different format than for previous versions of SURPH. SURPH 3.5 will still read input data files created for previous versions.

Figure A.1 shows the top portion of a sample SURPH input data file. The key parts are as follows:

- The **Surph3** keyword, which must be by itself on the first non-comment line of the data file, followed by a one-line description (required).
- The number of populations indicated by the keyword **npop** (required).
- The number of intervals indicated by the keyword **nint** (required).
- A group covariate declaration followed by its name (**middle**, in this case) and then followed by the values for the group covariate. The group covariate values must be numbers, and there can be either:
 - one value per population, in which case the same value is used for all intervals of a population, or
 - one value per population per interval—all values for the first population are listed first from the first interval to the last, followed by the values for the second population, and so on.
- A second group covariate declaration for the group covariate named **late**.
- An individual covariate declaration for the individual covariate **length**. The values for the individual covariates follow the capture histories. In the example in Figure A.1, the first individual of population **early1** has a **length** of 118.

- The **population** keyword followed by the name of the population. The population declaration is followed by all capture histories for the population. A new population is declared with another line with the **population** keyword followed by a name. The capture histories can begin with an optional tag ID as shown in Figure 1; if the first field is not a 0 or a 1, it is assumed to be a tag ID. If there are individual covariates, the covariate values for the current individual follow the capture history in the order that they were declared.

```

Surph3
Dworshak Hatchery 4/8/93, 4/22/93, and 5/6/93 Releases
npop 9
nint 3

gcov middle
0 0 0 1 1 1 0 0 0

gcov late
0 0 0 0 0 0 1 1 1

icov length

population early1
7F7B011B19 1 0 0 0 118
7F7B01341A 1 0 0 0 138
7F7B013E69 1 0 0 0 115
7F7B013F49 1 0 0 0 119

```

Figure A.1. Sample SURPH 3.5 input data file with two group covariates and one individual covariate.

Figure A.2 shows the top portion of a SURPH input file with multiple individual covariates. The order of the individual covariate values for each individual must match the order in which they are declared. In Figure A.2, the first individual of population **one** is assumed to have **length** of 100, **weight** of 10.8, and **condix** of 0.0108.

```

Surph3
Yakima: Lake Releases

npop 3
nint 2

icov length
icov weight
icov condix

population one
1 1 1 100 10.8 0.0108
1 1 0 116 19.4 0.0124287588667022
1 1 1 97 10.1 0.01106639508345267
1 1 0 115 18.8 0.01236130516972138
1 1 0 98 11.4 0.01211230014704757

```

Figure A.2. Sample SURPH 3.5 input data file with three individual covariates.

Figure A.3 shows an example with a count after each capture history instead of specifying one individual's capture history per line. For population **treatment**, there were 29,925 individuals with capture history "1 0 0 0 0 0," 590 individuals with capture history "1 0 0 0 0 1," and so on. Clearly this method cannot be used when there are individual covariates.

```

Surph3
Example Data Set from Burnham "Blue Book"

npop 2
nint 5

population treatment
1 0 0 0 0 0 : 25925
1 0 0 0 0 1 : 590
1 0 0 0 1 0 : 508
1 0 0 0 1 1 : 18
1 0 0 1 0 0 : 1500
1 0 0 1 0 1 : 48
1 0 0 1 1 0 : 37

```

Figure A.3. Sample SURPH 3.5 input data file with counts per capture history specified.

Figure A.4 shows a portion of an input data file showing how multiple populations are specified. All capture histories above the population `control` line belong to the previous population. All capture histories after it until the next **population** keyword or end-of-file belong to population `control`.

```
1 1 1 2 0 0 : 0
1 0 2 0 0 0 : 14
1 1 2 0 0 0 : 0
1 2 0 0 0 0 : 29

population control
1 0 0 0 0 0 : 24605
1 0 0 0 0 1 : 641
1 0 0 0 1 0 : 522
1 0 0 0 1 1 : 24
1 0 0 1 0 0 : 1678
```

Figure A.4. Portion of a SURPH 3.5 input data file showing how multiple populations are specified.

Capture Histories

A capture history is a string of “0s,” “1s,” and “2s,” representing the detection history for a single individual, as follows:

- “1” represents a detection and rerelease. The occasion of initial release is also represented by a “1,”
- “2” represents a detection and removal (censoring), and
- “0” represents a nondetection.

There must be one capture history field for each detection site including the initial release site. For example, in a study with three intervals:

- “1 1 0 1” would represent an individual released at the initial release, detected and re-released at event 1, and next detected at event 3, and
- “0 1 2 0” represents an individual released for the first time at the first resampling event, then detected and removed at the second resampling event.

Appendix B: SURPH

Statistical Theory

B.1 Release-Recapture Theory

B.1.1 Definition

Figure B.1 illustrates a release-recapture study. There is an initial release event where animals from the population of interest are captured, given uniquely identifying tags, and released. This is followed by subsequent sampling events where animals are captured, the tag IDs of previously tagged individuals are recorded, and the individuals are re-released. Optionally, previously untagged individuals may be tagged and released at each sampling event.

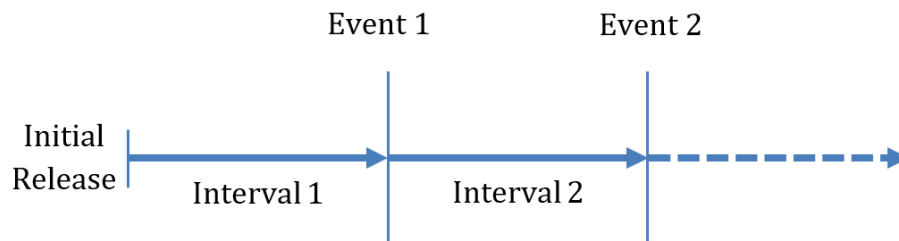


Figure B.1. Schematic of release-recapture study.

There are two basic types of release-recapture studies: a time-oriented study and a distance-oriented study.

In a time-oriented study, the sampling events all occur at the same general location at regular time intervals and the interval between events are periods of time.

A distance-oriented study refers to studies of migrating fish in a river system such as salmon in the Columbia Basin River system. There is an initial release event, and the

subsequent sampling events are locations downstream along the river of interest. In this case, the intervals between events are distances along the river.

The sampling events after the initial event can be active events, where the animals are physically recaptured, examined, and re-released, or they can be passive events where the animals are simply observed rather than physically recaptured.

B.1.2 An Intuitive Derivation of the Release-Recapture Parameters

In a release-recapture study, the parameter of interest is survival over time or distance. Figure B.2 shows a hypothetical study with one sampling event after the initial release event with 100% detection probability at Event 1. Ten individuals are released, and 3 of these are detected at Event 1. We can thus estimate survival (S) from Event 0 to Event 1 as

$$\hat{S} = \frac{m}{R} = \frac{3}{10} = 0.3 \quad (\text{B.1})$$

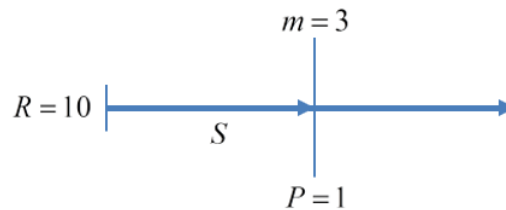


Figure B.2. A hypothetical study with perfect detection probability at event 1.

In reality, we do not have 100% detection probability at the sampling events. Figure B.3 illustrates a hypothetical study with two sampling events after the initial event. The detection probabilities are represented as P_1 and P_2 . In this case, we can start with the simple survival estimate in Equation (B.1) and adjust it for the estimate of detection probability \hat{P}_1 .

$$\hat{S}_1 = \left(\frac{m_1}{R}\right) \left(\frac{1}{\hat{P}_1}\right) \quad (\text{B.2})$$

In order to estimate P_1 , we can look at the detections at Event 2. An estimate of P_1 can be constructed as a ratio of those detected at both events 1 and 2 to those detected only at Event 2:

$$\hat{P}_1 = \frac{m_{12}}{m_2} = \frac{2}{5} = 0.4 \quad (\text{B.3})$$

The estimate of survival from initial release to Event 1 then becomes:

$$\hat{S}_1 = \frac{0.3}{0.4} = 0.75 \quad (\text{B.4})$$

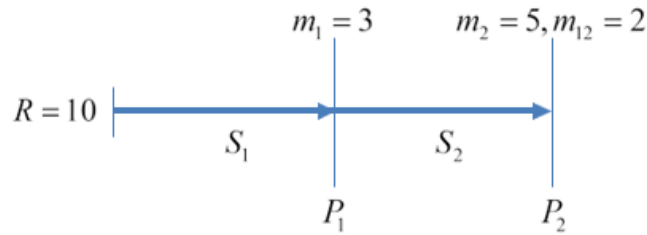


Figure B.3. A hypothetical study with two sampling events after the initial event. m_1 = detections at Event 1, m_2 = detections at Event 2, and m_{12} = detections at both events 1 and 2.

This illustrates that in order to estimate survival for a given period, we need one detection event beyond the final period of interest; for the final period of a release-recapture study, mortality cannot be distinguished from nondetection, and we can only estimate the product of detection and survival (S_2P_2).

B.1.3 Multinomial Likelihood Model

A multinomial likelihood has the form

$$L \propto \prod_{i=1}^I \theta_i^{n_i},$$

where

I = number of cells,

θ_i = the probability of cell i , and

n_i = the number of observations from cell i .

The cells must be mutually exclusive and exhaustive, i.e., each observation must fall into one and only one cell, and the cell probabilities must sum to 1.0. In this section, we construct the cell probabilities for a release-recapture study.

Figure B.4 shows a release-recapture study with two sampling events after the initial release event, along with the estimable parameters. Table B.1 gives some corresponding hypothetical counts. 100 individuals are released at the initial event, 30 of which are next detected at Event 1. Five are removed (say, because of handling mortality) and 25 are re-released. Of the 100 in the initial release, 50 are next recaptured at Event 2 (i.e., without detection at Event 1), and the remaining 20 are never recaptured. Of the 25 re-released at Event 1, 20 are captured at Event 2 and the remaining 5 are never recaptured.

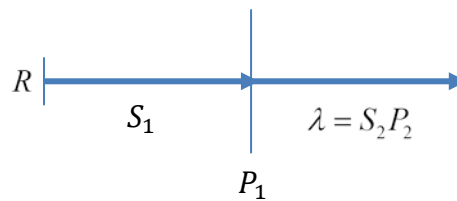


Figure B.4. A release-recapture study with two detection sites.

Table B.1. Sample data (M-array) for release-recapture study with two periods.

Event	Total Released	Event Next Recaptured 1	Event Next Recaptured 2	Never Recaptured
Initial	100	30:5	50	20
1	25		20	5

The cell probabilities and corresponding counts for a multinomial model corresponding to the initial release site are shown in Table B.2. In a multinomial model, the cells, represented by rows in Table B.2, must be mutually exclusive and exhaustive. Hence, the cell probabilities must sum to 1.0, and the cell counts sum to the total release size for initial release. Note that for the first row, the probability for individuals never recaptured after initial release takes into account the possibility of not surviving the first period $(1 - S_1)$ and the possibility of surviving but not being detected at Event 1, and not surviving and being recaptured at the final event $(S_1(1 - P_1)(1 - \lambda))$.

Table B.2. Hypothetical Multinomial cell probabilities and counts for the initial release event in a two-period release-recapture study.

Next Recaptured	Probability	Counts
Never	$1 - S_1 + S_1(1 - P_1)(1 - \lambda)$	$n_{00} = 20$
Event 1	S_1P_1	$n_1 = 30$
Event 2	$S_1(1 - P_1)\lambda$	$n_{01} = 50$

The likelihood for individuals released at the initial event is

$$L_0 = [1 - S_1 + S_1(1 - P_1)(1 - \lambda)]^{n_{00}} [S_1P_1]^{n_1} [S_1(1 - P_1)\lambda]^{n_{01}}.$$

In the same way, a likelihood can be constructed for those released at Event 1. Table B.3 shows the cell probabilities and counts.

Table B.3. Hypothetical multinomial cell probabilities and counts for Event 1 in a two-period release-recapture study.

Next Recaptured	Probability	Counts
-----------------	-------------	--------

Never	$1 - \lambda$	$n_{10} = 5$
Event 2	λ	$n_{12} = 20$

The likelihood for individuals release at Event 1 is

$$L_1 = \lambda^{n_{12}}(1 - \lambda)^{n_{10}}.$$

The full likelihood for the three-period release-recapture study is

$$L = L_0 \times L_1.$$

Thus, for a release-recapture study with K periods, the likelihood is

$$L = L_0 \times L_1 \times \dots \times L_{K-1}. \quad (\text{B.5})$$

B.1.4 Multinomial Likelihood for a Multiple-Population Study

For a release-recapture study with I populations and K sampling events after the initial release event, the parameters to be estimated are:

S_{ik} = the probability of survival to sampling event k for population i , conditional on survival to the event $k - 1$, ($i = 1, \dots, I; k = 1, \dots, K - 1$),

P_{ik} = the probability of detection at event k for population i , conditional on being alive at event k , ($i = 1, \dots, I; k = 1, \dots, K - 1$), and

λ_i = the probability of survival to and being detected at event K for population i , conditional on being alive at event $K - 1$, ($i = 1, \dots, I$).

For clarity, we define the following:

χ_{ik} = the probability of an individual of population i , not being seen after event k , conditional on being alive at event k ,

$$= 1 - S_{i,k+1} + S_{i,k+1}(1 - P_{i,k+1})\chi_{i,k+1}, \quad (i = 1, \dots, I; k = 0, \dots, K - 2), \text{ and}$$

$$\chi_{i,K-1} = 1 - \lambda_i.$$

The data can be summarized as follows.

m_{ik} = the number of previously tagged individuals of population i , recaptured at event k ($i = 1, \dots, I; k = 1, \dots, K - 1$),

R_{ik} = the number of tagged individuals of population i “released” at event k ($i = 1, \dots, I; k = 1, \dots, k - 1$), including both those that are initially tagged and released and those previously released that are captured and re-released,

r_{ik} = the number R_{ik} that are detected at least at one subsequent event, and

z_{ik} = the number of individuals of population i that were released prior to event k , but were not detected at event k ($i = 1, \dots, I; k = 1, \dots, K - 1$).

The likelihood of Equation (B.5) can now be rewritten as

$$L(\vec{S}, \vec{P}, \vec{\lambda}) \propto \prod_{i=1}^I \left\{ \left(\prod_{k=1}^{K-1} S_{ik}^{m_{ik}+z_{ik}} P_{ik}^{m_{ik}} (1 - P_{ik})^{z_{ik}} \right) \left(\prod_{k=0}^{K-1} \chi_{ik} \right) \lambda^{r_{i,K-1}+z_{i,K-1}} \right\} \quad (\text{B.6})$$

B.1.5 Model Assumptions

For the multinomial likelihood to be valid, the following assumptions must be met.

- Individuals marked for the study are a representative sample from the population of interest.
- Survival and capture probabilities are not affected by tagging or resampling. That is, tagged animals have the same probabilities as untagged animals.
- All sampling events are “instantaneous.” That is, sampling takes a negligible amount of time or distance relative to the length of the intervals between sampling events.
- The fate of each tagged individual is independent of the fate of all others.
- All tagged individuals alive at the beginning of a sampling interval have the same probability of surviving until the end of that interval.

- All tagged individuals alive at the time or location of a sampling event have the same probability of being detected at that event.
- Each individual detected at a particular sampling event has the same probability of being removed at that event, and the probability of removal (censoring) is independent of the survival process.

B.1.6 Cormack-Jolly-Seber Parameter Estimates

Cormack (1964), Jolly (1965), and Seber (1965) developed closed-form estimates of the parameters of the likelihood in Equation (B.6). The following derivations of the Cormack-Jolly-Seber (CJS) estimates are based on Burnham et al. (1987). For clarity of notation, the

The following are defined for convenience:

A_k = defined as r_k/R_k for $k = 0, \dots, K - 1$, defined as 1.0 for $k = K$,

T_k = the number known alive at occasion $k = 1, \dots, K - 1$,

$T_k = m_k + z_k$, and

B_k = defined as m_k/T_k for $k = 0, \dots, K - 1$, defined as 1.0 for $k = K$.

The CJS estimates can now be defined as follows:

$$\hat{S}_k = A_{k-1} \left(B_k + \frac{1 - B_k}{A_k} \right), k = 1, \dots, K - 1 \quad (\text{B.7})$$

$$\hat{P}_k = \frac{B_k}{B_k + \frac{1 - B_k}{A_k}}, k = 1, \dots, K - 1 \quad (\text{B.8})$$

$$\hat{\lambda} = \frac{r_{K-1}}{R_{K-1}} \quad (\text{B.9})$$

The non-zero variance and covariances are as follows.

$$\text{Var}(\hat{S}_k) = \hat{S}_k^2 \left[\left(\frac{1}{r_{K-1}} - \frac{1}{R_{K-1}} \right) + (1 - p_k)^2 \left(\frac{1}{r_k} - \frac{1}{R_k} \right) + p_k(1 - p_k) \frac{(1 - A_k)^2}{A_k T_k} \right], \quad (\text{B. 10})$$

$$k = 1, \dots, K - 1.$$

If m_k or z_k are 0, then $\widehat{\text{Var}}(\hat{P}_k) = 0$. Otherwise,

$$\widehat{\text{Var}}(\hat{P}_k) = \left(\hat{P}_k(1 - \hat{P}_k) \right)^2 \left(\frac{1}{r_k} - \frac{1}{R_k} + \frac{1}{m_k} + \frac{1}{z_k} \right), k = 1, \dots, K - 1 \quad (\text{B. 11})$$

$$\widehat{\text{Var}}(\hat{\lambda}) = \hat{\lambda}^2 \left(\frac{1}{r_{K-1}} - \frac{1}{R_{K-1}} \right) \quad (\text{B. 12})$$

$$\widehat{\text{Cov}}(\hat{S}_k, \hat{S}_{k+1}) = -\hat{S}_k \hat{S}_{k+1} (1 - \hat{P}_{K-1}) \left(\frac{1}{r_{K-1}} - \frac{1}{R_{K-1}} \right) \quad (\text{B. 13})$$

$$\widehat{\text{Cov}}(\hat{S}_{K-1}, \hat{\lambda}) = -\hat{S}_{K-1} \hat{\lambda} (1 - \hat{P}_{K-1}) \left(\frac{1}{r_{K-1}} - \frac{1}{R_{K-1}} \right) \quad (\text{B. 14})$$

$$\widehat{\text{Cov}}(\hat{S}_k, \hat{P}_{K-1}) = \hat{S}_k \hat{P}_{K-1} (1 - \hat{P}_{K-1}) \left(\frac{1}{r_{K-1}} - \frac{1}{R_{K-1}} \right), k = 2, \dots, K - 1 \quad (\text{B. 15})$$

If $z_k = 0$, then $\widehat{\text{Cov}}(\hat{S}_k, \hat{P}_k) = 0$. Otherwise,

$$\widehat{\text{Cov}}(\hat{S}_k, \hat{P}_k) = -\hat{S}_k \hat{P}_k (1 - \hat{P}_k)^2 \times \dots \times \left[\left(\frac{1}{r_k} - \frac{1}{R_k} \right) + \left(1 - \frac{r_k}{R_k} \right) \left(\frac{1}{z_k} \right) \right], \quad (\text{B. 16})$$

$$k = 1, \dots, K - 1$$

$$\widehat{\text{Cov}}(\hat{\lambda}, \hat{P}_{K-1}) = \hat{\lambda} \hat{P}_{K-1} (1 - \hat{P}_{K-1})^2 \left[\left(\frac{1}{r_{K-1}} - \frac{1}{R_{K-1}} \right) \right] \quad (\text{B. 17})$$

All other covariances are zero.

B.2 Modeling Survival Main Effects

In this section, the discussion focuses on modeling the survival probabilities (S_{ik}) of

likelihood Equation (B.6). Detection probabilities are modeled in the same way.

B.2.1 Modeling Main Effects

In SURPH 3, the survival probabilities (S_{ik}) are modeled as a function of population effects and interval effects as follows:

$$S_{ik} = f(S_0, \pi_{ik}), \quad (\text{B.18})$$

where

$f(\)$ = the link function,

S_0 = the baseline survival, defined as the survival parameter for population 1, interval 1, and

π_{ik} = the main effect for population i , interval k .

The following sections illustrate how these parameters are used to create survival estimates for the fully parameterized model, the common model, the population-effects model, and the interval-effects model, and provide an example of a custom model that assumes interactions between population effects and interval effects.

B.2.1.1 Fully Parameterized Model

The fully parameterized model is equivalent to the CJS model, with unique estimates of survival for each population and interval. It is parameterized as shown in Table B.4.

Table B.4. Parameterization of the fully parameterized model for survival main effects with K intervals and I populations.

Population	Interval 1	Interval 2	...	Interval $K - 1$
1	$f(S_0)$	$f(S_0, \pi_{12})$		$f(S_0, \pi_{1, K-1})$
2	$f(S_0, \pi_{21})$	$f(S_0, \pi_{22})$		$f(S_0, \pi_{2, K-1})$
⋮				
I	$f(S_0, \pi_{I1})$	$f(S_0, \pi_{I2})$		$f(S_0, \pi_{I, K-1})$

The parameters to be estimated are

- S_0 ,
- π_{1k} , $k = 2, \dots, K - 1$, and
- π_{ik} , $i = 2, \dots, I, k = 1, \dots, K - 1$,

for a total of $I(K - 1)$ survival-related parameters, the same as the CJS model.

B.2.1.2 Common Model

The common model has a common parameter across all populations and intervals, as shown in Table B.5.

Table B.5. Parameterization of the common model for survival main effects with K intervals and I populations.

Population	Interval 1	Interval 2	...	Interval $K - 1$
1	$f(S_0)$	$f(S_0)$		$f(S_0)$
2	$f(S_0)$	$f(S_0)$		$f(S_0)$
⋮				
I	$f(S_0)$	$f(S_0)$...	$f(S_0)$

There is one parameter to be estimated with the common model: S_0 .

B.2.1.3 Population-Effects Model

If we assume that survival differs across populations but not from one interval to the next, the survival probabilities can be parameterized as shown in Table B.6.

Table B.6. Parameterization of the population-effects model for survival main effects with K intervals and I populations.

Population	Interval 1	Interval 2	...	Interval $K - 1$
1	$f(S_0)$	$f(S_0)$		$f(S_0)$
2	$f(S_{0,\pi_2})$	$f(S_{0,\pi_2})$		$f(S_{0,\pi_2})$
⋮				
I	$f(S_{0,\pi_I})$	$f(S_{0,\pi_I})$...	$f(S_{0,\pi_I})$

The parameters to be estimated are

- S_0 , and
- $\pi_i, i = 2, \dots, I$,

for a total of I survival-related parameters.

B.2.1.4 Interval-Effects Model

The interval-effects model assumes unique survival probabilities from one interval to the next, but the survival probabilities do not vary across populations. The interval-effects model is parameterized as shown in Table B.7.

Table B.7. Parameterization of the interval-effects model for survival main effects with K intervals and I populations.

Population	Interval 1	Interval 2	...	Interval $K - 1$
1	$f(S_0)$	$f(S_{0,\pi_2})$		$f(S_{0,\pi_{K-1}})$
2	$f(S_0)$	$f(S_{0,\pi_2})$		$f(S_{0,\pi_{K-1}})$
⋮				
I	$f(S_0)$	$f(S_{0,\pi_2})$...	$f(S_{0,\pi_{K-1}})$

The parameters to be estimated are:

- S_0 , and
- $\pi_k, k = 1, \dots, K - 1$,

for a total of $K - 1$ survival-related parameters.

B.2.1.5 Custom Model with Interaction Between Population Effects and Interval Effects

To demonstrate the flexibility of main-effects modeling in SURPH 3, we look at a hypothetical study with 3 intervals and 10 populations. We assume that the first 3 populations share common survival probabilities, as do populations 4 through 6, and populations 7 through 10. Furthermore, we assume that survival differs across intervals for populations 1 through 3, and populations 7 through 10; for populations 4 through 6, a common survival probability is used across intervals.

Table B.8: Parameterization of a custom model for survival main effects with 3 intervals and 10 populations and an interaction between population effects and interval effects.

Population	Interval 1	Interval 2
1	$f(S_0)$	$f(S_{0,\pi_1})$
2	$f(S_0)$	$f(S_{0,\pi_1})$
3	$f(S_0)$	$f(S_{0,\pi_1})$
4	$f(S_{0,\pi_2})$	$f(S_{0,\pi_2})$
5	$f(S_{0,\pi_2})$	$f(S_{0,\pi_2})$
6	$f(S_{0,\pi_2})$	$f(S_{0,\pi_2})$
7	$f(S_{0,\pi_{31}})$	$f(S_{0,\pi_{32}})$
8	$f(S_{0,\pi_{31}})$	$f(S_{0,\pi_{32}})$
9	$f(S_{0,\pi_{31}})$	$f(S_{0,\pi_{32}})$
10	$f(S_{0,\pi_{31}})$	$f(S_{0,\pi_{32}})$

The parameters to be estimated are:

- S_0 , and
- $\pi_1, \pi_2, \pi_{31}, \pi_{32}$,

for a total of $K - 1$ survival-related parameters.

B.2.2 Modeling Group-Covariate Effects

In addition to modeling survival as a function of main effects, survival can be modeled as a function of group-covariate data. Group covariates are factors that apply to an entire population.

SURPH models survival as a function of a group covariate as follows:

$$S_{ik} = f(S_0, \pi_{ik} + \beta_{ik}x_{ik}),$$

where

β_{ik} = the group-covariate effect for population i , interval k ; and

x_{ik} = the value of the group covariate for population i , interval k .

It is possible to model survival as a function of multiple group covariates:

$$S_{ik} = f\left(S_0, \pi_{ik} + \sum_{m=1}^M \beta_{ikm}x_{ikm}\right), \quad (\text{B. 19})$$

where

M = the number of group covariates;

β_{ikm} = the group-covariate effect for population i , interval k , and the m th covariate;
and

x_{ikm} = the value of the m th group covariate for population i , interval k .

Note that if we start with the fully parameterized main-effects model, the model will be overparameterized if covariate effects are then added; there will be too many parameters to estimate from the data.

There are two standard ways of modeling group-covariate data: (1) an interval-specific, group-covariate model and (2) a population-specific, group-covariate model.

Note that SURPH modeling capability is flexible enough to allow variations to these two basic models, but it is important to remember that for the model to be valid, the number of survival parameters in the model must not exceed the number of survival parameters in the fully parameterized model: $I(K - 1)$ where I is the number of populations and K is the number of intervals.

B.2.2.1 Interval-Specific Group Covariate Model

With an interval-specific, group-covariate model, the survival main effects are parameterized for the interval-effects model (Table B.7), assuming common survival main effects across populations but unique effects from one interval to the next. The group-covariate effects are also assumed to be unique from one interval to the next and common across populations.

Table B.9 shows the interval-specific parameterization for the interval-specific, group-covariate model for a dataset with one covariate. This is the parameterization for the Full Covariate model in ANODEV described in Appendix B.5.

Population	Interval 1	Interval 2	...	Interval $K - 1$
1	$f(S_0, \beta_1 x_{11})$	$f(S_0, \pi_2 + \beta_2 x_{12})$		$f(S_0, \pi_{K-1} + \beta_{K-1} x_{1,K-1})$
2	$f(S_0, \beta_1 x_{21})$	$f(S_0, \pi_2 + \beta_2 x_{22})$		$f(S_0, \pi_{K-1} + \beta_{K-1} x_{2,K-1})$
□	□			
I	$f(S_0, \beta_1 x_{I1})$	$f(S_0, \pi_2 + \beta_2 x_{I2})$...	$f(S_0, \pi_{K-1} + \beta_{K-1} x_{I,K-1})$

where

π_k = the main effect parameter for interval k ,

β_k = the group covariate effect parameter for interval k , and

x_{ik} = the value of the group covariate for population i , interval k .

B.2.2.2 Population-Specific Group Covariate

An alternative to the interval-specific, group-covariate model is the population-specific, group-covariate model. In this case, common main effects are assumed across intervals, but unique main effects across populations. Similarly, the group covariate is assumed to affect each population in a unique way, but each interval in the same way. The parameterization is shown in Table B.10.

Table B.10: Parameterization of the population-specific, group-covariate model for survival with K intervals, I populations, and one group covariate.

Population	Interval 1	Interval 2	...	Interval $K - 1$
1	$f(S_0, \beta_1 x_{11})$	$f(S_0, \beta_1 x_{12})$		$f(S_0, \beta_1 x_{1,K-1})$
2	$f(S_0, \pi_2 + \beta_2 x_{21})$	$f(S_0, \pi_2 + \beta_2 x_{22})$		$f(S_0, \pi_2 + \beta_2 x_{2,K-1})$
□	□			
I	$f(S_0, \pi_I + \beta_I x_{I1})$	$f(S_0, \pi_I + \beta_I x_{I2})$...	$f(S_0, \pi_I + \beta_I x_{I,K-1})$

where

- π_i = the main effect parameter for population i ,
- β_i = the group covariate effect parameter for population i , and
- x_{ik} = the value of the group covariate for population i , interval k .

B.2.3 Modeling Individual-Covariate Effects

An individual covariate is measured at an individual level, such as body size. With individual covariate modeling, survival must be estimated for each individual within a population:

$$S_{ijk} = f(S_0, \pi_{ik} + \gamma_{ik} y_{ijk}), \quad (\text{B. 20})$$

where

S_{ijk} = the survival probability for the j th individual of population i in interval k ;

γ_{ik} = the individual covariate effect for population i , interval k ; and

y_{ijk} = the value of the individual covariate for the j th individual of population i , interval k .

As with group covariates, survival can be modeled as a function of multiple individual covariates. Also, survival can be modeled as a function of both group covariates and individual covariates. Thus, the most general formulation for the survival probability is as follows:

$$S_{ijk} = f \left(S_0, \pi_{ik} + \sum_{m=1}^M \beta_{ikm} x_{ikm} + \sum_{n=1}^N \gamma_{ikn} y_{ijkn} \right), \quad (\text{B. 21})$$

where

N = number of individual covariates;

γ_{ikn} = the individual covariate effect for covariate n , population i , interval k ; and

y_{ijkn} = the value of the individual covariate n for individual j , population i , interval k .

To estimate the survival parameters for a model with individual covariate effects, an individual-based likelihood must be used rather than the likelihood in Equation (B.6).

The likelihood for individual j of population i can be written as:

$$L_{ij}(\vec{S}_{ij}, \vec{P}_{ij}, \lambda_{ij}) = \left[\prod_{k=t_{0ij}+1}^{\min(t_{1ij}, K-1)} S_{ijk} P_{ijk}^{D_{ijk}} (1 - P_{ijk})^{1-D_{ijk}} \right] \chi_{ijt_{1ij}}^{1-C_{ij}} \lambda_{ij}^{D_{ijk}},$$

where

S_{ijk} = the survival probability for individual j of population i in interval k ,

P_{ijk} = the detection probability for individual j of population i in interval k ,

λ_{ij} = the final survival and detection probability for individual j of population i ,

χ_{ijk} = the probability of individual j of population i not being detected again after interval k ,

with

$$\chi_{ijK} = 1,$$

$$\chi_{ij,K-1} = 1 - \lambda_{ij},$$

$$\chi_{ijk} = 1 - S_{ij,k+1} + S_{ij,k+1}(1 - P_{ij,k+1})\chi_{ij,k+1}, \text{ for } k = 0, \dots, k - 2,$$

and

t_{0ij} = the initial release event for individual j of population i ,

t_{1ij} = the event where individual j of population i was last detected,

$D_{ijk} = 1$ if individual j of population i was detected at event k , 0 otherwise, and

$C_{ij} = 1$ if individual j of population i was removed (censored) at event t_{1ij} , 0 otherwise.

We define the likelihood model for an individual-covariate based model as follows:

$$L(\vec{S}, \vec{P}, \vec{\lambda}) = \prod_{i=1}^I \prod_{j=1}^{J_i} L_{ij}(\vec{S}_{ij}, \vec{P}_{ij}, \lambda_{ij}), \quad (\text{B.22})$$

where J_i = the number of individuals in population i .

Note that because Equation (B.22) is an individual-based likelihood, there is not the problem of overparameterizing the model, as in the case with the group-covariate model.

After the parameters are estimated using the individual-based likelihood (Equation B.22), the survival probability for population i , interval k , is then calculated as

$$S_{ik} = \frac{\sum_{j=1}^{J_{ik}} S_{ijk}}{J_{ik}}, \quad (\text{B.23})$$

where J_{ik} = the number of individuals of population i at risk in interval k . An individual is considered to be at risk in a particular interval if it was released prior to k at the beginning of the interval and was detected at the beginning of the interval or later and was not removed.

B.2.4 Modeling the Product Probabilities

The discussion on modeling so far has focused on modeling the survival probabilities as a function of main effects, group covariates, and individual covariates.

The final product probabilities—the product of survival and detection in the final interval—can also be modeled as a function of main effects, group covariates, and individual covariates.

For a model that includes group covariates but no individual covariates, the product parameter is modeled as

$$\lambda_i = f \left(\lambda_0, \pi_i + \sum_{m=1}^M \beta_{im} x_{iKm} \right),$$

where

M = the number of group covariates,

β_{im} = the group covariate effect for population i for the m th covariate, and

x_{iKm} = the value of the m th covariate for population i , interval K (the final interval).

For a model that includes individual covariates, the product parameter is modeled as

$$\lambda_{ij} = f \left(\lambda_0, \pi_i + \sum_{m=1}^M \beta_{im} x_{iKm} + \sum_{n=1}^N \gamma_{in} y_{iKn} \right),$$

Where

N = the number of individual covariates,

γ_{in} = the individual covariate n , population i ,

x_{ikm} = the value of the individual covariate n for individual j of population i , interval K .

B.2.5 Link Functions

As shown in the previous sections, SURPH models the covariates as a linear predictor of survival (or detection).

$$S = f(S_0, \bar{\theta}'\vec{x}), \quad (\text{B. 24})$$

where

S = the survival parameter,

S_0 = the baseline survival, and

$\bar{\theta}'\vec{x}$ = the linear function of the main effects and the covariate data.

For example, in Equation (B.19), the survival parameter is S_{ik} and the linear function $\bar{\theta}'\vec{x}$ is

$$\pi_{ik} + \sum_{m=1}^M \beta_{ikm} x_{ikm}.$$

The function that links the baseline survival and the linear function to the survival probability is the link function. SURPH 3 supports three link functions: (1) the logit link, (2) the proportional hazards link, and (3) the identity link.

B.2.5.1 Logit link function

The logit link is defined as

$$S = \frac{\exp(S_0 + \bar{\theta}'\vec{x})}{1 + \exp(S_0 + \bar{\theta}'\vec{x})}. \quad (\text{B. 25})$$

Note that if the linear function $S_0 + \vec{\theta}'\vec{x}$ ranges in the interval $(-\infty, \infty)$, the logit link bounds the resulting survival probability in the interval $(0, 1)$. The logit link is the default link for the detection probabilities.

Note: the function of the linear predictor shown in the right-hand side of Equation (B.25) is the inverse logit function. Equation (B.25) can be rearranged to use the logit function to relate the linear predictor to the expected survival via

$$S_0 + \vec{\theta}'\vec{x} = \ln(S/(1 - S)).$$

We refer to both this formulation and the formulation shown in Equation (B.25) as the logit link function.

B2.5.2 Proportional hazards link function

SURPH takes its name from the proportional hazards model (SURvival with Proportional Hazards). The hazard function, $h(t)$, is the probability of a mortality at time t , given survival to time t . With a proportional hazards model, the covariates are assumed to have a multiplicative effect on the hazard function. If the covariates do not change through time, the ratio of the hazard functions will be constant through time.

For a survival probability, the proportional hazards link function is defined as

$$S = \tilde{S}_0^{\exp(\vec{\theta}'\vec{x})} \tag{B.26}$$

where \tilde{S}_0 is the inverse logit of S_0 , that is:

$$\tilde{S}_0 = \frac{\exp(S_0)}{1 + \exp(S_0)}.$$

Applying the logit link to S_0 ensures that \tilde{S}_0 is in the interval $(0, 1)$ and, thus, the resulting probability (S) is in the interval $(0, 1)$ as well. Note that the hazards link is the same as the log-log link.

B.2.5.3 Identify link function

The identity link function is defined simply as

$$S = S_0 + \vec{\theta}' \vec{x} \quad (\text{B. 27})$$

While the easiest link function to understand and the most intuitive, it does not constrain the resulting parameter in the interval (0, 1) and the use of it often causes the parameters to become inestimable.

B.3 Detection Probability Estimation with Manly-Parr Estimates

Manly and Parr (1968) suggest a way of estimating the detection probabilities of a release-recapture structure independent of the survival process. The summary statistics used in the Manly-Parr estimates are:

g_{ik} = the number of animals marked before sampling event k in population i that were detected both at event k and at least one later event,

z_{ik} = the number of animals marked before sampling event k in population i that were not detected at event k but were detected on at least one later event, and

ρ_{ik} = the proportions of animals detected at event k in population i that were removed.

B.3.1 Closed-Form Estimates

A closed-form estimate of the detection probability for population i , event k , can then be calculated as

$$\hat{P}_{ik} = \frac{g_{ik}}{g_{ik} + z_{ik}} \quad (\text{B. 28})$$

When there are known removals, Equation (B.28) is biased, underestimating the true detection probability. This is because only detected animals can be removed; thus, known removals decrease the number of animals that contribute to g_{ik} but not to z_{ik} . To correct the bias, z_{ik} must be adjusted downward:

$$z'_{ik} = (1 - \rho_{ik})z_{ik}$$

and

$$\hat{P}_{ik} = \frac{g_{ik}}{g_{ik} + z'_{ik}}, \quad (\text{B.29})$$

The estimator of the capture probability in Equation (B.29) is unbiased.

B.3.2 Modeling Detection Probabilities with Manly-Parr Estimates

The likelihood for the Manly-Parr estimate can be written as

$$L \propto \sum_{i=1}^I \sum_{k=1}^{K-1} \hat{P}_{ik}^{g_{ik}} (1 - \hat{P}_{ik})^{z_{ik}}, \quad (\text{B.30})$$

where

I = the number of populations, and

K = the number of intervals.

\hat{P}_{ik} = can be modeled as described in Appendix B.2. If \hat{P}_{ik} is modeled as a function of individual covariates, then the likelihood for individual j of population i can be written as

$$L_{ij} \propto \prod_{k=t_{ij}+1}^{w_{ij}} P_{ijk}^{D_{ijk}} (1 - P_{ijk})^{(1-D_{ijk})(1-\rho_{ijk})}, \quad (\text{B.31})$$

where

t_{ij} = the release event for individual j of population i ,

w_{ij} = the last detection event (or last detection event minus 1 if last detection event = K) for individual j of population i , and

D_{ijk} = 1 individual j of population i detected at event k ; 0, otherwise.

The overall likelihood is then

$$L \propto \prod_{i=1}^I \prod_{j=1}^{J_i} L_{ij},$$

where J_i is the number of individuals in population i .

The overall estimate of detection probability is then estimated using Equation (B.23).

B.4 Diagnostic Plot Calculations

B.4.1 Individual Covariate Plots

In Section 5.2, we show how SURPH allows the user to plot the survival (or detection) probability as a function of a selected individual covariate. The parametric function (i.e., based on the chosen model parameterization) is plotted with the nonparametric function (i.e., not based on any model assumptions). In this section, we describe how the functions are calculated.

B.4.1.1 Nonparametric estimate of survival

The nonparametric estimate of survival is based on a moving average of the covariate values. For the desired covariate, the covariate values of the individuals at risk are sorted from the smallest value to the largest. A “data window” encompasses all covariate values to be used for a given point on the nonparametric survival curve, starting with the minimum size of eight individuals and growing to the right (larger values) until it reaches its maximum size of one-fifth of the total covariate data for the population and interval. It then shrinks to the minimum size of eight again as it reaches the right boundary (larger covariate values). The x -value used to calculate the nonparametric estimate of survival is, thus, the average of all covariate values within the current data window.

The nonparametric estimate of survival is based on the number of individuals whose covariate value falls within the data window defined above, and is defined as

$$S_{ikp}(y_p) = \left(\frac{\text{number of individuals at risk}}{\text{number of individuals at risk and detected at next event}} \right) \cdot \frac{1}{\hat{P}_{ik}},$$

where

y_p = the average of individual covariate p in the moving “data window,”

S_{ikp} = the estimate of survival for population i , interval k , based on covariate p , and

\hat{P}_{ik} = the CJS estimate of detection probability for population i , interval k .

B.4.1.2 Nonparametric estimate of detection

The nonparametric estimate of detection uses the same moving “data window” as the nonparametric estimate of survival does in the previous section. The detection probability is estimated as

$$P_{ikp}(y_p) = \frac{n_1}{n_2},$$

where

n_1 = the number at risk and detected at next event, and

n_2 = the number at risk and detection at next event *and* at least one later event.

B.4.1.3 Parametric function of an individual covariate

This section discusses survival but is equally applicable to detection as a function of an individual covariate.

Survival as a function of an individual covariate is denoted as

$$S_{ikp}(y_p),$$

where y_p is the given value of individual covariate p .

Survival as a function of a given individual covariate value is calculated as follows:

$$S_{ikp}(y_p) = f \left(S_0, \pi_{ik} + \sum_{m=1}^M \beta_{ikm} x_{ikm} + \sum_{n=1}^{p-1} \gamma_{ikn} \bar{y}_{ikn} + \sum_{n=p+1}^N \gamma_{ikn} \bar{y}_{ikn} + \gamma_{ikp} y_p \right),$$

where \bar{y}_{ikn} = the average individual covariate value for individual covariate n for all individuals at risk in population i in interval k .

Detection as a function of a given individual covariate is calculated in an analogous manner.

B.4.2 Group Covariate Plots

In Section 5.2.2, we show how SURPH allows the user to plot the survival probability as a function of a selected group covariate. The group covariate may be either interval-specific or population-specific. Again, all discussion is equally applicable to detection probabilities.

B.4.2.1 Interval-specific group covariate function

Let $S_{kp}(x_p)$ be the survival at interval k for group covariate p as a function of x_p :

$$S_{kp}(x_p) = f \left(S_0, \bar{\pi}_k + \frac{1}{I} \sum_{i=1}^I \sum_{n=1}^N \gamma_{ikn} \bar{y}_{ikn} + \sum_{\substack{m=1 \\ m \neq p}}^M \beta_{km} \bar{x}_{km} + \beta_{kp} \chi_{kp} \right),$$

where

f = the link function,

I = the number of populations,

N = the number of individual covariates,

M = the number of group covariates,

$$\bar{\pi}_k = \frac{\sum_{i=1}^I \pi_{ik}}{I},$$

\bar{y}_{ikn} = the average value of individual covariate n for all individuals of population i at risk in interval k , and

$$\bar{x}_{km} = \frac{\sum_{i=1}^I x_{ikm}}{I}.$$

B.4.2.2 Population-specific group covariate function

Let $S_{ip}(x_{ip})$ be the survival for population i for group covariate p as a function of x_p :

$$S_{ip}(x_p) = f \left(S_0, \bar{\pi}_i + \frac{1}{K} \sum_{k=1}^{K-1} \sum_{n=1}^N \gamma_{ikn} \bar{y}_{ikn} + \sum_{\substack{m=1 \\ m \neq p}}^M \beta_{im} \bar{x}_{im} + \beta_{ip} x_p \right),$$

where

f = the link function,

$$\bar{\pi}_i = \frac{\sum_{k=1}^{K-1} \pi_{ik}}{K}, \text{ and}$$

$$\bar{x}_{im} = \frac{\sum_{k=1}^{K-1} x_{ikm}}{K}.$$

B.5 Analysis of Deviance

Analysis of Deviance requires at a minimum three baseline models:

1. The **Full** model is fully parameterized with N_F parameters and log-likelihood l_F . Because it is fully parameterized, it fits the data the best but uses all degrees of freedom and, thus, is uninformative. In the Full model, the survival probabilities are parameterized like the CJS estimates with a unique survival probability for each interval and population.
2. The **Null** model is the baseline that omits all covariates, with N_0 parameters and log-likelihood l_0 .
3. The **Full Covariate** model is based on the Null model but with effects included for all covariates of interest. It has N_C parameters and log-likelihood l_C .

The Null model must be nested in both the Full model and the Full Covariate model.

From the three baseline models, we can construct a basic analysis of deviance as shown in Table B.11.

Table B.11. Basic analysis of deviance model

Source	DF	DEV	MDEV	<i>F</i>	<i>P</i>
Corrected total	$N_F - N_0$	$2(l_F - l_0)$	DEV (Corr)/ DF (Corr)		
Full covariate	$N_C - N_0$	$2(l_C - l_0)$	DEV (Full Cov)/ DF (Full Cov)	MDEV (Full Cov)/ MDEV (Resid)	Prob (<i>F</i>)
Residual	$N_F - N_C$	$2(l_F - l_C)$	DEV (Resid)/ DF (Resid)		

The three rows in the table are as follows:

Corrected total is the total amount of deviance to be explained, i.e., the difference between the deviance of the Full model and the Null model.

Full covariate is the portion of the Corrected total explained by the covariates.

Residual is the portion of the Corrected total not explained by the Full Covariate model.

In Table B.11, the deviance (“DEV”) is defined as two times the difference in the log-likelihood values between two models. The mean deviance (“MDEV”) is the deviance is divided by the degrees of freedom, defined as the difference between the number of parameters for the given model and the number of parameters for the nested model. In computing the mean deviance, the nested model is either the Null model or the Full Covariate model, depending on the row of the table. Under standard normal theory, the mean deviance has a χ^2 distribution. Thus, the ratio of mean deviances has an *F* distribution if both models are equivalent, with respective numerator and denominator degrees of freedom.

Analysis of Deviance uses the fact that the ratio of the mean deviances is a known distribution to perform hypothesis testing. There is one hypothesis test in the basic

ANODEV table as shown in Table B.11. The mean deviance for the Full Covariate model is divided by the mean residual deviance to test

H_0 : The Full Covariate model and Null model are equivalent – that is, the Full Covariate model that includes the group covariate(s) is no better than the Null model that does not include the group covariate(s),

versus

H_a : The Full Covariate model is significantly better.

The F -statistic is given in the F column with the corresponding P -value in the P column.

Beyond the basic ANODEV model in Table B.11, the user can define models to further partition the total deviance and test for significance, as shown in Table B.12.

In Table B.12, the deviance in the Full Covariate model is further partitioned by models M_1 through M_n . Model M_1 must be nested in the Null model, M_2 must be nested in M_1 , and so on. Finally, model M_n must be nested in the Full Covariate model. The F -test in the line “ M_1 vs Null” tests the significance of the model M_1 versus the Null model. The F -test for the line “ M_2 vs M_1 ” tests the significance of the model M_2 versus M_1 , and so on.

Table B.12. Analysis of Deviance table with the Full Covariate deviance partitioned by n submodels.

Source	DF	DEV	MDEV	F	P
Corrected total	$N_F - N_0$	$2(l_F - l_0)$	Deviance/DF	–	–
Full covariate	$N_C - N_0$	$2(l_C - l_0)$	Deviance/DF	MDEV (Full Cov)/ MDEV (Resid)	Prob (F)
M_1 vs. Null	$N_{M_1} - N_0$	$2(l_{M_1} - l_0)$	Deviance/DF	MDEV (M_1)/ MDEV (Resid)	Prob (F)
M_2 vs. M_1	$N_{M_2} - N_{M_1}$	$2(l_{M_2} - l_{M_1})$	Deviance/DF	MDEV (M_2)/ MDEV (Resid)	Prob (F)
□	□	□	□	□	□
$M_n - M_{n-1}$	$N_{M_n} - N_{M_{n-1}}$	$2(l_{M_n} - l_{M_{n-1}})$	Deviance/DF	MDEV (M_n)/ MDEV (Resid)	Prob (F)
Full covariate vs. M_n	$N_C - N_{M_n}$	$2(l_C - l_{M_n})$	Deviance/DF	MDEV (Full Cov)/ MDEV (Resid)	Prob (F)
Residual	$N_F - N_C$	$2(l_F - l_C)$	Deviance/DF	–	–

B.5.1 Analysis of Deviance Preferred for Group Covariates

SURPH models survival as a function of group covariates, as shown in Equation (B.19), which can be simplified for the purpose of illustration to

$$S_{ij} = f(\pi_j + \beta_j x_{ij}), \quad (\text{B.32})$$

where

f = the link function,

S_{ij} = the survival for population i , interval j ,

π_j = the main effects for interval j ,

β_j = the covariate effect for interval j , and

x_{ij} = the measured value of the covariate for population i , interval j .

Models are constructed in SURPH, and the parameters estimated for both the alternative hypothesis ($H_a : \beta_{ij} \neq 0$) and the null hypothesis ($H_o : \beta_{ij} = 0$). It is possible to test the hypothesis with the Likelihood Ratio Test, but Smith (1991) has shown that the ANODEV is the preferred method for testing the significance of group covariates in a model.

Both ANODEV and the F -statistic test the significance of the effect parameter β . The Likelihood Ratio Test, however, assumes that there is no variability between populations. The Likelihood Ratio Test chooses between two alternatives: either there is no variability between populations or the variability is explained by the covariate effect being tested.

In reality, there may be variability between populations unrelated to the covariate effect. The denominator of the F -statistic in the ANODEV accounts for variability not related to the covariate. In a sense, the F -test tests the significance of β_j in the following model structure:

$$S_{ij} = f(\pi_j + \beta_j x_{ij} + \varepsilon_{ij}), \quad (\text{B.33})$$

where ε_{ij} is the variability in S_{ij} not accounted for by covariate.

Smith (1991) showed with simulations that the Likelihood Ratio Test rejected the null hypothesis far too often when the null hypothesis was true. The ANODEV F -test rejected the null hypothesis at the nominal level of the test, thus making it the preferred method for testing the effect of group covariates.

Note that Analysis of Deviance is not valid with models that include individual covariate parameters.

B.6 Literature Cited

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 5.

Cormack, R. M. 1964. [Estimates of survival from the sighting of marked animals](#). *Biometrika* 51: 429–438.

Jolly, G. M. 1965. [Explicit estimates from capture-recapture data with both death and immigration-stochastic model](#). *Biometrika* 52: 225–247.

Manly, B.F. J., and M. J. Parr. 1968. A new method of estimating population size, survivorship, and birth rate from capture-recapture data. *Transactions of the Society for British Entomology* 18: 81–89.

Seber, G. A. F. 1965. [A note on the multiple recapture census](#). *Biometrika* 52: 249–259.

Smith, S. G. 1991. Assessing hazards in wild populations using auxiliary variables in tag-release models. Dissertation, University of Washington, Seattle, Washington, USA.

