# Part 6. Planning Experiments

## 6.1. Introduction

The careful design of survival experiments is important. The design must consider the assumptions (Section 1.4.4), the feasibility of different experimental protocols (Part 2), a set of statistical issues, and practical constraints. Our objective here is to provide a discussion of some of these general considerations. We make no attempt to cover field procedures because they are specific to each experiment and taxonomic group of interest. Before a more formal approach to the design of survival experiments can be given, we must have more experience with the analysis of real data. Until the methods presented here have been used, we can only provide a general discussion of some of the important issues, a mention of certain compromises, and a few "rules of thumb."

There is not much literature concentrating on design issues in release-recapture (see Manly 1977). Moreover, much of the existing literature on classical design issues for capture studies (Skalski 1985) is for simpler situations than we consider here. As experimental design is usually approached, one aspect involves examining tradeoffs of sampling effort and costs, and optimal allocation of resources to achieve preselected goals of precision or test power (the other aspect of design involves principles of validity such as randomization and replication). Nonlinear programming (Hadley 1964, Mangasarian 1969) provides an optimal approach for evaluating some of the tradeoffs in experimental design. However, in general, we lack a priori knowledge as to which design variables are known versus unknown. For a particular study, the total budget, number of fish available, the recapture rates ($p_i$), number of sampling sites ($k$), and costs of marks, handling, and marking may be known approximately. The design may then focus on the number of lots to use, the allocation of the proportion of fish to $R_{t1}$ and $R_{c1}$, and issues such as stratification by size and timing of releases. Nonlinear programming provides a class of methods for assessing such compromises; we recommend this approach for achieving a nearly optimal design, at least for large, expensive experiments.

### 6.1.1. Desirability of a Pilot Experiment

The conduct of a small pilot experiment is important as it may provide needed insight into a variety of design questions. Critical to such a preliminary experiment is $k \geq 3$, which allows the assumption $p_{t2} = p_{c2}$ to be assessed by TEST 1.*T*2. Other statistical tests and field experience will aid in preliminary model selection. Rough estimates of $\phi_{vi}$, $p_{vi}$, and $S$ will be available following a pilot experiment. These estimates are useful in design considerations covered in the following chapters. Ideally, the complete capture history protocol should be used in the preliminary study.

In addition to the statistical considerations, a pilot experiment will provide evidence concerning feasibility of the field aspects of the full-scale experiment. Cost and labor requirements can be better projected from the experience gained during the pilot experiment. Difficulties with marking, handling, transporting, recapturing, and rereleasing animals will become apparent. The need for stratification and other issues can be assessed in preparation for the main experiment. We strongly recommend that such a preliminary experiment be conducted as an integral part of the study design.

### 6.1.2. Review of Assumptions

The general assumptions that underly the statistical methods presented here are repeated from Section 1.4.4.

(1) The test fish used are representative of the population of fish about which one seeks treatment mortality information.

(2) Test conditions are representative of the conditions of interest.

(3) Treatment and control fish are biologically identical prior to release at dam 1. A strong version of assumption 3 is that initial handling, marking, and holding do not affect survival rate.

(4) The numbers of fish released, by lots, are known exactly.

(5) Marking (tagging) is accurate; there are no mark (tag) losses and no misread marks (tags).

(6) All releases and recaptures occur in brief time intervals and recaptured fish are released immediately.

Assumptions 7-8 relate to the stochastic component of the models.

(7) The fate of each individual fish, after any known release, is independent of the fate of any other fish.

(8) With multiple lots (or other replication), the data are statistically independent over lots.

Assumptions 9-12 relate to model structure.

(9) Statistical analysis of the data is based on the correct model.

(10) Treatment and control fish move downstream together.

(11) Captured fish that are rereleased have the same subsequent survival and capture rates as fish alive at that site which were not caught, i.e., capture and rerelease do not affect their subsequent survival or recapture.

(12) All fish (in the study) of an identifiable class (e.g., treatment or control, size, or replicate) have the same survival and capture probabilities; this is an assumption of parameter homogeneity.

The first two assumptions are fundamental; they are not statistical nor testable, rather they are biological judgments. If these assumptions are not valid, there is little point in proceeding. If the treatment effect $S$ is the only objective, then a weaker version of assumption 1 suffices: the treatment effect must be the same for the test fish and the population of fish being investigated.

The investigator has considerable control over the validity of assumptions 3-6 (some additional insight into these matters is provided in Part 7). It is important to assign randomly fish to treatment groups and lots within any stratifications, such as by age or size. If several persons are handling and marking fish, then these persons should be randomly rotated (e.g., marker A should not mark all of the control fish while marker B marks all of the treatment fish). The handling and transport methods should be identical for all treatment groups. Often, it is appropriate to hold marked fish for a period of time prior to release. Deaths during the holding period can be recorded and their total subtracted from the total numbers marked such that $R_{v1}$ accurately reflects the actual numbers released (assumption 4).

A study should be designed so that any ill effects of marking on survival are minimized, otherwise, survival estimates ($\hat{\phi}_{vi}$) will be biased negatively. If the marking effect is equal for treatment and control fish, comparisons may still be reasonable. However, such comparisons may be invalid if an interaction between the marking and treatment-induced mortality occurs. Thus, treatment fish would have higher mortalities relative to control fish than if no marking mortality occurred. If marking is known to affect survival, then protocols that require a second batch mark (partial capture history schemes A and B) should be avoided.

Minimization of mark loss is important when study is designed. Once again, however, inferences about a treatment effect will still be valid if treatment and control fish lose their marks at the same rate. Another possibility with larger fish or other animals might be to use a double-marking scheme so that one could adjust for mark loss (Seber 1982:94). Accurate marking is critical; marks should not become lost or unreadable, and should be accurately recorded. Again, the validity of the treatment survival rate estimator $\hat{S}$ rests on somewhat weaker assumptions than estimators of $\phi_{vi}$ (see Section 1.4.4).

Treatment effects ($S$) and environmental and sampling effects ($\phi$, $p$) should not be confounded with time effects, which are possibly associated with the release and recapture of fish; this is the essence of assumption 6. A given lot of treatment fish should be released in as short a time interval as possible so that all fish experience identical passage conditions. It is also desired that all recaptures occur in a brief time interval. In some types of studies, this objective can be accomplished; however, the recapture process at a given site usually is spread over time. This component of assumption 6 then gets replaced by assumption 10. When recaptures do occur, those fish should be immediately released. The longer the time recaptured fish are held, the more there may be a capture effect. Thus, captured and uncaptured fish that pass that dam will have different subsequent survival and capture rates.

Assumptions (7) and (8) relate to independence among fish and lots, respectively. The investigator usually has little control over these issues. Fortunately, these are relatively minor assumptions that are probably met approximately in most cases. Moreover, failure of

assumption (7) does not cause any bias in parameter estimation; it only causes the theoretical variance to be too low. By using empirical estimators of variance, we avoid the need for assumption (7).

Substantial bias may arise if an incorrect model is used in the analysis of data (assumption 9). For example, if the recapture rate at dam 2 differs between treatment and control groups (i.e., $p_{t2} \neq p_{c2}$), then model $H_{2p}$ is appropriate if other parameters for downstream dams are the same for each group. The use of the estimator of treatment survival under model $H_{1\phi}$ will result in a biased estimate of the treatment survival rate $S$. The design should allow adequate replication and sample sizes such that statistical tests will have high power. Thus, correct model selection will be likely, and use of the correct model is critical to valid, efficient inferences.

Treatment and control fish should move downstream together (assumption 10). This assumption, as the previous one, is closely tied with model selection. Ideally, fish in all the groups move at similar rates. If differential movement occurs, a more complex model may be required (e.g., $H_{4p}$ rather than $H_{1\phi}$). In the extreme case where fish in the different groups never move simultaneously past some point $j$, then valid estimation may not be possible, although testing for a treatment effect remains valid.

Sometimes the treatment and control groups might have the same parameters ($p_j$ and $\phi_j$), but some groups move ahead of the others. In this case, it is important to sample at each dam $j$ for a sufficient period of time such that all groups are appropriately sampled. We have seen one instance where sampling was terminated after, say 3 weeks at dam $j$ and it appeared that the control fish had all passed the dam, but only perhaps 90% of the treatment fish had passed the dam by the time sampling efforts were terminated. This type of situation must be avoided.

Recapture and rerelease of fish should not affect their subsequent survival or recapture rates (assumption 11). This assumption is important for many of the models proposed. TESTs 2 and 3 provide powerful assessments of this assumption. In particular, this assumption may fail commonly with young fish. In that case, the first capture history protocol should be considered. An interesting alternative is provided by the use of PIT tags whereby recapture and rerelease do not involve handling. Instead, the tag number and, therefore, group membership, are recorded automatically as the fish pass through special recorders in the bypass system of a large hydroelectric dam.

When recapture is at dams, the ability to meet assumption 11 may depend on where the captured fish are rereleased. If fish are released below the dam, they are not exposed to the same risks as fish that pass through the dam by some route and are not caught. If the fish are released above the dam, they will have experienced more risk than fish not caught. The captured fish could be released into some bypass structure, but even then they would tend to experience slightly different risks than fish not caught that will take a variety of routes through the dam (turbines, bypasses, spillways, fish ladders). Perhaps the best strategy is to release the fish directly into the bypass route that most fish are expected to use.

Assumption 12 relates to homogeneity of fish within a group. Again, the investigator can stratify the sample fish by age, sex, size, or other variable to limit the degree of

heterogeneity. If homogeneity is slightly violated, few problems will result (see Part 5). The main effect will be that the theoretical estimates of sampling variances will be slightly altered. However, this problem tends to vanish if replication is done to allow empirical estimates of variances.

## 6.2. Selection of an Experimental Protocol

The investigator must consider the type of tag or mark to use and whether fish are to be removed or rereleased upon capture at dam $j$ ($j > 1$). A summary of the experimental protocols is presented in Table 6.1 for review here. Some considerations for each protocol follow.

Table 6.1. – Summary of the important elements involved in the different experimental protocols.

| Protocol | Marking scheme | Remove or rerelease | Number of recaptures |
|---|---|---|---|
| Complete capture history | Unique marks | Either[a] | Up to $k$ - 1 |
| Complete capture history | $k$ - 1 batch marks | Either[a] | Up to $k$ - 1 |
| Partial capture history, scheme A | $k$ - 1 batch marks | Remove after second recapture | 2 |
| Partial capture history, scheme B | 2 batch marks | Remove at sites 3 - $k$ | 2 |
| Unknown capture history | 1 batch mark | Rerelease at sites 2 - $k$ | Up to $k$ - 1 |
| First capture history | 1 batch mark | Remove at sites 2 - $k$ | 1 |

[a]Some fish, but not all, can be intentionally removed. Enough fish must be rereleased so that some are recaptured in future samples.

### 6.2.1. Complete Capture History Protocol

Setting aside costs and other nonstatistical considerations, the ideal experiment would use the complete capture history protocol. Under this protocol, the survival ($\phi_{vi}$) and recapture ($p_{vi}$) rates can be estimated and several critical hypotheses can be tested (TESTs 2 and 3).

These estimates and test results provide flexibility in the analysis of a treatment effect. In addition, internal replication can be achieved by post-stratifying the data based on the last digit of the tag number. This procedure permits the estimation of empirical sampling variances based on the 10 internal replicates (see Chapter 7.3 for an example).

If resampling does not involve handling animals, which could be avoided by the use of PIT tags on fish or resighting methods on terrestrial vertebrates, the complete capture history protocol seems ideal if the recapture rates are high. If the $p_{vi}$ are low, few recaptures are made and little information is gained.

### 6.2.2. Partial Capture History Protocol

The partial capture history protocol offers many of the same advantages as the complete capture history protocol. An advantage is that batch marks, identifying only group membership, can be used. However, if handling adversely affects the subsequent survival or recapture rates, the partial capture history protocol should not be used.

The recapture rates must be high or few animals will be remarked and available for subsequent recapture. We believe this protocol will be quite valuable for terrestrial vertebrates, but may see limited application in fisheries investigations.

### 6.2.3. Unknown Capture History Protocol

The unknown capture history protocol may sometimes be considered when the recapture rates are relatively low and are known a priori to be unaffected by the treatment. Only the parameter $S = \phi_{t1}/\phi_{c1}$ can be estimated from data under this protocol. This protocol has been used in the past in fisheries investigations where the cost of fish for the study and of batch-marking have been fairly low and the recapture rates have been low. In general, we do not recommend this protocol, especially when $k = 2$.

### 6.2.4. First Capture History Protocol

The first capture history protocol is most appropriate if $H_{1\phi}$ is the true model; i.e., there is a direct, acute treatment effect. Only batch marks are required. Only the treatment survival rate is estimable and the method is increasingly precise as the $p_{vi}$ increase. The estimator of $S$ is identical to those of the complete and partial capture history protocols under model $H_{1\phi}$. As this protocol relies on removal data, it is appropriate (actually necessary) when a handling effect is present.

The extended sequence of models $H'_{2\phi}$, $H'_{3\phi}$, ..., $H'_{k-1,\phi}$ should be considered if the treatment is chronic and it is known that treatment does not affect the $p_{vi}$. We doubt that this

series of models will be generally useful, except as approximations, as mentioned in Chapter 3.9.

## 6.3. Effort and Sample Size Considerations

### 6.3.1. Introduction

The question most often asked of statisticians during a study design is "How big a sample do I need?" This question is not easy to answer, even for simple studies. For the experiments considered here, the question should be rephrased, "How much effort will this study require to achieve a given level of precision for $\hat{S}$ and to have high power of the statistical tests used to evaluate the assumptions of the experiment?" The standard approach is to assume a model structure (which cannot be done with certainty), specify the desired cv($\hat{S}$), specify all the parameters of the model, write down the formula for var($\hat{S}$), and solve for sample size. For example, under model $H_{1\phi}$, the theoretical large-sample variance of $\hat{S}$ is

$$\text{var}(\hat{S}) = (S)^2 \left[ \frac{1}{E(r_{t1})} - \frac{1}{R_{t1}} + \frac{1}{E(r_{c1})} - \frac{1}{R_{c1}} \right].$$

This formula can be rewritten as

$$[\text{cv}(\hat{S})]^2 = \frac{1}{R_{t1}} \left[ \frac{1}{S\lambda_{c1}} - 1 \right] + \frac{1}{R_{c1}} \left[ \frac{1}{\lambda_{c1}} - 1 \right]$$

where

$$\lambda_{c1} = E\left( \frac{r_{c1}}{R_{c1}} \right).$$

If the investigator (1) assumes the model structure of $H_{1\phi}$, (2) assumes the theoretical variance is appropriate, (3) assumes a value for $S$, such as 0.9, (4) assumes a value for $\lambda_{c1}$, such as 0.15, (5) assumes, for example, $R_{t1} = R_{c1} = R$ (so $2R$ total fish are released), and (6) specifies cv($\hat{S}$), say as 0.05, then the required sample size $R$ can be found. In this example, the relevant equation becomes

$$[\text{cv}(\hat{S})]^2 = \frac{1}{R} \left[ \frac{1}{(0.9)(0.15)} - 1 \right] + \frac{1}{R} \left[ \frac{1}{0.15} - 1 \right];$$

hence,

$$(0.05)^2 = \frac{12.074}{R},$$

or $R$ = 4,830. A total of about 10,000 fish is required. However, this sample size is only as reliable as the assumptions used to compute it. If the model structure is reasonable but the theoretical variance is too low (compared to the actual variance) by a factor or two or more (a common situation), then $R$ can be substantially wrong. Further consideration of absolute numbers of fish to release is in Section 6.3.6.

The sample size questions must be generalized here to be a series of questions about (1) how much total effort is needed and (2) how effort should be allocated proportionally to different aspects of the experiment. Quantities that are at least partially controllable are $R_{t1}$, $R_{c1}$, and recapture effort. It is much easier to consider optimal allocation of total effort than to try and determine what the total effort should be. Also, knowing the optimal relative allocation makes it easier to determine total effort. For example, the sample size question has two components: (1) what is the necessary total size, $R_{t1} + R_{c1}$, and (2) what is the optimal ratio of treatment to control releases, $R_{t1}$ versus $R_{c1}$? We will first consider questions of allocation of effort, given some total effort, and then return to the question of determining total effort.

### 6.3.2. Relative Numbers of Releases

*6.3.2.1. Two groups.* – Given a total number of available fish, one of the easiest questions to address is the optimal relative numbers of treatment and control fish, $R_{t1}$ and $R_{c1}$, respectively. In particular, for two groups, it will be nearly optimal to have equal numbers of fish released for both groups, i.e., $R_{t1} = R_{c1}$. We explore here the two-group case in some detail.

Under model $H_{1\phi}$, the variance of $\hat{S}$ may be taken as

$$\text{var}(\hat{S}) = c\,(S)^2 \left[\frac{1}{R_{t1}}\left(\frac{1}{S\lambda_{c1}} - 1\right) + \frac{1}{R_{c1}}\left(\frac{1}{\lambda_{c1}} - 1\right)\right].$$

Here, $c$ is an unknown variance inflation factor that reflects the phenomenon often observed of excess variation (see Part 4). Letting the total $R_{.1} = R_{t1} + R_{c1}$ be fixed but arbitrary, we use calculus to find the optimal allocation ratio $R_{t1}/R_{c1}$ to minimize $\text{var}(\hat{S})$. This minimization is facilitated by realizing the form of $\text{var}(\hat{S})$ is

$$\text{var}(\hat{S}) = \frac{A}{R_{t1}} + \frac{B}{R_{c1}},$$

where

$$A = c(S)^2 \left[\frac{1}{S\lambda_{c1}} - 1\right]$$

and

$$B = c(S)^2 \left[\frac{1}{\lambda_{c1}} - 1\right].$$

Using $R_{c1} = R_{.1} - R_{t1}$, write the above as

$$\text{var}(\hat{S}) = \frac{1}{R_{.1}} \left[\frac{A}{\gamma} + \frac{B}{1-\gamma}\right],$$

where

$$\gamma = \frac{R_{t1}}{R_{.1}}, \qquad 0 < \gamma < 1.$$

Taking the derivative of $\text{var}(\hat{S})$ with respect to $\gamma$, gives

$$\frac{\partial \text{var}(\hat{S})}{\partial \gamma} = \frac{1}{R_{.1}} \left[\frac{-A}{\gamma^2} + \frac{B}{(1-\gamma)^2}\right].$$

Setting this derivative equal to zero and solving it yields the value of $\gamma$ that gives the minimum $\text{var}(\hat{S})$ for any fixed $R_{.1}$:

$$\gamma = \frac{1}{1 + (B/A)^{1/2}}.$$

Translating this general result back into terms of the original variables gives

$$\frac{R_{t1}}{R_{.1}} = \left[1 + \left(\frac{1 - \lambda_{c1}}{\frac{1}{S} - \lambda_{c1}}\right)^{1/2}\right]^{-1}. \tag{6.1}$$

From (6.1) we find the optimal ratio of $R_{t1}/R_{c1}$:

$$\frac{R_{t1}}{R_{c1}} = \left(\frac{\frac{1}{S} - \lambda_{c1}}{1 - \lambda_{c1}}\right)^{1/2}. \tag{6.2}$$

From (6.2) it is clear that under optimal allocation of release numbers, we have $R_{t1} \geq R_{c1}$; they are equal if and only if $S = 1$.

Formulae (6.1) and (6.2) give the optimal ratio of treatment to total, or to control, releases (respectively) when $V = 2$ and model $H_{1\phi}$ is assumed. One interesting feature of this result is that the variance inflation factor, $c$, drops out. Thus, the actual magnitude of $\text{var}(\hat{S})$ does not influence the optimal allocations of fish to treatment and control groups. Rather, optimal allocation depends only on the true value of $S$ and the probability of recapturing a fish that is alive below dam 1 ($\lambda_{c1}$).

Table 6.2 shows the value of the optimal ratio of treatment to control fish, $R_{t1}/R_{c1}$, for a range of values of $S$ and $\lambda_{c1}$. For example, when $S = 0.9$ and $\lambda_{c1} = 0.20$, we have

$$\frac{R_{t1}}{R_{c1}} = \left(\frac{0.9611}{0.8500}\right)^{1/2}$$

$$= 1.063.$$

Table 6.2. – Values of the optimal ratio $R_{t1}/R_{c1}$, for minimizing var($\hat{S}$) under model $H_{1\phi}$, for a range of values of $S$ and $\lambda_{c1}$ (based on formula 6.2).

| | $\lambda_{c1}$ | | | | |
|---|---|---|---|---|---|
| $S$ | 0.10 | 0.15 | 0.20 | 0.25 | 0.30 |
| 0.95 | 1.029 | 1.030 | 1.032 | 1.034 | 1.037 |
| 0.90 | 1.060 | 1.063 | 1.067 | 1.072 | 1.076 |
| 0.85 | 1.094 | 1.099 | 1.105 | 1.111 | 1.119 |
| 0.80 | 1.130 | 1.138 | 1.146 | 1.155 | 1.165 |
| 0.75 | 1.171 | 1.180 | 1.190 | 1.202 | 1.215 |
| 0.70 | 1.215 | 1.226 | 1.239 | 1.254 | 1.270 |
| 0.50 | 1.453 | 1.475 | 1.500 | 1.528 | 1.558 |

Thus, for every 100 control fish, one would release 106 treatment fish. We believe this is not practical in the field. Moreover, the optimality of this allocation only applies exactly under model $H_{1\phi}$ with $S = 0.9$ and $\lambda_{c1} = 0.15$. We do not know if model $H_{1\phi}$ is true; we do not know $S$ or $\lambda_{1c}$. Thus, if $S$ is large, say $S \geq 0.8$, we recommend equal release numbers $R_{t1}$ and $R_{c1}$.

From Table 6.2 we see that optimal allocation to treatment and control groups depends mostly on the true value of $S$. (This result is typical in optimality theory: the optimal experiment requires knowing the exact parameter or parameters that one is trying to estimate.) Some release ratios other than 1:1 could be practical, for example, 2:1 ($R_{t1}/R_{c1} = 2.0$), 3:2 (1.500), 4:3 (1.333), or 5:4 (1.25). If, for example, $S$ was believed to be about 0.5, one should use a ratio of 3:2 treatment to control fish.

A quick rule of thumb is that $R_{t1}/R_{c1}$ should be approximately equal to $1/\sqrt{S}$; thus, at $S = 2/3$, 5:4 is about optimal. However, even for $S = 0.667$, a 1:1 allocation results in a se($\hat{S}$) that is not much larger than what is achieved at optimal allocation. With $\gamma = R_{t1}/R_{.1}$, the variance of $\hat{S}$ as a function of $\gamma$ is

$$\text{var}(\hat{S}) = \frac{c(S)^2}{R_{.1}}\left[\frac{1}{\gamma}\left(\frac{1}{S\lambda_{c1}} - 1\right) + \frac{1}{1-\gamma}\left(\frac{1}{\lambda_{c1}} - 1\right)\right].$$

Using this formula, we computed the results below for $S = 0.667$ and $\lambda_{c1} = 0.20$:

| $\gamma$ | $R_{t1}/R_{c1}$ | var($\hat{S}$)/[optimal var($\hat{S}$)] |
|---|---|---|
| 0.500 | 1.000 | 1.015 |
| 0.560 | 1.275 | 1.000 |
| 0.600 | 1.500 | 1.006 |
| 0.670 | 2.000 | 1.051 |

The unknowns $c$ and $R_{.1}$ drop out of the ratio

$$\frac{\text{var}(\hat{S})}{\text{optimal var}(\hat{S})}.$$

In this example, the optimal ratio is nearly 5:4, yet using 1:1 or even 2:1 results in only a trivial loss in precision.

The conclusion is that a release ratio of 1:1 is safe; we recommend it unless one knows $S$ is <0.4. If $S < 0.4$, then use a release ratio of 2:1 for $R_{t1}$:$R_{c1}$.

*6.3.2.2. The case of more than two groups.* – With three or more groups the problems of allocation of release numbers to $R_{v1}$, $v = 1, ..., V$ are more difficult. The problems are not primarily of mathematical difficulty, but rather of deciding on appropriate criteria to use as a basis for determining optimal allocation. There are now $V$ - 1 independent treatment effects, whereas for $V = 2$ we dealt with only one treatment effect. One simple situation is to have one control and then compare each treatment release to that control. For this purpose, we let the "last" group, $V$, be the control group (i.e., $R_{c1} = R_{V1}$) and define the $v$th treatment effect as

$$S_v = \frac{\phi_{v1}}{\phi_{c1}}, \quad v = 1, ..., V - 1.$$

The second situation that we will consider here corresponds to a series of releases at different locations along the river, with release 1 being upriver from release 2, which is upriver from release 3, and so forth. Treatment effects of interest are then

$$S_v = \frac{\phi_{v1}}{\phi_{v+1,1}}, \quad v = 1, ..., V - 1.$$

This situation corresponds (roughly) to a system-wide study.

Now there are $V$ - 1 sampling variances, var($\hat{S}_v$), as well as covariances among these $\hat{S}_v$, and we have to specify an optimality criterion; the possibilities are endless, so we consider here only minimizing the sum of the $V$ - 1 variances. It is also easy to deal with a weighted sum, if some treatment effects are deemed more important than others. In general, one could use numerical optimization (i.e., nonlinear programming) for more complex objective functions.

For the first situation, $S_v = \phi_{v1}/\phi_{c1}$ and, under model $H_{1\phi}$,

$$\text{var}(\hat{S}_v) = \frac{A_v}{R_{v1}} + \frac{B_v}{R_{c1}}, \qquad v = 1, \ldots, V$$

with

$$A_v = c(S_v)^2 \left( \frac{1}{S_v \lambda_{c1}} - 1 \right),$$

and

$$B_v = c(S_v)^2 \left( \frac{1}{\lambda_{c1}} - 1 \right).$$

Subject to $R_{.1} = R_{11} + R_{21} + \cdots + R_{V1}$ being fixed, we want to minimize

$$\sum_{v=1}^{V-1} \text{var}(\hat{S}_v)$$

in terms of the allocation ratios

$$\gamma_v = \frac{R_{v1}}{R_{.1}}, \qquad v = 1, \ldots, V.$$

An alternative representation for the solution is in terms of the ratios

$$\frac{\gamma_v}{\gamma_V} = \frac{R_{v1}}{R_{c1}}, \qquad v = 1, \ldots, V - 1.$$

The objective function to be minimized is now

$$f(\gamma_1, \ldots, \gamma_{V-1}) = \sum_{j=1}^{V-1} \frac{1}{R_{.1}} \left( \frac{A_j}{\gamma_j} + \frac{B_j}{\gamma_V} \right),$$

where

$$\gamma_V = 1 - \sum_{j=1}^{V-1} \gamma_j .$$

Take the $V - 1$ partial derivatives of $f(\gamma_1, ..., \gamma_{V-1})$, with respect to the $\gamma_v$, to obtain the equations to be solved:

$$\frac{1}{R_{.1}} \left[ \frac{-A_v}{(\gamma_v)^2} + \sum_{j=1}^{V-1} \frac{B_j}{(\gamma_V)^2} \right] = 0 , \quad v = 1, ..., V .$$

The solution for $\gamma_v/\gamma_V$ is obvious; thus, we get

$$\frac{R_{v1}}{R_{c1}} = \left( \frac{A_v}{\sum_{j=1}^{V-1} B_j} \right)^{1/2} , \quad v = 1, ..., V - 1 .$$

A little more algebra yields

$$R_{c1} = R_{.1} \left[ 1 + \sum_{v=1}^{V-1} \left( \frac{A_v}{\sum_{j=1}^{V-1} B_j} \right)^{1/2} \right]^{-1} .$$

Thus, we can compute $R_{c1}$ given $R_{.1}$, then get the $R_{11}, ..., R_{V-1,1}$.

It is important to remember that this is the solution to optimal allocation for given $S_v = \phi_{v1}/\phi_{c1}$, $v = 1, ..., V - 1$ under model $H_{1\phi}$, and for just one simple optimality criterion. The unknown constant $c$ will drop out of these ratios. However, we are still left with

$$\frac{R_{v1}}{R_{c1}} = \frac{S_v}{\left( \sum_{j=1}^{V-1} (S_j)^2 \right)^{1/2}} \left( \frac{\frac{1}{S_v} - \lambda_{c1}}{1 - \lambda_{c1}} \right)^{1/2} , \quad v = 1, ..., V - 1 .$$

A useful approximation to this solution, provided $\lambda_{c1}$ is not too large (say, $\lambda_{c1} \leq 0.3$), is

$$\frac{R_{v1}}{R_{c1}} = \left( \frac{S_v}{\sum_{j=1}^{V-1} (S_j)^2} \right)^{1/2} .$$

If the $S_v$ do not vary too much (i.e., all $S_v$ are in the range 0.6 to 0.9), then a suitable allocation would be to have the ratio $R_{v1}/R_{c1}$ constant over $v = 1, ..., V - 1$. As with the case of $V = 2$ groups, the variance of $\hat{S}_v$ will be only weakly dependent on the allocation, so little loss in precision is seen with allocations that are only approximately optimal. This gives us the useful approximations

$$\frac{R_{v1}}{R_{c1}} = \frac{1}{\sqrt{V-1}} ;$$

$$R_{c1} = R_{.1} \left[ 1 + \sqrt{V-1} \right]^{-1} .$$

Thus, if one had four groups, a good starting allocation would be

$$R_{v1} = (0.577) R_{c1} ;$$

$$R_{c1} = (0.366) R_{.1} .$$

For practical purposes, we take this as

$$R_{v1} = \frac{2}{9} R_{.1} , \quad v = 1, 2, 3;$$

$$R_{c1} = \frac{R_{.1}}{3} .$$

For example, if $R_{.1}$ was taken as 90,000 fish, then we get

$$R_{c1} = 30{,}000 ;$$

$$R_{v1} = 20{,}000 , \quad v = 1, 2, 3 .$$

One conclusion here is that, with one control group and multiple treatments, optimal allocation requires proportionally more fish in the control group than in the treatment groups.

Now we consider the situation where treatment group $(v + 1)$ is the control for group $v$; thus,

$$S_v = \frac{\phi_{v1}}{\phi_{v+1,1}}, \quad v = 1, ..., V - 1.$$

This could, of course, be just an alternative way of defining the treatment effect from an experiment where each $\phi_{v1}/\phi_{V1}$ is also of interest. The two formulations will, however, give different optimal allocation results. Now

$$\text{var}(\hat{S}_v) = (S_v)^2\left[\frac{A_v}{R_{v1}} + \frac{A_{v+1}}{R_{v+1,1}}\right], \quad v = 1, ..., V - 1,$$

and

$$A_v = c\left[\frac{1}{\lambda_{v1}} - 1\right], \quad v = 1, ..., V - 1.$$

We take as our objective the minimizing of

$$\sum \text{var}(\hat{S}_v) = \frac{1}{R_{.1}} \sum_{j=1}^{V-1} (S_v)^2 \left[\frac{A_j}{\gamma_j} + \frac{A_{j+1}}{\gamma_{j+1}}\right],$$

where, as before,

$$\gamma_j = \frac{R_{j1}}{R_{.1}}, \quad j = 1, ..., V - 1;$$

$$\gamma_V = 1 - \gamma_1 - \cdots - \gamma_{V-1}.$$

Now take the $V$ - 1 partial derivatives of the above $\Sigma \text{var}(\hat{S}_v)$ with respect to $\gamma_1, ..., \gamma_{V-1}$ to get the equations

$$(S_1)^2 \frac{A_1}{(\gamma_1)^2} = (S_{V-1})^2 \frac{A_V}{(\gamma_V)^2};$$

$$\left[(S_{j-1})^2 + (S_j)^2\right] \frac{A_j}{(\gamma_j)^2} = (S_{V-1})^2 \frac{A_V}{(\gamma_V)^2}, \quad j = 2, ..., V-1.$$

The algebra and solution are easier if we use symbols $S_0$ and $S_V$ and define them to be $S_0 = 0$, $S_V = 0$. We now define

$$Q_j = \left\{ \left[ \frac{(S_{j-1})^2 + (S_j)^2}{(S_{V-1})^2} \right] \left[ \frac{A_j}{A_V} \right] \right\}^{1/2}, \quad j = 1, ..., V.$$

Note that $Q_V = 1$. The optimal release proportions are

$$\gamma_j = \frac{Q_j}{\sum_{j=1}^{V} Q_j}, \quad j = 1, ..., V.$$

This result is for model $H_{1\phi}$ with the objective function being to minimize the sum of the $\text{var}(\hat{S}_v)$. When there is more than one treatment ($V > 2$), there are many possible optimality criteria. We cannot investigate them all here; however, for a given study, with $V > 2$, extensive analytical and numerical investigations of optimal release ratios could be done.

The optimal $\gamma_j$ for this case and problem formulation depend on $V$ parameters, $S_1, ..., S_{V-1}$ and $\lambda_{V1}$, because $\lambda_{v1} = S_v \cdots S_{V-1}\lambda_{V1}$. The relative ratios, $Q_1, ..., Q_{V-1}$, depend weakly on $\lambda_{V1}$ but strongly on the treatment effects $S_1, ..., S_{V-1}$. The more treatment groups there are, the more sensitive $Q_v$ appears to be to variations in the $S_v$. Consequently, a simple rule of thumb for nearly optimal allocation does not seem possible here. As a starting point to think about these results, note that when all $S_v = 1$, $v = 1, ..., V-1$, then $Q_1 = 1$, $Q_{V-1} = 1$, and $Q_v = \sqrt{2}$, $v = 2, ..., V-2$. One then finds the ratios $\gamma_v$ from the normalized $Q_v$. Finally, we give the $Q_v$ below for the case of $V = 4$, $\lambda_{V1} = 0.16$, and $S_1 = S_2 = S_3 = S$ for several values of $S$:

| $S$ | $Q_1$ | $Q_2$ | $Q_3$ | $Q_4$ |
|---|---|---|---|---|
| 1 | 1 | 1.4 | 1.4 | 1 |
| 0.9 | 1.2 | 1.6 | 1.5 | 1 |
| 0.8 | 1.5 | 1.8 | 1.6 | 1 |
| 0.7 | 1.8 | 2.1 | 1.7 | 1 |

*6.3.2.3. Comments.* – The results on relative allocations to treatment groups were derived assuming model $H_{1\phi}$. They will hold reasonably well even under model $H_{2p}$ because the variance of $S$ is similar under these two models. If some other model holds, then one should explore allocations under that model. Analytical results are only easy to get under model $H_{1\phi}$. For more complex models, numerical procedures are recommended; the EXPECT option in program RELEASE can be used to explore alternative designs and models. The more difficult problem in complex, multiparameter situations is to specify a meaningful criterion to be optimized. Finally, the optimal solution will depend on the unknown parameters. Consequently, simple approximations may be more useful than exact results.

### 6.3.3. Relative Recapture Effort

Once it has been decided how many recapture sites to use; one needs to consider how to allocate recapture effort across sites 2 to $k$. There may not be any choices here due to practical constraints. If recapture is at dams, and one expends maximal effort at each dam, there are no choices to be made. Hence, we address here the case where some tradeoffs are possible about how much recapture effort is allocated to the $k$ - 1 recapture sites (or occasions in the general Jolly-Seber model). Effort will be represented by the capture probabilities, $p_2, ..., p_k$.

With no constraints operating, the optimum is to put all $p_i = 1$. However, resources are always limited, so assume the general constraint

$$C = \sum_{j=2}^{k} c_i p_i ,$$

where C is fixed. Here, $c_i$ can be thought of as cost-per-unit effort at site $i$. A useful case to consider will be all $c_i$ equal; then we are just saying that total capture effort is fixed. Let model $H_{1\phi}$ be true, and assume one wants to minimize var($\hat{S}$). The mathematics get complex. However, it turns out to have at least one simple solution; if the costs at dam 2 are not too large, the variance of $\hat{S}$ is minimized by putting all the recapture effort at dam 2. The problem with this solution (which assumes $H_{1\phi}$ is true) is that one cannot test any hypothesis about the extent of the treatment effect on the equality of, for example, $p_{t2}$ and $p_{c2}$. In fact, minimization of var($\hat{S}$) is a poor criterion for allocating effort to recapture sites.

An appropriate criterion for allocation of effort to recapture sites is to achieve high power of model selection tests, especially TEST 1.T2, which tests model $H_{1\phi}$ versus $H_{2p}$ (nominally it tests $p_{t2} = p_{c2}$). The power of this test can be investigated numerically using the EXPECT option of program RELEASE (see Section 6.3.7).

We considered the reverse problem of minimizing cost given a fixed variance of $\hat{S}$. The problem is not analytically solvable in a sufficiently simple form to give useful results. This allocation problem (to minimize costs) could be solved numerically. It is probably not worth doing, given all the other more critical facets of the design problem.

The general principle that emerges here is that recapture site 2 is an important site; if at all possible, it should get more recapture effort than the others. Thus, for example, if $k = 5$, one should not aim for $p_2 = p_3 = p_4 = p_5$. Rather, have $p_2 > p_3, p_4$, and $p_5$. Substantial numbers of releases are important at site 2 in order to get good power of hypothesis tests. This is especially true if the treatment effect is mostly acute so that $H_{1\phi}$ or $H_{2p}$ are likely the appropriate models.

A reasonable allocation of recapture effort over occasions $j = 2$ to $k$ is

$$\frac{p_2}{p_.} = \frac{\sqrt{k-1}}{\sqrt{k-1} + k - 2};$$

and

$$\frac{p_j}{p_.} = \frac{1}{\sqrt{k-1} + k - 2}, \quad j \geq 3.$$

Here $p_. = p_2 + p_3 + \cdots + p_k$. The ratios $p_j/p_.$ are relative capture rates; one can think of them as also indexing relative recapture effort at each site. With some rounding, the above formulae produce the following relative efforts:

| $j$ | $k = 3$ | $k = 4$ | $k = 5$ | $k = 6$ |
|---|---|---|---|---|
| 2 | 60% | 50% | 40% | 36% |
| 3 | 40% | 25% | 20% | 16% |
| 4 | | 25% | 20% | 16% |
| 5 | | | 20% | 16% |
| 6 | | | | 16% |

In any study where allocation of recapture effort to sites is a critical feature, numerical studies must be done to find a good allocation scheme.

In the extreme case, think of allowing some $p_i$ to be zero. This consideration raises the question of how many recapture sites to use; $k = 3$ is a minimum in order to test the crucial hypothesis of $p_{t2} = p_{c2}$. If all effort could be concentrated at two recapture sites, appropriately located, $k = 3$ would be excellent when the effect is mostly acute and the only releases at site 2 are recaptures made there. If capture probabilities (i.e., effort) are unavoidably low at any site, then increase the number of sampling sites.

### 6.3.4. Relative Allocation of Effort to Releases and Recaptures

There is a tradeoff between number of animals released versus total effort expended to recapture them. If it is easy (inexpensive) to release large numbers but difficult (expensive) to recapture them, then release thousands or tens of thousands and settle for low recapture rates. However, as recapture probabilities increase, release fewer individuals and still achieve the same precision of $\hat{S}$ with some test powers. If it is difficult to release large numbers of animals but easy to recapture them, then a good study should also be possible. When both options are difficult, then evaluate carefully any possible tradeoff between numbers released versus recapture efforts.

First we give an example of the relationship between release numbers and recapture rates. Under model $H_{1\phi}$, with $R_{t1} = R_{c1}$, the theoretical ($c = 1$) coefficient of variation of $\hat{S}$ is

$$[\mathrm{cv}(\hat{S})]^2 = \frac{1}{R_{\cdot 1}}\left[\frac{1}{\lambda_{c1}}\left(\frac{1}{S} + 1\right) - 2\right].$$

For illustrative purposes, let $\mathrm{cv}(\hat{S}) = 0.025$ and $S = 0.8$. The following pairs of values will then all produce a cv of 0.025:

| $\lambda_{c1}$ | $R_{\cdot 1}$ |
|---|---|
| 0.01 | 356,800 |
| 0.05 | 68,800 |
| 0.10 | 32,800 |
| 0.15 | 20,800 |
| 0.20 | 14,800 |
| 0.40 | 5,800 |
| 0.80 | 1,300 |
| 1.00 | 400 |

These pairs of points lie along the hyperbola described by

$$R_{.1} = \frac{3{,}600}{\lambda_{c1}} - 3{,}200.$$

In general, there is an inverse relationship between sampling effort (capture probabilities) and release numbers.

The above example illustrates that it is highly desirable to avoid low recapture rates; inordinate numbers of animals ($R_{.1} = R_{t1} + R_{c1}$) are needed at low capture probabilities. Also note that a given proportional increase in $\lambda_{c1}$ corresponds to a greater proportional decrease in $R_{.1}$, especially at high capture rates. When $\lambda_{c1}$ increases 5 × from 0.01 to 0.05, $R_{.1}$ decrease 5.19 ×. Some cases of $\lambda_{c1}$ doubling and the corresponding proportional decreases of $R_{.1}$ follow:

| Change in $\lambda_{c1}$ from | to | Proportional decrease in $R_{.1}$ |
|---|---|---|
| 0.05 | 0.1 | 2.097 |
| 0.1 | 0.2 | 2.216 |
| 0.2 | 0.4 | 2.552 |
| 0.4 | 0.8 | 4.462 |

Thus, in the tradeoff between increasing capture probabilities and increasing release numbers, the preferred strategy should be to increase capture probabilities and allow a corresponding decrease in $R_{.1}$. At the least, we recommend that first consideration be given to doing all that is possible to achieve high capture probabilities before adopting the "brute-force" approach of releasing as many animals as possible at site 1.

If the cost of releasing fish and recapturing them is known, then fix the precision of $\hat{S}$ and find the values of $R_{.1}$ and $\lambda_{c1}$ that give minimum cost. What typically happens is that the recapture effort gets set in advance and then $R_{t1}$ and $R_{c1}$ are chosen to achieve the desired precision. We considered this approach but decided useful formulae could not be given without a specific context. Just the matter of cost functions themselves cannot be dealt with satisfactorily in the abstract (see Skalski 1985). Given a specific study, however, it is possible, in principle, to optimize over tradeoffs such as numbers released versus recapture effort.

Finally, we note a particular area where these ideas would be useful to pursue: the tradeoffs associated with using high technology marks, such as PIT or radio tags (Stier and Kynard 1986), versus batch marks. We would not be surprised to achieve a doubling or tripling of recapture rates with PIT tags and even greater corresponding reduction in releases. This reduction in release numbers needed might make the use of PIT tags not only feasible but quite attractive.

### 6.3.5. Numbers of Releases

Determining absolutes, such as the total number of fish to release, $R_{.1}$, is paradoxically both the simplest and hardest aspect of the problem. It is mathematically simple, but the results depend critically on many unknowns and cannot easily be made reliable. Consider that model $H_{1\phi}$ holds and our primary objective is to estimate $S$ with good precision. Then, from the formula for var($\hat{S}$), the "solution" is

$$R_{.1} = \frac{2c}{[\mathrm{cv}(\hat{S})]^2}\left[\left(\frac{1}{S}+1\right)\frac{1}{\lambda_{c1}}-2\right]. \tag{6.3}$$

This is predicated on using $R_{t1} = R_{c1}$, which is certainly appropriate if $S \geq 0.8$. The general formula, based on using the optimal ratio of $\gamma = R_{t1}/R_{.1}$, is

$$R_{.1} = \frac{c}{[\mathrm{cv}(\hat{S})]^2}\left[\frac{1}{\gamma}\left(\frac{1}{S\lambda_{c1}}-1\right)+\left(\frac{1}{1-\gamma}\right)\left(\frac{1}{\lambda_{c1}}-1\right)\right]. \tag{6.4}$$

The determination of $R_{.1}$ is actually the final consideration. First, determine relative quantities like optimal $\gamma$ and allocation of effort to sites, as well as number of recapture sites, and optimize $\lambda_{c1}$. As the last step, plug these quantities into equation (6.4), along with the target cv($\hat{S}$) and the value of $c$. $R_{.1}$ depends critically on $\lambda_{c1}$ and c, neither of which, especially $c$, are well known. The survival rate $S$ is often known well enough that $R_{.1}$ does not critically depend on it (e.g., if one believes $0.8 < S < 1$, just take $S = 0.9$). Finally, the number of released fish needed also depends on the true model (and protocol). For models $H_{1\phi}$ and $H_{2p}$, equation (6.4) can be used. For model $H_{2\phi}$, substantially more releases are needed.

As an example, let $S = 0.9$, $\lambda_{c1} = 0.2$, and our target $\mathrm{cv}(\hat{S}) = 0.02$. This corresponds to an approximate 95% CI of about 0.86 to 0.94 on $S$. Using $R_{t1} = R_{c1}$, then from equation (6.3)

$$R_{.1} = c\,\frac{17.1111}{[\mathrm{cv}(\hat{S})]^2}$$

$$= c\ 42{,}778.$$

If binomial (or multinomial) variation would hold, then $c = 1$ would be used. However, experience has shown that empirical variances often exceed theoretical variances, especially for count data, even when the model structure is reasonable. It is common, especially in

ecological and environmental sampling, to find that empirical standard errors are double the theoretical errors (this corresponds to $c = 4$). Unfortunately, it is quite possible to have this variance inflation be as large as $c = 9$. In a carefully controlled study, it can be hoped to have $c \leq 4$. Therefore, in the previous example, the use of $c = 1$, hence, $R_{.1} = 43{,}000$, is likely to fail to achieve $\text{cv}(\hat{S}) = 0.02$. A more realistic sample size is $R_{.1} = 4(42{,}778) = 171{,}112$; rounded off, $R_{.1} = 170{,}000$ and $R_{t1} = R_{c1} = 85{,}000$.

Despite the uncertainties involved, it is still better to go through planning procedures such as the above, rather than just guess at a sample size. This procedure can be used to explore the relative sensitivity of results to the different factors. What if $\text{cv}(\hat{S}) = 0.03$ were acceptable? Then

$$R_{.1} = c\ 19{,}012\,,$$

which clearly allows a big decrease in sample size. In general, halving the target $\text{cv}(\hat{S})$ requires quadrupling the sample size.

If we continue with $\text{cv}(\hat{S}) = 0.02$, but explore sensitivity to $\lambda_{c1}$ and $S$, we have the results below:

| $S$ | $\lambda_{c1}$ | $R_{.1}/c$ |
|---|---|---|
| 0.95 | 0.25 | 31,053 |
| 0.90 | 0.25 | 32,222 |
| 0.85 | 0.25 | 33,529 |
| 0.95 | 0.20 | 41,316 |
| 0.90 | 0.20 | 42,778 |
| 0.85 | 0.20 | 44,412 |
| 0.95 | 0.15 | 58,421 |
| 0.90 | 0.15 | 60,370 |
| 0.85 | 0.15 | 62,549 |

Clearly, $\lambda_{c1}$ or, more precisely, capture probabilities, are a critical factor in selection of sample size.

Often the practical and financial constraints are such that the design consists of using the largest affordable sample size. Then these calculations should be reversed to compute precision, given $R_{.1}$, using

$$[\text{cv}(\hat{S})]^2 = c\,\frac{2}{R_{.1}}\left[\left(\frac{1}{S}+1\right)\frac{1}{\lambda_{c1}} - 2\right].$$

The above is still assuming model $H_{1\phi}$. If the budget would allow only $R_{.1} = 50{,}000$, then we have (for $S = 0.9, \lambda_{c1} = 0.2$)

$$[\text{cv}(\hat{S})]^2 = c\,\frac{2}{50{,}000}\left[\left(\frac{1}{.9}+1\right)\left(\frac{1}{.2}-2\right)\right]$$

$$= c\ (0.0003422)\,,$$

or

$$\text{cv}(\hat{S}) = \sqrt{c}\,(0.0185).$$

We see that 50,000 is acceptable (i.e., $\text{cv}(\hat{S}) \sim 0.02$) only if the theoretical variances hold (i.e., $c = 1$). Realistically, one should allow $c = 4$; thus, $\text{cv}(\hat{S})$ is more likely to be $\geq 0.037$; hence, a sample size of 50,000 is likely to cause a lower precision for $\hat{S}$ than desired.

An additional complication occurs when the design uses multiple lots (which it should) and the investigator gets $\text{var}(\hat{S})$ empirically. Then confidence intervals are based on only a few degrees of freedom, and, hence, are wider than intervals based on theoretical variances (even after inflation by $c$).

These types of precision and sample size calculations can be done assuming a model other than $H_{1\phi}$. It is then easiest to use the EXPECT option of PROC SIMULATE in program RELEASE to compute theoretical standard errors. Note that no variance inflation factor is used by RELEASE; users have to either accept the theory, $c = 1$, or make the simple modification by hand.

### 6.3.6. Numerical Evaluation of Some Design Features: an Example

The type of tradeoffs and total effort considerations in previous sections of this chapter can be investigated numerically using the EXPECT option of PROC SIMULATE in program RELEASE. Chapter 3.6 gives background details. Also, EXPECT has been used in Part 5. Consequently, rather than again explain EXPECT per se, we give here an example of its use. We encourage the user to enter the information in Tables 6.3-6.6 into RELEASE on their computer system and obtain the full output (none of which we give here). Such outputs will aid in understanding the material and in making comparisons. Let $k = 6$ and assume model $H_{1\phi}$ holds with $\phi_{t1} = 0.72$, $\phi_{c1} = 0.85$, $\phi_3 = \phi_4 = \phi_5 = 0.9$, $p_2 = 0.05$, $p_3 = 0.02$, $p_4 = 0.07$, $p_5 = 0.05$, and $p_6 = 0.09$. We use $R_{t1} = R_{c1}$ and, for convenience, set these to 50,000. Table 6.3 shows the inputs to RELEASE for this run. Note that treatment effect is $S = 0.847$, and the complete capture history protocol is used.

From the output, we find that the noncentrality parameter of the test for a treatment effect (TEST 1) is $\delta = 127.34$, with 9 df. Thus, from Table 3.4, the power of this test is one. Also, the theoretical standard error of $\hat{S}$ is $\text{se}(\hat{S}) = 0.0125$ (cv = 0.015). If $c = 4$, then the actual $\text{se}(\hat{S})$ would be $2 \times 0.0125 = 0.025$, which is probably acceptable. So one knows that in a carefully done study, if these are about the correct $\phi$ and $p$ parameters, results will be reasonable with $R_{t1} = R_{c1} = 50{,}000$ if model $H_{1\phi}$ holds.

Table 6.3. – Inputs to program RELEASE to generate information on theoretical standard errors and test powers for the parameter values shown; the complete capture history protocols is used.

```
proc title example of using EXPECT to help in study design, model H1phi;
proc simulate expect occasions=6 groups=2;
phi(1)=.72 .85;
phi(2)=.9;
phi(3)=.9;
phi(4)=.9;
phi(5)=.9;
p(2)=.05;
p(3)=.02;
p(4)=.07;
p(5)=.05;
p(6)=.09;
R=50000 50000;
```

We now vary the conditions and see what happens. Table 6.4 shows the inputs if we assume model $H_{2p}$ holds with $p_{t2} = 0.03$ and $p_{c2} = 0.05$, rather than model $H_{1\phi}$. So now there is a treatment effect on capture probabilities at dam 2. Using the output of RELEASE, we find out a number of things. The test for a treatment effect (TEST 1) still has power of one. In fact, component TEST 1.*R*1 still has power of one, and we find TEST 1.*T*2 also has power of one. This tests for $p_{t2} = p_{c2}$ in this case because there is no treatment effect on any parameters that apply downriver from dam 2. If the parameters in Table 6.4 were the true parameters of a study, the data analysis would, with high probability, select model $H_{2p}$ as the appropriate model.

Under model $H_{2p}$, $\hat{S}$ is unbiased and has theoretical se($\hat{S}$) = 0.0140. Although the se($\hat{S}$) has increased, it is only a moderate increase (over se($\hat{S}$) = 0.0125 if $H_{1\phi}$ were true). However, we can look at what would happen if model $H_{2p}$ were true and we used model $H_{1\phi}$ for data analysis. Then we find $E(\hat{S}) = 0.775$ with se($\hat{S}$) = 0.0117. Using the (wrong) model $H_{1\phi}$, we would be seriously misled: $S$ would have a bias of only -0.072; however, the ratio of absolute bias to theoretical se($\hat{S}$) would be 6.2. From this, we infer the 95% CI would almost never cover the true value of $S$. We conclude that we really should use a protocol that allows us to test $p_{t2} = p_{c2}$; this provides protection against a possible treatment effect on $p_2$.

Table 6.4. – Inputs to program RELEASE when the setup of Table 6.3 (model $H_{1\phi}$) is changed to model $H_{2p}$.

```
proc title example of using EXPECT to help in study design, model H2p;
proc simulate expect occasions=6 groups=2;
phi(1)=.72 .85;
phi(2)=.9;
phi(3)=.9;
phi(4)=.9;
phi(5)=.9;
p(2)=.03 .05;
p(3)=.02;
p(4)=.07;
p(5)=.05;
p(6)=.09;
R=50000 50000;
```

We go one step further and ask what happens if there is slight treatment effect on $\phi_2$. Table 6.5 shows the inputs to RELEASE after we modify model $H_{2p}$ of Table 6.4 to allow $\phi_{t2}$ to be 0.86 and $\phi_{c2}$ to remain at 0.9. We find that model $H_{2p}$ is still chosen; the power of the test of $\phi_{t2} = \phi_{c2}$ (TEST 1.R2) is only 0.07. However, under model $H_{2p}$, $E(\hat{S}) = 0.811$, so it is biased, and has theoretical se$(\hat{S}) = 0.0136$.

If model $H_{2\phi}$ is used as the basis of estimates in this case, results are imprecise because now the parameter estimators depend on the number of releases at dam 2, which are in the low thousands ($E(R_{t2}) = 1{,}079$, $E(R_{c2}) = 2{,}125$). Under model $H_{2\phi}$, we can estimate the treatment effects, the true values of which are

$$S_1 = \frac{\phi_{t1}}{\phi_{c1}} = 0.847$$

and

$$S_2 = \frac{\phi_{t2}}{\phi_{c2}} = 0.956.$$

Table 6.5. – Inputs to program RELEASE when the setup of Table 6.4 (model $H_{2p}$) is changed to model $H_{2\phi}$

```
proc title example of using EXPECT to help in study design, model H2phi;
proc simulate expect occasions=6 groups=2;
phi(1)=.72 .85;
phi(2)=.86 .9;
phi(3)=.9;
phi(4)=.9;
phi(5)=.9;
p(2)=.03 .05;
p(3)=.02;
p(4)=.07;
p(5)=.05;
p(6)=.09;
R=50000 50000;
```

Theoretical results under model $H_{2\phi}$ are

$$E(\hat{S}_1) = 0.845, \quad \text{se}(\hat{S}_1) = 0.0729;$$

$$E(\hat{S}_2) = 0.958, \quad \text{se}(\hat{S}_2) = 0.0842.$$

The large $\text{se}(\hat{S}_2)$ under model $H_{2\phi}$ shows why the test of $\phi_{t2} = \phi_{c2}$ has poor power. Estimates of $S_1$ and $S_2$ under model $H_{2\phi}$ are unbiased, but have large standard errors. In contrast, $\hat{S}_1$ under model $H_{2p}$ is biased (-0.037), but has a much smaller standard error (0.0136 versus 0.0729). Thus, the use of estimates under $H_{2p}$ is a reasonable compromise.

Given this situation, can we alter the design to protect ourselves against poor precision if there is a treatment effect on $\phi_2$? Increasing the number of fish released will not be cost-effective. It would take $R_{v1}$ = 2,200,000 under this model $H_{2\phi}$ to achieve a test power of 0.9 for testing $\phi_{t2} = \phi_{c2}$. This conclusion is based on knowing that the noncentrality parameter $\delta$ is 0.236 for $R_{v1}$ = 50,000, and $\delta$ depends directly on the number of releases. That is, to double $\delta$, one must double the releases. To get 90% power with a 1-df chi-square test, we must have (approximately) $\delta$ = 10.5. Therefore, to achieve this 90% power by "brute force," we must increase the releases by a factor of 10.5/0.236, or roughly 44. Thus, we would need $R_{v1}$ = 44(50,000).

We can try altering the design, if it is possible to do the following. Table 6.6 shows inputs for a new design based on the same survival rates and release numbers as in Table 6.5. We dropped site 6, and put that effort, plus a little more, into site 2. Thus, $p_{t2} = 0.09$ and $p_{c2} = 0.15$. The results include the following:

$$E(R_{t2}) = 3{,}241\,,$$

$$E(R_{c2}) = 6{,}375\,,$$

and, under analysis model $H_{2\phi}$,

$$E(\hat{S}_1) = 0.846, \quad \text{se}(\hat{S}_1) = 0.0507$$

and

$$E(\hat{S}_2) = 0.957, \quad \text{se}(\hat{S}_2) = 0.0608\,,$$

while, under analysis model $H_{2p}$,

$$E(\hat{S}_1) = 0.814, \quad \text{se}(\hat{S}_1) = 0.0163\,.$$

Table 6.6. – Inputs to program RELEASE for an alternative design to that in Table 6.5.

```
proc title example of using EXPECT to help in study design, model H2Phi;
proc simulate expect occasions=5 groups=2;
phi(1)=.72 .85;
phi(2)=.86 .9;
phi(3)=.9;
phi(4)=.9;
p(2)=.09 .15;
p(3)=.02;
p(4)=.07;
p(5)=.05;
R=50000 50000;
```

Also, the power of TEST 1.*R*2, which here tests $\phi_{t2} = \phi_{c2}$, is now 0.11. Although we tripled $p_{t2}$ and $p_{c2}$, there is little increase of power for this test, or improved precision of $\hat{S}_1$ (or $\hat{S}_2$) under model $H_{2\phi}$. This new design does not solve our problem. In fact, we probably cannot solve the problem of detecting that model $H_{2\phi}$ holds if it is the true model. This conclusion is both surprising and disturbing.

Similar "what-if" investigations should prove useful when study designs are contemplated.

## 6.4. Multiple Lots

### 6.4.1. Introduction

Multiple lots may be released simply at random (see Chapter 4.3), or as part of a larger design wherein different measured or preset conditions underly the release of each lot. If $n$ multiple treatment-control lots are released at $n$ different times, we let $S_1, ..., S_n$ stand for the true treatment effects at these times. In principle, there are explanatory variables $\underline{x} = x_1, ..., x_a$ that exist so we could write $S_i = f(\underline{x}) + \varepsilon_i$ ($\varepsilon$ being unexplained residual variation). Then a design can be imposed on the multiple released lots in terms of the identified variables $x_1, ..., x_a$. If no such explanatory variables are used in the analysis, or even recorded, we are then treating the $S_i$ as identically distributed random variables. This includes the possibility that $S_i = S$, which means, by definition, that these lots are replicates as regards the treatment effect.

### 6.4.2. Replication

A study with one treatment and one control is a poor one, even if batch sizes are $R_{t1} = R_{c1} = 100{,}000$. A point estimate of $S$ can be computed, but only a theoretical variance estimator is possible unless uniquely numbered marks are used. It is much better, with 200,000 fish, to use 10 paired treatment-control lots, each of size 10,000. These might then be released on 10 consecutive days, either at a standardized time or under standardized (or prescribed) conditions. Replication of paired treatment-control lots also allows one to sample a wider range of conditions or times.

Because the theoretical variance often underestimates true variance, the best approach is to take $n$ replicates of the treatment and control fish lots and estimate the variance empirically. Suppose that we obtain estimates

$$\hat{S}_1, \hat{S}_2, \ldots, \hat{S}_n$$

from equally sized replicates. Then an approximate $(1 - \alpha)$ percent confidence interval for $E(S)$ is given by

$$\overline{\hat{S}} \pm t_{\alpha/2,n-1}\, \hat{\text{se}}(\overline{\hat{S}}),$$

where

$$\overline{\hat{S}} = \sum_{i=1}^{n} \hat{S}_i/n,$$

$$\hat{\text{se}}(\overline{\hat{S}}) = \frac{\hat{\sigma}}{\sqrt{n}},$$

and

$$\hat{\sigma}^2 = \sum_{i=1}^{n} (\hat{S}_i - \overline{\hat{S}})^2/(n - 1).$$

The value $t_{\alpha/2,n-1}$ is the appropriate value from the Student's $t$-table with $n - 1$ df (Table 6.7).

If we assume that the total number of treatment and control fish is fixed, then the important question is how many replicates should be taken. Also, taking replication will probably cost more money than only releasing one batch each of treatment and control fish because the replicates would have to have different batch marks, unless each fish has a unique mark. The variance $\sigma^2$ will not be affected by the number of replicates; thus, the only consideration is how $t_{\alpha/2,n-1}$ changes as $n$ increases. Consideration of a Student's $t$-table (Table 6.7) shows that this quantity decreases from 12.706 to 1.96 as $n$ goes from two to infinity (if we use $1 - \alpha = 0.95$ or a 95% CI). Cost considerations aside, it is better to take a large number of replicates. Notice, however, that the $t$ value has decreased to 2.262 when $n = 10$ and to 2.093 when $n = 20$. Therefore, from a practical viewpoint, it is probably adequate to have about 10 replicate samples. We advise no less than six replicates.

Table 6.7. – Critical values of Student's *t*-test for various levels of confidence $(1 - \alpha)$.

| | $1 - \alpha$ | | |
|---|---|---|---|
| df | 0.90 | 0.95 | 0.99 |
| 1 | 6.314 | 12.706 | 63.657 |
| 2 | 2.920 | 4.303 | 9.924 |
| 3 | 2.353 | 3.183 | 5.841 |
| 4 | 2.132 | 2.776 | 4.604 |
| 5 | 2.015 | 2.571 | 4.032 |
| 6 | 1.943 | 2.447 | 3.707 |
| 7 | 1.895 | 2.365 | 3.499 |
| 8 | 1.860 | 2.306 | 3.355 |
| 9 | 1.833 | 2.262 | 3.250 |
| 10 | 1.812 | 2.228 | 3.169 |
| 11 | 1.796 | 2.201 | 3.106 |
| 12 | 1.782 | 2.179 | 3.055 |
| 13 | 1.771 | 2.160 | 3.012 |
| 14 | 1.761 | 2.145 | 2.977 |
| 15 | 1.753 | 2.131 | 2.947 |
| 20 | 1.725 | 2.086 | 2.845 |
| 30 | 1.697 | 2.042 | 2.750 |
| $\infty$ | 1.645 | 1.960 | 2.576 |

### 6.4.3. Use of More Complex Designs

The simplest study has only one treatment factor, hence, one effect $S$. It is unrealistic to ignore the effect of varying environmental and engineering conditions that can affect $S$. One could release batch fish in lots at different (known) flow rates. Fish size could be included as a design factor. If a dam had, say, five presumably identical turbines, it would be better statistically to release 10,000 fish through each turbine than 50,000 fish through only one turbine.

The inclusion of several known factors in a design with multiple lots is highly desirable, and we recommend it. However, it does not seem possible to make specific suggestions on how to incorporate such additional factors into the design. Each situation requires detailed evaluation (general theoretical developments are given by Grizzle et al. 1969; McCullagh and Nelder 1983). The key point is that more information can be gained out of multiple, smaller lots in a properly designed study than is possible with fewer but larger lots that do not account for the many factors that can influence the experimental outcome. This is a general principle in the design of experiments.

## 6.5. Coping with External Variables

### 6.5.1. Introduction

Thus far, we have concentrated on technical issues in the design of fish survival experiments. There are many nontechnical and nonstatistical matters to keep in mind, including fundamental design principles. Numerous texts exist on experimental design (e.g., Cox 1958) that discuss issues such as selection of experimental units, randomization, blocking, avoidance of confounding variables, use of covariates, and so forth. Most of these references are oriented toward settings where the investigator has a great deal more control over matters than exists in environmental impact studies. Some useful references oriented to statistical aspects of environmental studies include Eberhardt (1976), Green (1979), Armour et al. (1983), Hurlbert (1984), and Stewart-Oaten et al. (1986). We will mention only a few general concerns that this literature covers in more detail (see, in particular, Green 1979).

Two overall design principles are to use multiple experimental units and to not confound variables of interest (treatments here) with external variables (such as time, location, equipment, personnel). Much of the science of study design and conduct has to do with eliminating the influence of external variables by blocking, stratifying, randomizing treatments to units, recording covariates so their influence can be adjusted out in the analysis of the data, standardizing equipment and conditions, and careful training of personnel.

### 6.5.2. Stratification

It is important to consider stratifying the fish by any recognizable variable that might affect the parameters $\phi$, $p$, and $S$. The most important variable is probably fish size, as measured, for example, by length; others are age, sex, strain, and so forth. The purpose of stratification is twofold. First, fish in a lot should be homogeneous in their response parameters. Homogeneity reduces extraneous sources of variation. Second, we want to identify factors that might influence treatment effects. If treatment effect varies by size (over the range of size of interest), we want to know this. We make size a design factor by releasing two or more size classes.

We recommend three size classes if stratification on size is used. Further, we recommend eliminating extremes of size, which provides further control of outliers and should reduce excess variation in the resultant data (i.e., helps achieve $c$ near 1). Within each size stratum, there are $V$ treatment groups of fish; fish are to be assigned randomly to treatment groups within strata.

### 6.5.3. Blocking

Given that there are, say, 10 lots each of treatment and control fish, it is best (statistically) to release lots as pairs with a day or two between releases. It is not good to release all 10 lots at once, as there is no advantage to pairing in this case. With paired treatment-control lots released over a span of time, the generality of the inference also improves.

### 6.5.4. Randomization and Balance

Random selection of experimental units and allocation of treatments is a fundamental principle for valid inferences. However, what can be randomized in these studies is constrained. Primarily, one should randomly assign fish to treatments. Randomization is often used to avoid confounding; so is balancing variables. For example, if there are two people marking fish, half of each lot should be marked by each person. Such a procedure avoids the possibility of confounding a person-effect with treatment effects. If the first person marked all the treatment fish and the second person marked all the control fish, marking and treatment effects are confounded and cannot be separated. There are many field details, such as this one, that are important in a study.

### 6.5.5. Timing of Recapture Effort

It takes days or even weeks for all the experimental fish to move past any downstream dam (or capture site). It is therefore critical that sampling at a recapture site begin before any of the fish arrive there and continues until all the study fish have passed that point. If treatment and control fish move together, the concern about the timing of sampling effort is less critical. The only way to know the temporal distribution of the fish passing a capture site is to extend one's effort until they have all passed. If recapture effort had been stopped at a site when 100% of control survivors had passed, but only 90% of treatment survivors had passed, important bias could result in $\hat{S}$.

There is much concern expressed in the literature about the temporal movement distributions of fish. Capture probabilities could change rapidly at a site as flow and other conditions change. This concern is not a problem if movements are independent of treatment status. If treatment and control movements differ, the analyses presented here are still valid if the time-averaged recapture probabilities at the site are the same for treatment and control fish passing that site.

### 6.5.6. Use of Sublots

All the fish in a lot must have a distinguishing batch mark as a minimum level of marking. From a statistical-information point of view, it is far better to give each fish a uniquely coded tag. The use of unique tags allows one to partition the lot into sublots (as well as use more protocols); these sublots provide a useful approximation to true replication, thereby allowing empirical estimation of (within-lot) sampling variance. When unique marks are not possible, a compromise is possible: use more than one distinct batch mark for each lot. Even having only two distinct batch marks in each lot allows a basis for estimating sampling variation if multiple lots are used. Because of the additional information that can be gained by having sublots, we recommend use of two or more (rather than one) distinct batch marks, even within a lot.

### 6.5.7. Partitioning Recaptures by Site

Under the complete capture history protocol, it is possible to partition both the releases and recaptures based on the last digit of the tag number. Such partitioning is not possible under the other sampling protocols. However, the recaptures $m_{wij}$ can be partitioned within a site or occasion in some cases. In many fisheries studies, the fish are recaptured in gatewells within each dam. Recording the number of fish in each gatewell is simple. These partitioned data are useful in variance estimation using the quasi-likelihood methods given in Section 4.1.3. The number of partitions should be moderately large, say six or more.

### 6.5.8. Attention to Detail and Quality Control

Numerous steps and details are involved in any study involving thousands of fish. There are many ways for biases to arise in the handling, transporting, and releasing of fish, as well as in reading and recording of data from recaptures. It is beyond our scope or expertise to deal with these matters here. These sorts of details do not get published in the refereed literature, but there are numerous study reports that provide details on field considerations (e.g., Olson and Kaczynski, unpublished report, 1980; Heinle and Olson, unpublished report, 1981; Ruggles and Collins, unpublished report, 1981).

## 6.6. Refining the Design by Simulation

Many issues concerning the design of a survival experiment can be addressed by numerical methods, either Monte Carlo simulations or the EXPECT option of PROC SIMULATE in RELEASE. Program RELEASE provides a powerful simulator useful in answering questions about the finer points of experimental design. Most experiments are conducted under resource limitations; these limitations impose compromises in the design.

Once a preliminary design has been reached, the power of the proposed experiment is of concern. What is the probability of finding a treatment effect of size $\Delta$, if it exists? This question concerns the power of the test of $H_0$ versus $H_{1\phi}$ or $H_0$ versus $H_{k\text{-}1,\phi}$. The use of simulation to approximate the power of such tests is straightforward with program RELEASE. Also, analytical approaches are sketched in Chapter 3.6 and are implementable with RELEASE.

The power of the above tests may be particularly important if the treatment survival rate is near one (i.e., little treatment effect). An example might be an experiment to estimate fish survival over a spillway at a hydroelectric dam. For a reasonable number of fish released ($R_{1t}$ and $R_{1c}$), the ability to detect a significant treatment effect may be quite low (e.g., if $S = 0.98$ the power might be only 0.12). Low power should result in either the redesign of the proposed experiment or the decision not to conduct the experiment.

In some experiments, the confidence interval width and expected coverage are of particular concern. Often, these concerns relate to the treatment survival rate $S$, but the $\phi_{vi}$ may also be of interest. Monte Carlo simulation might be used to assess a proposed design by providing an estimate of the expected confidence interval widths and achieved confidence interval coverage. The shape and parameters of the sampling distribution of the estimator can be analyzed by specifying options within the SIMULATE procedure in program RELEASE. The sampling distribution of the $\phi_{vi}$ and $S$ estimators are only normal in large samples; RELEASE provides means to study these distributions for particular sample sizes and parameter values. Results of Monte Carlo simulations also provide a basis for the parametric bootstrap method of establishing confidence intervals (Buckland 1984).

Simulation can be done to explore experimental designs where replicate lots are used (as we recommend strongly). Simulated data can be used to study the possible estimation of variance components, model selection, and interval estimation.

Heterogeneity, a violation of assumption 12, can be simulated to investigate its effect on the power of tests, bias of estimators, and interval estimation. RELEASE allows heterogeneity to be simulated in various ways (see Part 9), and heterogeneity can be hypothesized for the $\phi_{vi}$, $p_{vi}$, or both.

Proper experimental design must consider the tests used in selection of an appropriate model. If a chronic treatment effect is suspected, simulation can be done to explore a model such as $H_{3\phi}$. Such simulation would allow examination of the power of tests and estimator bias if an incorrect model is used. The EXPECT option of PROC SIMULATE can be used to facilitate the scrutiny of a particular experimental design with little computer time.

In all cases, we urge investigators to use program RELEASE in fine tuning the design of experiments. The assessment of a tentative design is made easy if the interactive version of RELEASE is used.