Technical Report

Massachusetts Division of Marine Fisheries
Technical Report TR-25

A Guide to Statistical Sampling for the Estimation of River Herring Run Size Using Visual Counts

Gary A. Nelson

Massachusetts Division of Marine Fisheries
Department of Fish and Game
Executive Office of Environmental Affairs
Commonwealth of Massachusetts

February 2006

Massachusetts Division of Marine Fisheries Technical Report Series

Managing Editor: Michael P. Armstrong

The Massachusetts Division of Marine Fisheries Technical Reports present information and data pertinent to the management, biology and commercial and recreational fisheries of anadromous, estuarine, and marine organisms of the Commonwealth of Massachusetts and adjacent waters. The series presents information in a timely fashion that is of limited scope or is useful to a smaller, specific audience and therefore may not be appropriate for national or international journals. Included in this series are data summaries, reports of monitoring programs, and results of studies that are directed at specific management problems.

All Reports in the series are available for download in PDF format at: http://www.mass.gov/marinefisheries/publications/technical.htm or hard copies may be obtained from the Annisquam River Marine Fisheries Station, 30 Emerson Ave., Gloucester, MA 01930 USA (978-282-0308).

TR-1 McKiernan, D.J., and D.E. Pierce. 1995. **The *Loligo* squid fishery in Nantucket and Vineyard Sound**.
TR-2 McBride, H.M., and T.B. Hoopes. 2001. **1999 Lobster fishery statistics**.
TR-3 McKiernan, D.J., R. Johnston, and W. Hoffman. 1999. **Southern Gulf of Maine raised footrope trawl experimental whiting fishery**.
TR-4 Nelson, G.A, M.P. Armstrong, and T.B. Hoopes. 2001. **Massachusetts 2000 striped bass monitoring report**.
TR-5 Chase, B.C., and A.R. Childs. 2002. **Rainbow smelt (*Osmerus mordax*) spawning habitat in the Weymouth-Fore River**.
TR-6 Chase, B.C., J. Plough, and W. Castonguay. 2002. **A study of the marine resources of Salem Sound, 1997**.
TR-7 Estrella, B.T., and R.P. Glenn. 2001. **Massachusetts coastal commercial lobster sampling program May-November 2000**.
TR-8 Estrella, B.T. 2002. **Techniques for live storage and shipping of American lobster, third edition**.
TR-9 McBride, H.M., and T.B. Hoopes. 2002**. 2000 lobster fishery statistics**.
TR-10 Sheppard, J.J, M.P. Armstrong, D.J. McKiernan and D.E. Pierce 2003. **Characterization of the Massachusetts scup (*Stenotomus chrysops*) fisheries**.
TR-11 Nelson, G.A., and T.B. Hoopes. 2002. **Massachusetts 2001 striped bass fisheries monitoring report**.
TR-12 Howe, A. B., S. J. Correia, T. P. Currier, J. King, and R. Johnston. 2002. **Spatial distribution of ages 0 and 1 Atlantic cod (*Gadus morhua*) off Eastern Massachusetts coast, relative to 'Habitat Area of Special Concern'**.
TR-13 Dean, M.J., K.A. Lundy, and T.B. Hoopes. 2002. **2001 Massachusetts lobster fishery statistics**.
TR-14 Estrella, B.T., and R.P. Glenn. 2002. **Massachusetts coastal commercial lobster trap sampling program, May-November 2001**.
TR-15 Reback, K.E., P.D. Brady, K.D. McLauglin, and C.G. Milliken. 2004. **A survey of anadromous fish passage in coastal Massachusetts: Part 1. Southeastern Massachusetts**.
TR-16 Reback, K.E., P.D. Brady, K.D. McLauglin, and C.G. Milliken. 2004. **A survey of anadromous fish passage in coastal Massachusetts: Part 2. Cape Cod and the Islands**.
TR-17 Reback, K.E., P.D. Brady, K.D. McLauglin, and C.G. Milliken. 2004. **A survey of anadromous fish passage in coastal Massachusetts: Part 3. South Coastal**.
TR-18 Reback, K.E., P.D. Brady, K.D. McLauglin, and C.G. Milliken. 2004. **A survey of anadromous fish passage in coastal Massachusetts: Part 4. Boston and North Coastal**.
TR-19 Nelson, G.A. 2003. **2002 Massachusetts striped bass monitoring report**.
TR-20 Dean, M.J., K.A. Lundy, and T.B. Hoopes. 2003. **2002 Massachusetts lobster fishery statistics**.
TR-21 Nelson, G.A. 2004. **2003 Massachusetts striped bass monitoring report**.
TR-22 Lyman, E.G. and D.J. McKiernan. 2005. **Scale modeling of fixed-fishing gear to compare and quantify differently configured buoyline and groundline profiles: an investigation of entanglement threat**.
TR-23 Dean, M.J., K.A. Lundy, and T.B. Hoopes. 2005. **2003 Massachusetts lobster fishery statistics**.
TR-24 Nelson, G.A. 2005. **2004 Massachusetts striped bass monitoring report**.
TR-25 Nelson, G.A. 2006. **A Guide to Statistical Sampling for the Estimation of River Herring Run Size Using Visual Counts.**

Massachusetts Division of Marine Fisheries
Technical Report TR-25

A Guide to Statistical Sampling for the Estimation of River Herring Run Size Using Visual Counts

Gary A. Nelson

Massachusetts Division of Marine Fisheries
Annisquam River Marine Fisheries Station
30 Emerson Ave.
Gloucester, MA 01930

February 2006

Massachusetts Division of Marine Fisheries
Paul Diodati, Director
Department of Fish and Game
Dave Peters, Commissioner
Executive Office of Environmental Affairs
Stephen R. Pritchard, Secretary
Commonwealth of Massachusetts
Mitt Romney, Governor

Acknowledgements

Funding for this effort was provided by the Massachusetts Division of Marine Fisheries and Sportfish Restoration Funds Grant F-49-T.

CONTENTS

ABSTRACT

In recent years, public interest in the status of river herring in Massachusetts has been growing. Community groups have established visual counting programs at several rivers along the Massachusetts coast, but few, if any, adhere to and use proper statistical sampling techniques required to produce reliable estimates of herring run size. In this document, basic statistical concepts required for the estimation of herring run size, statistical sampling designs that offer alternatives to the sampling requirements of Rideout et al. (1979), and consequences of departures from sampling requirements are reviewed. Based on results of the 2005 herring counting workshop, it is recommended that counting programs follow a two-way stratified random sampling design in which volunteers make 3 ten-minute counts during each of three daily periods (7-11 am, 11-3 pm, and 3-7 pm) from April 1st to mid-June.

Introduction

In Massachusetts, more than 100 coastal rivers and streams are home to the anadromous alewife (*Alosa pseudoharengus*) and blueback (*Alosa aestivalis*) herrings. Also known colloquially by anglers as "river herring", these fishes are ecologically-important because they are forage for many marine and freshwater fish predators such striped bass (*Morone saxatilis*), cod (*Gadus morhua*), and yellow perch (*Perca fluvescens*) as well as birds (Loesch, 1987). In addition, they are a key link in the transfer of nutrients from freshwater to marine systems (Mullen et al., 1986).

Monitoring of herring abundances is essential to the management of these important fisheries resources. The Massachusetts Division of Marine Fisheries (DMF) monitors the absolute abundance at two locations in the State: the Essex Dam fish-lift in Lawrence, MA on the Merrimack River and at the Bournedale ladder in Bourne, MA. Run size is computed by counting the number of fish observed in each bucket lift at the Essex Dam, or by using an electronic fish counter at the Bournedale run. Since herring can be passed only through the lift system or through the electronic counter, no statistical sampling design is required to estimate run size because an exact count is achieved.

In recent years, public interest in the status of river herring in Massachusetts has been growing. Community watershed groups have established herring counting programs at several rivers along the Massachusetts coast. Typically, participants count for a short interval of time herring passing at an observation point, usually at the top of a fish ladder. Since passage is observed in samples of counts, this visual counting method requires a statistical sampling design to provide accurate and precise estimates of run size. Some watershed groups try to adhere to the sampling scheme of Rideout et al. (1979), but the required daily coverage and hourly sampling is often not achieved due to insufficient numbers of participants or scheduling difficulties. Other groups have conducted projects without consideration of an appropriate sampling design. Without the adherence to and use of proper sampling techniques, run size estimates derived by these watershed groups may not be useful to biologists and managers responsible for the management of herring in Massachusetts.

The objective of this report is to act as a guide for community watershed groups currently conducting or starting sampling efforts to estimate herring run sizes using visual counts. In this document, basic statistical concepts required for the estimation of herring run size, statistical sampling designs that offer alternatives to the sampling requirements of Rideout et al. (1979), and consequences of departures from sampling requirements are reviewed. If the procedures outlined in this document are followed, then statistically-sound estimates of herring run size will be achieved.

Basic Statistical Concepts

The main objective of statistical sampling is to estimate some characteristic of a population from only a small subset or *sample* of observations. With herring, the quantity of interest is the total number of fish passing during a spring spawning run. Since passage occurs over time, the population to be sampled is the number of herring passing in every unit of time (e.g., 5 or 10-minute intervals) throughout the run. Because every time unit can not be observed due to lack of participants or money, a sample of time units is selected without replacement (the same interval can not be sampled again) and counts are made. The statistic that is estimated and used to extrapolate to the total number passing is the mean number of fish passing per unit time:

$$\hat{\bar{y}} = \frac{\sum_{i=1}^{n} y_i}{n}$$

where $\hat{\bar{y}}$ is the mean number of fish passing per time unit, y_i is the count of herring during the ith time unit, and n is the number of time units sampled (sample size). The symbol $\sum$ means to sum values of all observations. For example, suppose three 5-minute herring counts were made and count values were 20, 3, and 10 fish per interval. Each of the three observations can be referenced by a letter and its position i: y_1 is 20, y_2 is 3, and y_3 is 10. With these references, the numerator signifies $y_1+y_2+y_3$ or 20+3+10. n refers to the total number of observations collected or the last i

if the observations were sequentially numbered from one. To get the total number of fish passing ($\hat{Y}$), the mean is multiplied by the total number of time units, i.e.,

$$\hat{Y} = N \cdot \hat{\bar{y}}$$

This is essentially what is done in every sampling design except the selection of units will differ by design. The hat (^) above the mean and total symbols indicates that the value represents an estimate of a population characteristic, not the true value.

Producing Reliable Estimates

To produce reliable estimates of the mean and total, several statistical assumptions related to accuracy and precision must be met. An estimate is said to be accurate if repeated estimates of the same quantity are centered around the true value (Figures 1A and 1B), whereas those that are distant from the true value are said to be inaccurate or biased (Figures 1C and D). Because all sampling designs are derived from the probability theory (Cochran, 1963), all time units must have equal chances of being selected (e.g., morning time units are equally likely to be selected as evening time units) and selection of one unit must be independent of the selection of another unit to ensure accuracy. In reality, only one estimate is made, but it would still be considered accurate if the estimation procedure is correctly followed.

A biased estimate can be produced either by deviating from sampling theory or by making errors when measuring characteristics. With the estimation of herring run size, it is assumed that selection of time intervals is made either randomly or systematically (Jessop and Harvie, 1990), but if deviations occur, accuracy of the estimates can not be guaranteed. Similarly, it is assumed that all individual herring are correctly identified and counted, but this may not be true if murky water hinders visibility, other herring-like species are

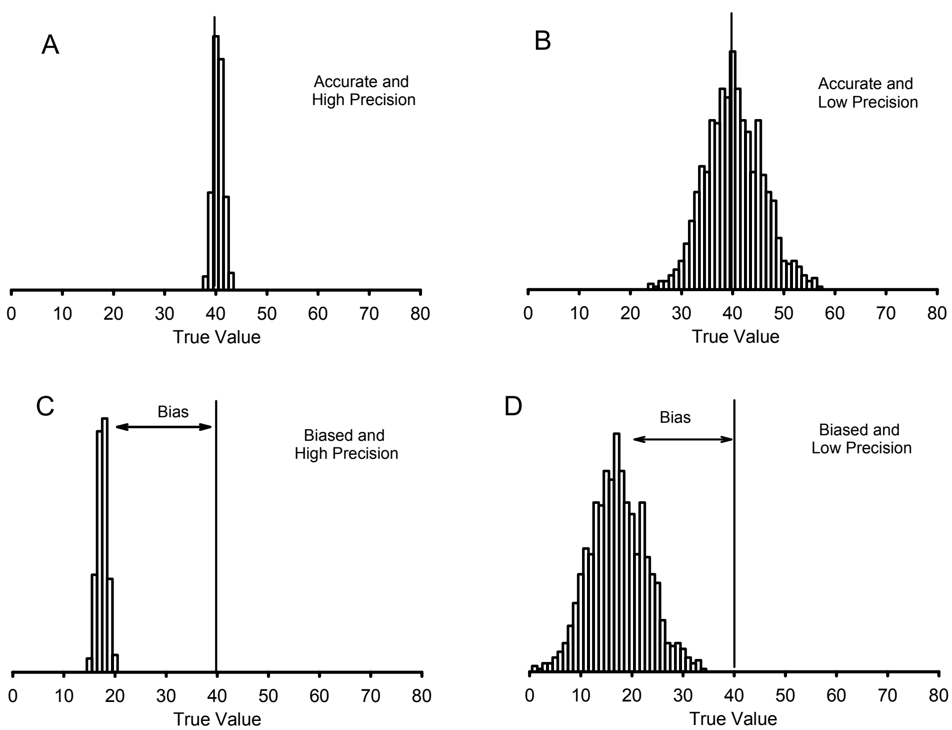

Figure 1. Diagrams demonstrating the concepts of accuracy and precision in the estimation of population characteristics.

migrating concurrently, or too many herring pass at once. In such events, the interval counts will be falsely high or falsely low and will produce inaccurate estimates of daily totals.

An estimate is said to be precise if repeated estimates of the mean are close to each other (Figures 1A and 1C); whereas those that show lots of scatter are said to be imprecise (Figure 1B and 1D). To ensure reasonably good precision, the number of samples taken must be large enough to reduce the effect of random deviations. In practice, as the sample size increases, the less likely the estimate would be affected by random deviations, and the more certain we are that it will lie close to the true population value. The determination of the sample size needed to attain a certain level of precision is an important aspect of sampling theory that must be considered before sampling begins.

An important measure of variability that is used to derive variance of estimates is the sample variance (s^2), given by

$$\hat{s}^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{\bar{y}})^2}{n-1}$$

It is calculated by first computing the mean, subtracting the mean from each observation, squaring the deviations, summing all squared deviations, and then dividing the sum by the number of observations (n) less one. The sample variance summarizes the differences among characteristics of individuals in a population. In nature, biological processes dictate the amount of variability that is observed. For herring run size estimation, variation of observed numbers of herring passing at a single location will depends on factors influencing migration, such as water temperature and velocity, the reproductive and energy state of herring, the ability of herring to find and use fish ladders, etc. (see Mullen et al., 1986). If herring pass relatively consistently throughout a day, differences among sample observations (variation) will be low (Figure 2A). If migration becomes more concentrated during specific times of the day, variation increases (Figures 2B-C). The highest variation would occur if all herring passed in a relatively short period of time (Figure 2D).

The sample variance is important because the variance of the total is computed directly from it:

$$\text{var}(\hat{Y}) = N \cdot (N-n) \cdot \frac{\hat{s}^2}{n}$$

Notice that the var($\hat{Y}$) decreases as the number of samples taken increases. The square-root of the variance is used to produce confidence intervals for the total,

$$\hat{Y} \pm t_{\alpha/2} \cdot \sqrt{\text{var}(\hat{Y})}$$

where t is the two-tailed student-t critical value for α (the allowable probability of error) which provides 100(1- α)% confidence intervals given the degrees of freedom (see Sokal and Rohlf, 1981). A table of t values for α ranging from 0.01 to 0.20 to get 80%-99% confidence intervals is given in Appendix A for various degrees of freedom. The interpretation of these intervals for a given level of confidence (say 95%) is that, if sampling were repeated 100 times and confidence intervals were generated for each sampling event, 95 out of 100 confidence intervals would encompass the true estimate. It cannot be stated that there is a probability of 0.95 that the true mean is contained within any particular observed confidence interval (Sokal and Rohlf, 1981). Large confidence intervals mean that the estimate is not very precise; therefore, increasing the number of observed time units (sample size) or increasing the length of count intervals are the only options for reducing the width of the intervals for a given α.

Review of Sampling Designs

Common sampling designs that can be used for run size estimation are reviewed in standard sampling texts (e.g., Cochran, 1963; Thompson, 2002) as well as Rideout et al. (1979) and Jessop and Harvie (1990). The sampling designs reviewed here are 1) simple random sampling, 2) one-way stratified random sampling, 3) two-way stratified random sampling, 4) stratified systematic sampling and 5) stratified two-stage random sampling. To demonstrate how each design is implemented, schematic examples for two days, broken into 20 minute counting intervals, will be used (Figures 3 -5).

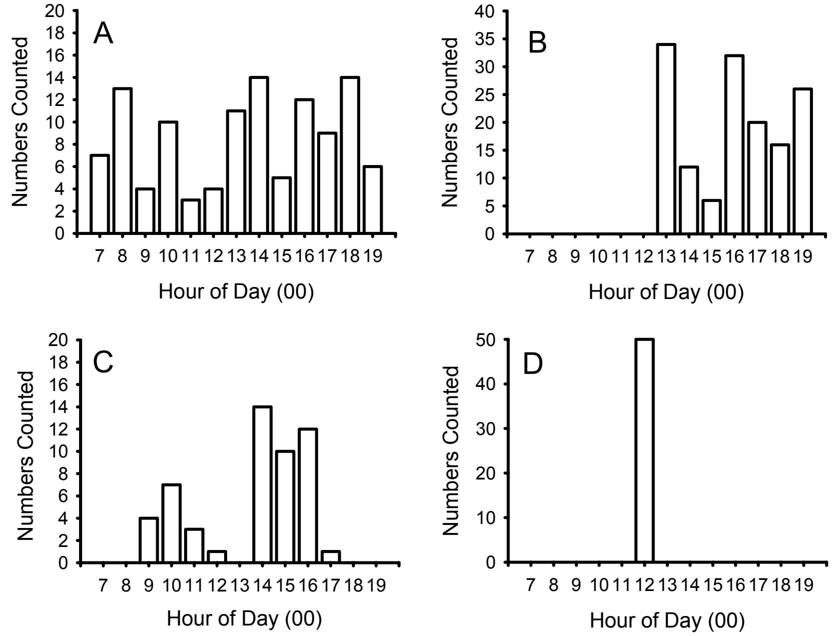

Figure 2. Hypothetical daily herring counts demonstrating possible hourly patterns in migration: A) no pattern, B) afternoon migration, C) bimodal migration, and D) single-hour migration

In simple random sampling (SRS), a sample of the total number of time units in the entire duration of the run is randomly selected without regard to the day of run (Figure 3A). During each interval, herring are counted. The mean number of herring counted per time interval under simple random sampling is

$$\hat{\bar{y}}_{SRS} = \frac{\sum_{i=1}^{n} y_i}{n} \tag{1}$$

and the total run size ($\hat{Y}_{SRS}$) is

$$\hat{Y}_{SRS} = N \cdot \hat{\bar{y}}_{SRS} \tag{2}$$

where y_i is the ith count, n is the number of time units sampled and N is the total number of units during the run. The variance of $\hat{Y}_{SRS}$ is given by:

$$\mathrm{var}(\hat{Y}_{SRS}) = N \cdot (N - n) \cdot \frac{\hat{s}^2}{n} \tag{3}$$

where s^2 is the sample variance defined as

$$\hat{s}^2 = \frac{\sum_{i=1}^{n} (y_i - \hat{\bar{y}}_{SRS})^2}{n-1}$$

Confidence intervals of the total are calculated by

$$\hat{Y}_{SRS} \pm t_{\alpha/2} \cdot \sqrt{\mathrm{var}(\hat{Y}_{SRS})} \tag{4}$$

The advantage of this design is the simplicity of the approach and it may not require volunteers for every day of the run. A disadvantage is that the duration of the run has to be known in advance, otherwise the run size could be under- or over-estimated if sampling is stopped early, or extended beyond the end of the actual migration (the addition of a false zero count will affect the overall estimate). Another disadvantage is that, due to the random chance, the selected intervals could be clumped during certain time periods of the run (Figure 3A) and thereby potentially produce estimates that are over- or under-estimated. This design is generally never used in fish counting due to these disadvantages.

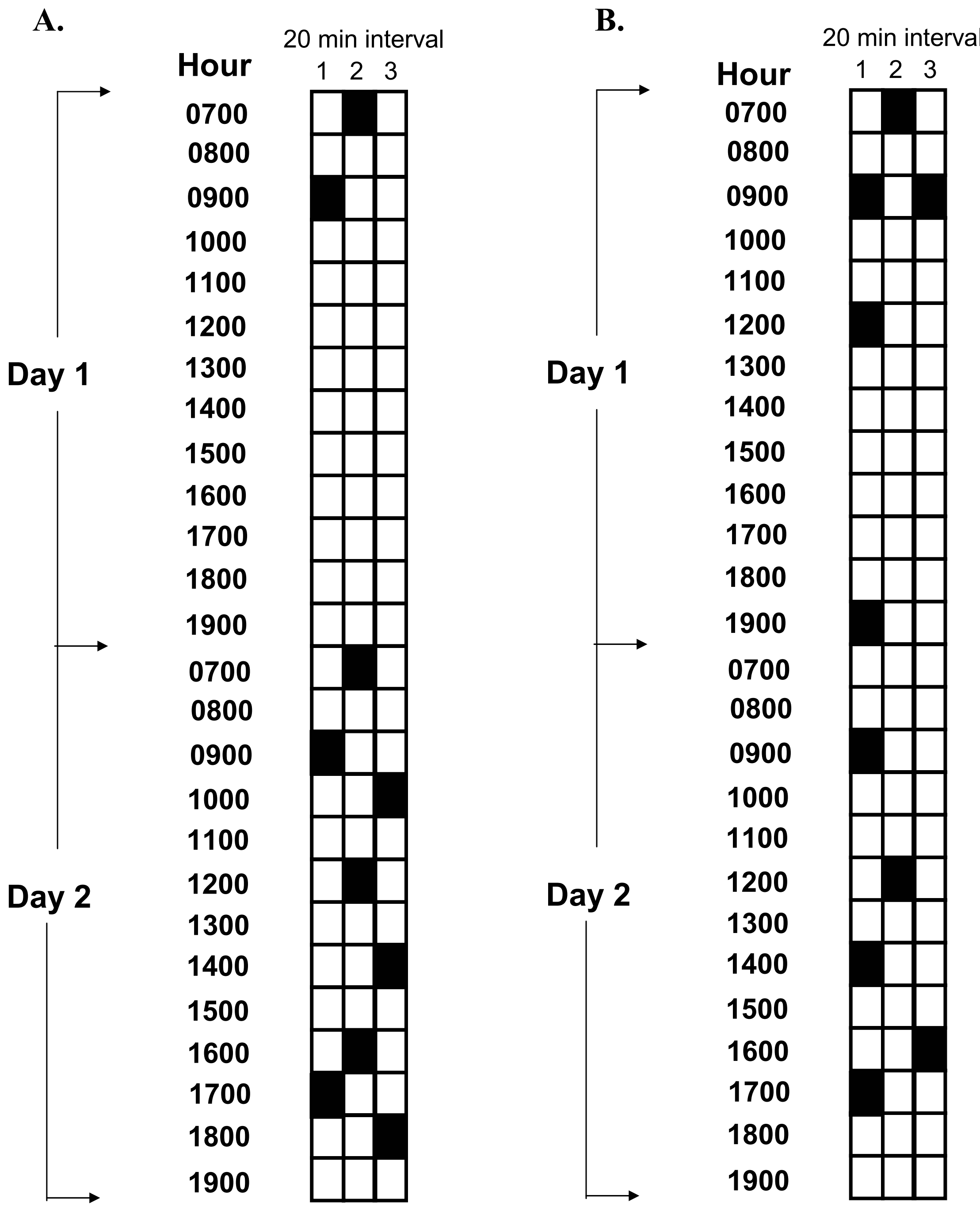

Figure 3. Diagram showing the selection of ten 20-minute intervals under A) simple random sampling and B) one-way stratified (by day) random sampling. In A, notice that the selection of 8 intervals occurs within one day of the run.

In one-way stratified random sampling (St1WRS), selection of time units occurs with regard to each day (the stratum) of the run and, within each day, a sample of the units is selected randomly (Figure 3B). The mean number of herring counted per time unit for each day (k) is

$$\hat{\bar{y}}_k = \frac{\sum_{i=1}^{n_k} y_{k,i}}{n_k} \qquad (5)$$

where $\hat{\bar{y}}_k$ is the mean on day k, $y_{k,i}$ is the ith count on day k, and n_k is the daily number of time units sampled. The total run size ($\hat{Y}_{St1WRS}$) is estimated as

$$\hat{Y}_{St1WRS} = \sum_{k=1}^{L} N_k \cdot \hat{\bar{y}}_k \qquad (6)$$

where L is the number of days during the run, and N_k is the total number of time units in day k. The variance of $\hat{Y}_{St1WRS}$ is

$$\mathrm{var}(\hat{Y}_{St1WRS}) = \sum_{k=1}^{L} N_k \cdot (N_k - n_k) \cdot \frac{\hat{s}_k^2}{n_k} \qquad (7)$$

where

$$\hat{s}_k^2 = \frac{\sum_{i=1}^{n_k} (y_{k,i} - \hat{\bar{y}}_k)^2}{n_k - 1}$$

The confidence intervals are given by

$$\hat{Y}_{St1WRS} \pm t_{\alpha/2} \cdot \sqrt{\mathrm{var}(\hat{Y}_{St1WRS})} \qquad (8)$$

The degrees of freedom (*df*) are calculated using the Satterthwaite approximation (Cochran, 1963) which is given by

$$\hat{df} = \frac{(\sum_{k=1}^{L} a_k \cdot \hat{s}_k^2)^2}{(\sum_{k=1}^{L} \frac{(a_k \cdot \hat{s}_k^2)^2}{n_k - 1})} \qquad (9)$$

where

$$a_k = \frac{N_k \cdot (N_k - n_k)}{n_k}$$

If N_k and n_k are the same for each day, then the degrees of freedom are

$$df = (\sum_{k=1}^{L} n_k) - L \qquad (10)$$

With St1WRS, all days of the run have to be sampled. Several advantages of St1WRS are that the clumping of sampled time units across days can not occur, that sampling can occur well beyond the extent of the run because the addition of a daily zero count will not affect the overall estimate, and that the total number of intervals within a day can change over time (e.g., length of daylight changes over time). This stratified design will generally produce more precise estimates of run size than SRS if the within-day variation changes from day-to-day. A disadvantage is that clumping within a day can occur and, if there is an hourly pattern in migration, daily estimates of the number of fish passing could be over- or underestimated. At least two time units must be sampled to produce estimates of variance for each day.

In two-way stratified random sampling (St2WRS) with day and periods defined as strata, selection of time units occurs with regard to each day (the first stratum) and then with regard to each period (hours grouped into strata). For example, periods could be morning (06:00-11:59) and evening (12:00-18:00) hours. Every period, a sample of time units is then selected randomly (Figure 4A). The mean number of herring per time unit of each period of day k is

$$\hat{\bar{y}}_{k,p} = \frac{\sum_{i=1}^{n_{k,p}} y_{k,p,i}}{n_{k,p}} \qquad (11)$$

where $\hat{\bar{y}}_{k,p}$ is the mean, $y_{k,p,i}$ is the ith count during period p on day k and $n_{k,p}$ is the number of time units sampled during period p on day k. The total run size ($\hat{Y}$) under St2WRS is estimated as

$$\hat{Y}_{St2WRS} = \sum_{k=1}^{L} \sum_{p=1}^{P} N_{k,p} \cdot \hat{\bar{y}}_{k,p} \qquad (12)$$

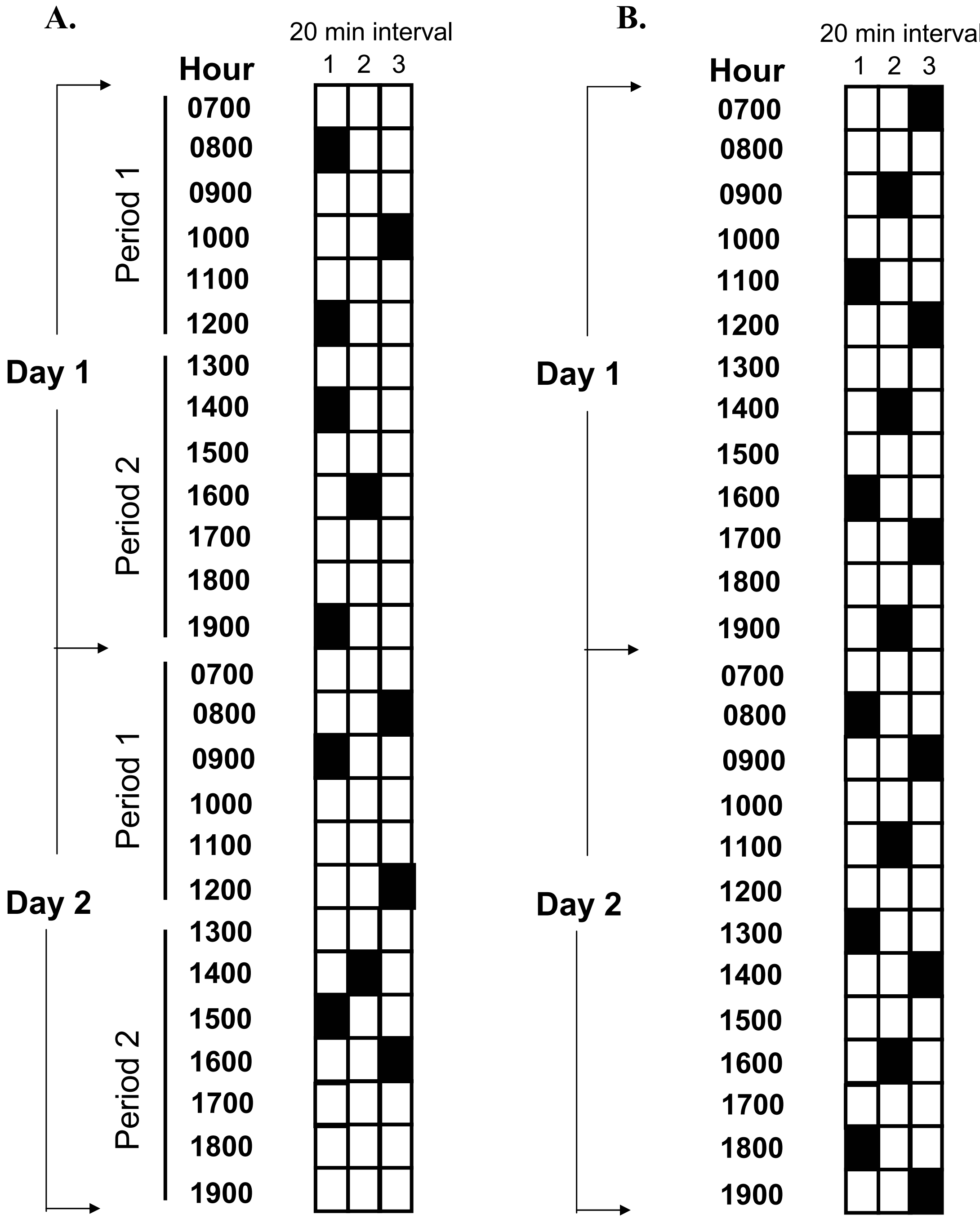

Figure 4. Diagram showing the selection of 20-minute intervals under A) two-way stratified sampling for two period and B) stratified systematic sampling for every 5th interval.

where $N_{k,p}$ is the total number of time units during period p and P is the number of periods during day k, and L is the number of days of the run. The variance estimate is given by

$$\text{var}(\hat{Y}_{St2WRS}) = \sum_{k=1}^{L}\sum_{p=1}^{P} N_{k,p} \cdot (N_{k,p} - n_{k,p}) \cdot \frac{\hat{s}_{k,p}^2}{n_{k,p}} \qquad (13)$$

where

$$\hat{s}_{k,p}^2 = \frac{\sum_{i=1}^{n_{k,p}} (y_{k,p,i} - \hat{\bar{y}}_{k,p})^2}{n_{k,p} - 1}$$

Confidence intervals can be approximated by

$$\hat{Y}_{St2WRS} \pm t_{\alpha/2} \cdot \sqrt{\text{var}(\hat{Y}_{St2WRS})} \qquad (14)$$

If $N_{k,p}$ and $n_{k,p}$ are the same for every period during each day, then the degrees of freedom are

$$df = L \cdot (\sum_{p=1}^{P} n_p - 1) \qquad (15)$$

Otherwise, equation 9 is used to calculate *df* for each day substituting p for k, and the total degrees of freedom are derived by summing *df* over all days.

The advantages of St2WRS are similar to St1WRS. In addition, selection of time units can be spread across each day which minimizes potential clumping, and if there are patterns in migration, this design can further reduce variance of the total estimate. One disadvantage is that at least two samples must be taken every period, so the minimum number of counts made per day is $2P$.

In stratified systematic sampling (StSYS), selection of time units occurs with regard to each day (the stratum) of the run and the units are selected systematically instead of randomly. For systematic selection, the desired sampling interval v (every v^{th} time unit) for a day is first chosen, and then a random time unit j between 1 and v is selected (it is assumed that the daily time units are sequentially numbered from the beginning to end of the day). Next, every interval at the appropriate v spacing from j is selected. For example, if there are thirty-nine 20-minute intervals within a day, and the desired sampling interval is 5, a random unit from the first five intervals of the day is selected and sampled, and then every 5th interval after that is sampled subsequently (Figure 4B). The mean number of herring counted per time interval during day k is estimated as

$$\hat{\bar{y}}_k = \frac{\sum_{i=1}^{n_k} y_{k,i}}{n_k} \qquad (16)$$

and the total run size ($\hat{Y}$) under StSYS is estimated as

$$\hat{Y}_{StSYS} = \sum_{k=1}^{L} N_k \cdot \hat{\bar{y}}_k \qquad (17)$$

The variance of $\hat{Y}_{StSYS}$ is approximately

$$\text{var}(\hat{Y}_{StSYS}) = \sum_{k=1}^{L} N_k \cdot (N_k - n_k) \cdot \frac{\hat{s}_k^2}{n_k} \qquad (18)$$

where

$$\hat{s}_k^2 = \frac{\sum_{i=1}^{n_k} (y_{k,i} - \hat{\bar{y}}_k)^2}{n_k - 1}$$

The confidence intervals are calculated as

$$\hat{Y}_{StSYS} \pm t_{\alpha/2} \cdot \sqrt{\text{var}(\hat{Y}_{StSYS})} \qquad (19)$$

The degrees of freedom are the same as St1WRS. Because of the systematic selection of time units, $\hat{Y}_{StSYS}$ will be an unbiased estimate of the population value only if the ratios N_k/v are integers (e.g., 1,2, etc.). If the ratios N_k/v are not integers (e.g., 1.5, 2.6, etc.) then $\hat{Y}_{StSYS}$ will be biased. To avoid producing biased estimates, the following procedure should be used (Levy and Lemeshow, 1999):

1. Each day, choose a random number between 1 and N_k where N_k is the total number of time units in the day.

2. Divide the random number by the desired sampling interval v. Express this quotient as an integer plus a remainder in fractional form

(e.g., 3⅔).

3. If the numerator of the remainder is 0, take a systematic sample corresponding to v beginning at the vth interval. If the numerator is nonzero, take a systematic sample corresponding to v beginning at the interval equal to the numerator of the remainder.

Equation 18 is appropriate only for estimating the variance of the total run size if herring pass randomly throughout a day. If this assumption is violated, then equation 18 will produce a biased estimate of variance (Cochran, 1963). This is a major disadvantage of systematic sampling (Levy and Lemeshow, 1999).

In stratified two-stage random sampling (St2STRS), selection of time units occurs with regard to each day (the stratum) of the run. Within each day, the time units are grouped into a larger time frame such as an hour. Several hours are then selected randomly (stage 1) and within each selected hour, a sample of time units is randomly taken (stage 2: Figure 5). The mean number of herring counted per time interval within hour i during day k is calculated as

$$\hat{\bar{y}}_{k,i} = \frac{\sum_{j=1}^{m_{k,i}} y_{k,i,j}}{m_{k,i}} \tag{20}$$

and total count for each hour i is

$$\hat{Y}_{k,i} = M_{k,i} \cdot \hat{\bar{y}}_{k,i} \tag{21}$$

The mean number of herring counted per hour is

$$\hat{\bar{\bar{y}}}_k = \frac{\sum_{i=1}^{n_k} \hat{Y}_{k,i}}{n_k} \tag{22}$$

and total count for day k is estimated as

$$\hat{Y}_k = N_k \cdot \hat{\bar{\bar{y}}}_k \tag{23}$$

The total run size ($Y_{St2STRS}$) is computed as

$$\hat{Y}_{St2STRS} = \sum_{k=1}^{L} \hat{Y}_k \tag{24}$$

For the above equations, $m_{k,i}$ is the number of time units sampled within hour i, M_k is the total number of time units within each hour, n_k is the number of hours sampled during day k, N_k is the total number of hours during day k, and L is the number of days of the run.

The variance of $\hat{Y}_{St2STRS}$ is given by

$$\text{var}(\hat{Y}_{St2STRS}) = \sum_{k=1}^{L} N_k (N_k - n_k) \frac{\hat{s}_{k,u}^2}{n_k} + \frac{N_k}{n_k} \sum_{i=1}^{n_k} M_{k,i} (M_{k,i} - m_{k,i}) \frac{\hat{s}_{k,i}^2}{m_{k,i}} \tag{25}$$

where

$$\hat{s}_{k,u}^2 = \frac{\sum_{i=1}^{n_k} (\hat{Y}_{k,i} - \hat{\bar{\bar{y}}}_k)^2}{n_k - 1}$$

and

$$\hat{s}_{k,i}^2 = \frac{\sum_{j=1}^{m_{k,i}} (y_{k,i,j} - \hat{\bar{y}}_{k,i})^2}{m_{k,i} - 1}$$

such that $s^2_{k,u}$ is the sample variance of the hourly totals during day k, and $s^2_{k,i}$ is the sample variance of the time units with each hour. Confidence intervals are approximated by

$$\hat{Y}_{St2STRS} \pm t_{\alpha/2} \cdot \sqrt{\text{var}(\hat{Y}_{St2STRS})} \tag{26}$$

Approximate total degrees of freedom are

$$df = \sum_{k=1}^{L} n_k \cdot (m_k - 1) \tag{27}$$

assuming $m_{k,i}$ is the same within n_k hours.

An advantage of this design is that it offers greater flexibility in the arrivals times of participants over St1WRS. For example, under St1WRS, interval counts, due to random selection, will be spread out over the day, and the time between counting could be large. Few volunteers would be willing to stay over several hours to make those counts. With St2STRS, the volunteers

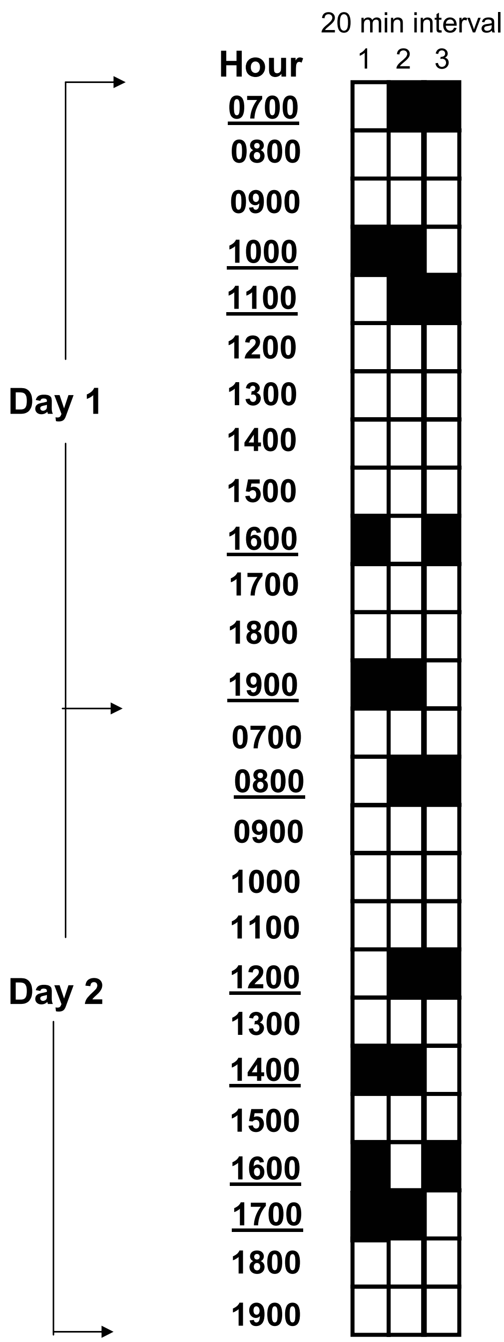

Figure 5. Diagram showing the selection of ten 20-minute intervals under stratified two-stage random sampling for two counts per hour (randomly selected hours are underlined).

would have to stay for 1 hour only (assuming arrival is random) to make two or more random counts. A disadvantage of this design is that, due to the random chance, the selected hours and time units within hours could be clumped, and if there is an hourly pattern in migration, daily estimates of the number of fish passing could be over- or under-estimated. The minimum number of time units that would have to be sampled each day of the run is 4 (2 hours x 2 intervals per hour).

Which Sampling Design Should Be Used?

Which sampling design should be used by watershed groups to estimate herring run size? The answer will depend on the human resources of each group. In the past, groups have tried to adhere to Rideout et al.'s systematic design in which one count (10 minute interval) is made every hour for a total of 13 counts per day. However, volunteers were often unavailable during certain hours or days to accomplish the required sampling (Figure 6). Such deficiencies cause estimates to be biased because not all time units have an equal chance being selected. A better strategy would be for watershed groups to assess realistically the minimum reliable volunteer effort that can be guaranteed daily, and then to select a design that is more suitable for the level of available effort.

Jessop and Harvie (1990) evaluated the accuracy and precision of various sampling designs for estimating population means of herring counts for alewives through a fishway on the Gaspereau River, Nova Scotia. The designs included in their evaluation were SRS, simple systematic sampling (a sample of the total number of time units in the entire duration of the run are selected in a systematic way), St1WRS, and StSYS. They showed that the designs that used random selection (over systematic) with stratification required lower sample sizes to achieve equivalent levels of precision. Therefore, watershed groups should choose St1WRS, St2WRS, or St2STRS. The choice will depend on advantages/disadvantages discussed earlier and on the willingness and numbers of the volunteers to make counts. For instance, if there are numerous volunteers willing to make only one count per visit per day, then St1WRS would be more appropriate. If there are only a few volunteers willing to make more than one count per visit per day, then St2STRS would be more appropriate. But remember, regardless of the stratified design, time units still have to be selected randomly with respect to time!

How Many Time Intervals Should Be Sampled?

The number of time units that have to be sampled will depend upon the desired level of precision for the total run size. A prior estimate of the mean number of fish passing per time units and associated sample variance either for the whole run (for SRS) or for a number of days (for St1WRS, St2WRS, StSYS, or St2STRS) is needed before sample size can be determined. Initial guesses of the estimates can be obtained from scientific papers that have conducted similar studies or from a preliminary study.

For SRS, the sample size necessary to estimate total herring run size to within *r (proportion)* of the true total with *1-α%* confidence is

$$n = \frac{1}{r^2/(t_{\alpha/2}^2 \cdot CV^2) + 1/N} \qquad (28)$$

where N is the total number of count intervals during the run, t is the t-distribution value for α under a two-tailed test, and CV is the coefficient of variation. The coefficient of variation is defined as

$$CV = \frac{\sqrt{s^2}}{\hat{\bar{y}}} \qquad (29)$$

The student t-distribution values are those for df = ∞ (Appendix A). If there are no estimates of mean and sample variance in the literature, or a preliminary study can not be conducted, then use a best guess for the CV in equation 28. Again, however, SRS is never used for herring run size estimation because the duration of the run is not known in advance.

For St1WRS, St2WRS, and StSYS, the issue is more complex because the ultimate goal is to produce an estimate of total run size given daily sample sizes and estimates of totals that are made every day of the run. Formulae are available (Cochran, 1963; Thompson, 2002) that can be used to allocate total n to each stratum assuming that total n is known. How is total n derived? The best strategy is to use daily data from the literature or derived in a pilot study to calculate an average of the CVs and assume that the total number of intervals per day (N_k) is the same throughout the run. Then to estimate total n, the following equation is used:

$$n = \frac{1}{r^2/(t_{\alpha/2}^2 \cdot CV^2) + 1/(N_k \cdot L)} \tag{30}$$

where L is the expected duration of the run and the remaining parameters are as previously defined. If N_k will change daily throughout the run, $N_k * L$ would be the sum of all N_ks. Because daily sample variances and totals will not be known in advance, n is allocated equally to each day such that daily sample size is calculated as

$$n_k = \frac{n}{L} \tag{31}$$

For St2STRS, determination of sample size is much, more complicated than the above. A pilot study must be conducted to estimate the within-hour as well as the between-hour sample variances. Derivation of sample size requires calculus to solve for the optimal number of time units and hours to sample. For those interested, see Thompson, 2002 (p.150-151) or Cochran (1963: p. 283-285).

Relationship between Precision, Coefficient of Variation, and Sample Size For Stratified Sampling

It may be of interest to determine what the resulting precision of the total run size will be for the average CV and the sample sizes estimated from equation 30. Although precision was stated in the equations (i.e., r), a measure of precision that is more easily interpreted is the proportional standard error (PSE) calculated as

$$PSE = \frac{\sqrt{\text{var}(\hat{Y})}}{\hat{Y}} \tag{32}$$

The PSE can be estimated given daily sample size, the average CV, and expected duration of the run.

Suppose for day k, the variance of the daily total under St1WRS is

$$\text{var}(\hat{Y}_k) = N_k \cdot (N_k - n_k) \cdot \frac{\hat{s}_k^2}{n_k}$$

Taking the square-root of both sides, rearranging the equation, and dividing by $\hat{Y}_k$ gives

$$\frac{\sqrt{\text{var}(\hat{Y}_k)}}{\hat{Y}_k} = \sqrt{\frac{N_k \cdot (N_k - n_k)}{n_k}} \cdot \frac{\hat{s}_k}{\hat{Y}_k}$$

Since $\hat{Y}_k = N_k \hat{\bar{y}}_k$, then

$$\frac{\sqrt{\text{var}(\hat{Y}_k)}}{\hat{Y}_k} = \frac{\sqrt{\frac{N_k \cdot (N_k - n_k)}{n_k}}}{N_k} \cdot \frac{\hat{s}_k}{\hat{\bar{y}}_k}$$

Substituting equations 29 and 32 gives

$$PSE_k = \frac{\sqrt{\frac{N_k \cdot (N_k - n_k)}{n_k}}}{N_k} \cdot CV_k \tag{33}$$

The true PSE for the total run size is not the sum PSE_k over all days, but

$$PSE_{St1WRS} = \frac{\sqrt{\sum_{k=1}^{L} \text{var}(\hat{Y}_k)}}{\sum_{k=1}^{L} \hat{Y}_k} = \frac{\sqrt{\sum_{k=1}^{L} \frac{N_k \cdot (N_k - n_k)}{n_k} \cdot s_k^2}}{\sum_{k=1}^{L} N_k \cdot \hat{\bar{y}}_k}$$

which shows that the PSE_{St1WRS} is a function of the duration of the run and daily variances and totals which are not known in advance. However, if it is assumed that N_k and n_k, are constant over the run, a PSE for the total can be approximated from the average CV of the total run by

$$PSE = \frac{\sqrt{L \frac{N \cdot (N - n)}{n}}}{L \cdot N} \cdot \overline{CV} \tag{34}$$

where $\overline{CV}$ is the average coefficient of variation over L days, N is the daily total number of time units, and n is the daily sample size.

The Power To Detect Trends in Herring Run Size

The ability to detect changes in run size over time is important for management purposes. If abundance is declining precipitously over a number of years, it may be a sign of trouble and therefore, managers may have to take regulatory action to avoid collapse. The ability to detect changes when they are occurring is known as *power* in the field of statistics. Power is related to the natural variability in fish passing, the sample size taken to estimate daily totals, the significance level (α) of inference testing, and the size of the change to be detected (also known as the "effect size"). Basically, as power increases, the more likely change will be detected. What this means is that if run size is imprecisely estimated each year, we may not be able to conclude statistically that an observed increasing or decreasing trend is occurring. Further discussion on statistical power can be found in Sokal and Rohlf (1981).

Although it may be difficult to allocate effort each year to produce annual run size estimates that are precise enough to detect year-to-year changes in abundance, trends over longer periods of time may be statistically detectable. The power analysis of Gerrodette (1987;1991) can be used to determine the minimum percent change in herring run size over time that can be detected at a given level of power based on the precision of one year's estimate. Gerrodette's procedure uses linear regression to test whether the slope of an assumed linear or exponential trend in run size differs from zero for a given rate of change over time, level of precision (PSE), significance level (α), and the null hypothesis of the test.

Gerrodette's formulation for power is complicated; therefore, I provide power curves for linear and exponential changes in run size over time for 3-6 years of surveys, for PSEs ranging from 0.1 to 0.3, for α =0.05 and α=0.10, and for one- and two-tailed tests (Figure 7). Power was calculated assuming that PSEs change inversely with the square-root of the run size (see Gerrodette, 1987), and using the non-centralized *t*-distribution because of the small number of years (3-6 years of surveys) explored (see Gerrodette, 1991). If the exact PSE or desired α level is not available in Figure 7, additional calculations can be made using the *R* code provided in Appendix B.

To determine the minimum percent change over time that can be detected, the following procedure should be used:

1. Obtain a PSE for the total run size or estimate using equation 34 for the expected number of days of the run.

2. Select the significance level (usually α =0.05)

3. Choose to examine either a linear or exponential changes in abundance

4. Determine what null hypothesis is to be tested. It may be more important to determine that run size is changing in one direction (decreasing or increasing) over time; therefore, a one-tailed test should be used because the null hypotheses tested are slope$\geq$0 (alternate hypothesis is slope<0) for a decreasing trend, or slope$\leq$0 (alternate hypothesis is slope>0) for an increasing trend. For a two-tailed test, the null hypothesis tested for is slope = 0 (alternate hypothesis is slope $\neq$ 0) meaning it can be concluded that change is occurring but not in a particular direction.

5. Select a desired power (most statistical textbooks recommend power should be $\geq$0.80).

6. Read the percent change that can be detected for the power probability and number of survey years from a plot in Figure 7 corresponding to the parameters selected above.

If the percent change is not at a desirable level for a given power and number of survey years, the only recourses are to increase daily sample sizes to increase precision of the estimate, or increase the counting interval which will reduce sample variation.

As an example, suppose that a watershed group is interested in determining the percent decline in run size that can be detected over time. The group's sampling efforts generate an estimate of

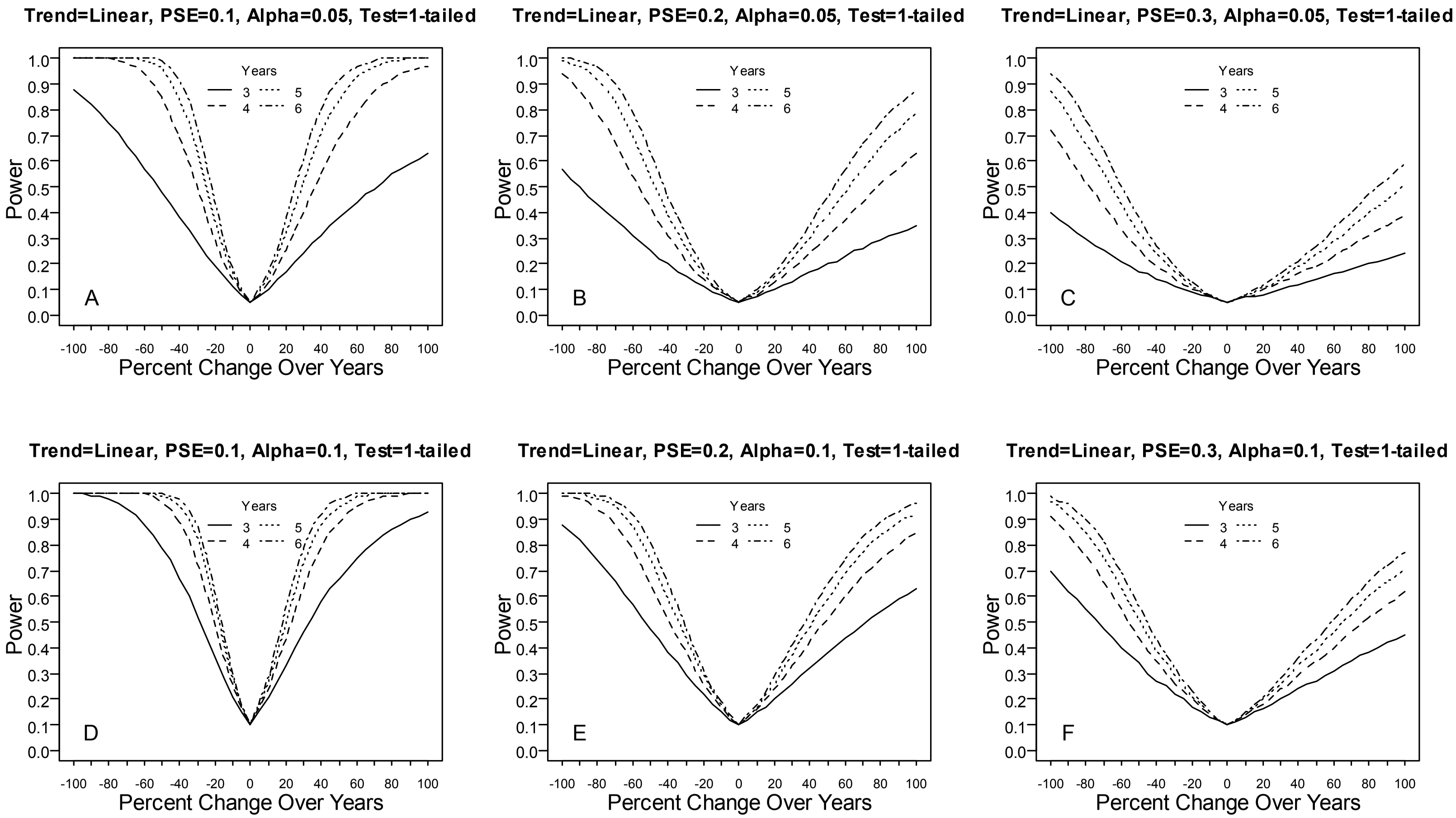

Figure 7. Power versus percent change derived via the method of Gerrodette (1987; 1991) for linear and exponential (Exp) trends in abundance, proportional standard errors from 0.1 to 0.3, α of 0.05 and 0.10, one– and two-tailed hypothesis tests, and 3-6 survey years.

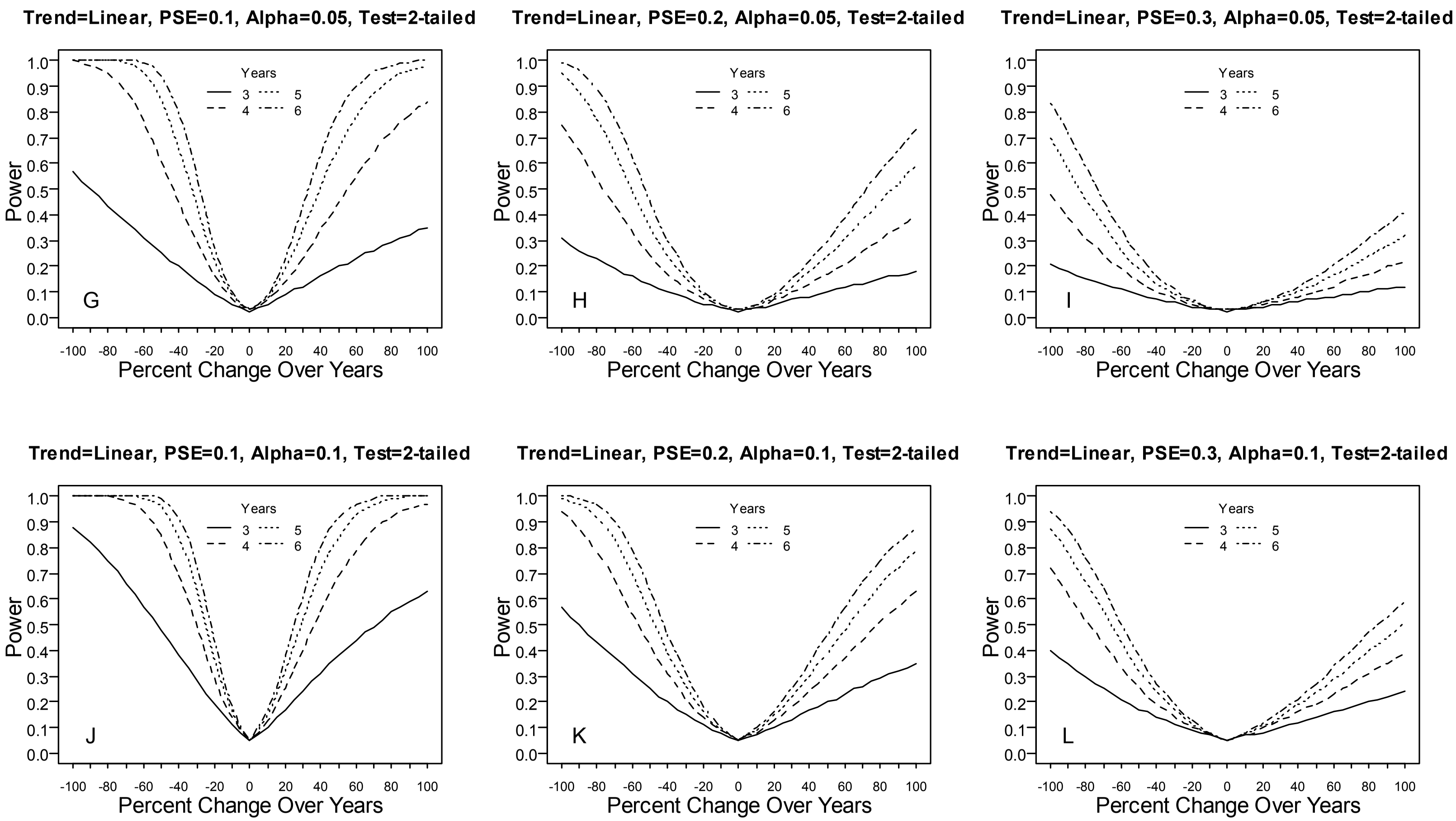

Figure 7 cont'd.

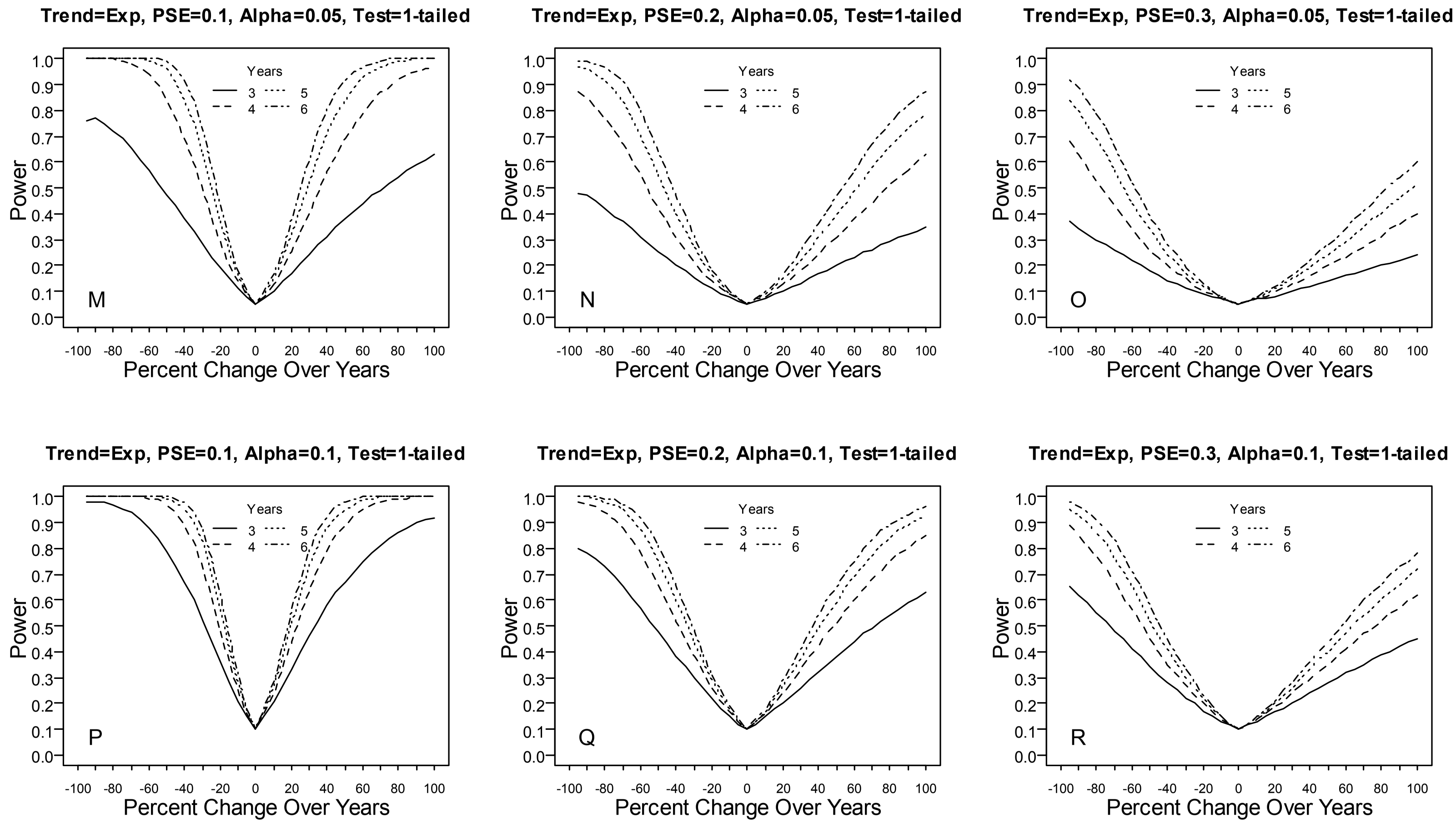

Figure 7 cont'd.

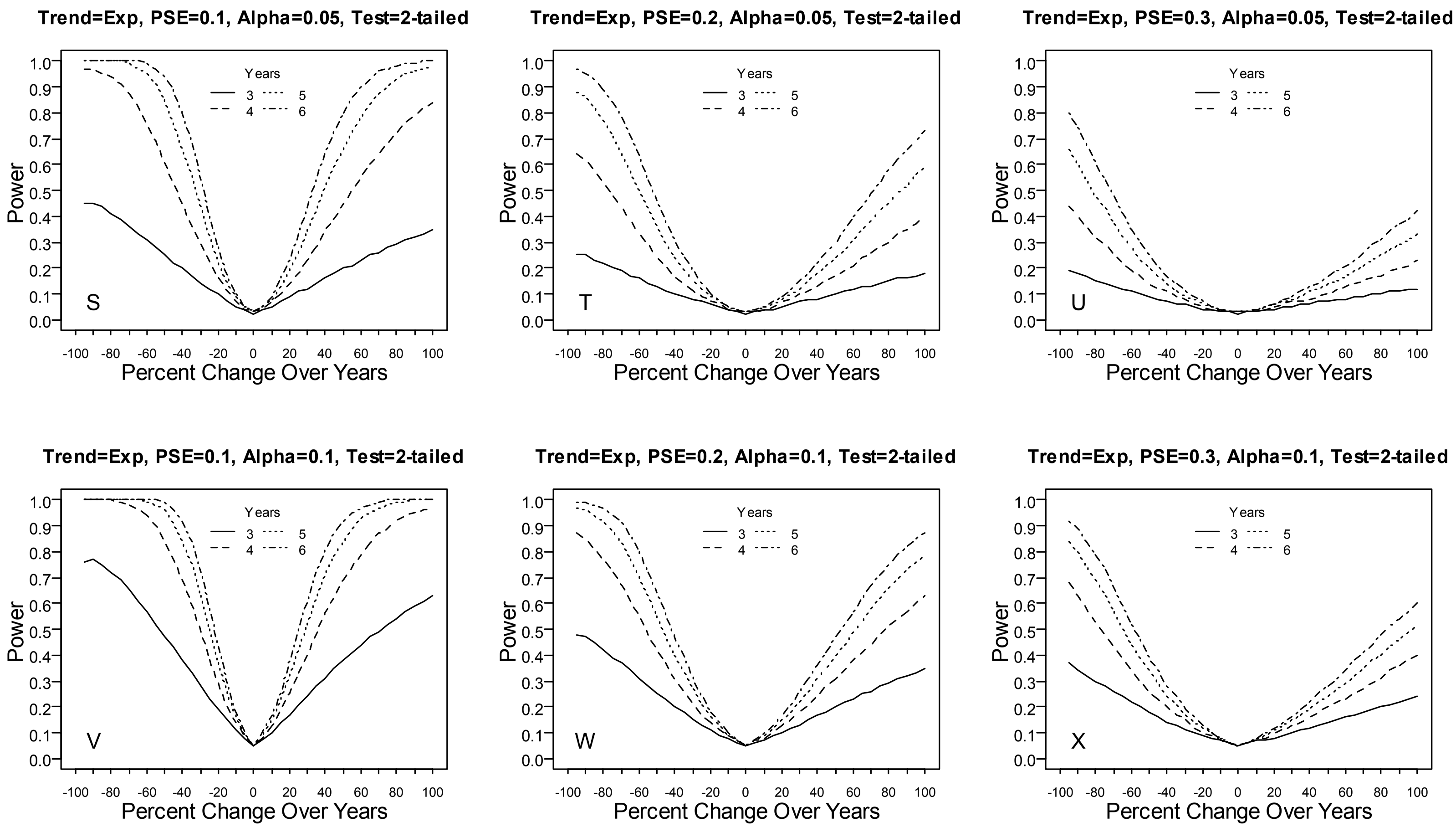

Figure 7 cont'd.

total run size with a PSE=0.10. They select a one-tailed test at α =0.05, want power to be 0.80, believe an exponential decline would be the underlying trend form, and want to detect changes over 4 years. Referring to Figure 7M, the minimum percent decline that can be detected is 42% over 4 years.

What If The Sampling Procedures Aren't Followed?

Three deficiencies that seriously affect the accuracy and precision of run size estimates were identified in a review of efforts made by several watershed groups; these were 1) low sample sizes, 2) patterned sampling, and 3) interpolation of daily counts for missing days. Often, volunteers were unavailable to make counts every hour as required under the sampling scheme of Rideout et al. (1979) thus lower expected sample sizes were achieved. As discussed previously, low sample sizes will reduce the precision of the daily estimate and affect the overall precision of the run size estimate and the power to detect a change. Patterned sampling occurred because volunteers were often available only after work hours, and counting became aggregated during those times (Figure 6). Such patterned sampling can cause biased estimates of run size, particularly if herring migrate differently throughout the day, because not all time intervals are available for selection. Daily counts for missing days estimated by using linear interpolation of counts from adjacent non-missing days can introduce bias and error if dramatic changes in the numbers passing occur between days, and if the number of missing days is large. Unless these deficiencies are minimized by each watershed group, the usefulness of their run size estimates in the management of herring will be questionable.

To demonstrate the effects of the three deficiencies on the estimation of run size, sampling simulations were conducted on a computer-generated herring run. The benefit of this approach is that, by using the computer to sample, bias and error can be exposed by comparing the estimates to known daily and total run sizes. The hypothetical herring run was created from actual count data collected at the Central Street ladder on the Parker River during 1999 and 2000. Assuming a 13 hour day and a run duration of 33 days, the count of herring in every five minute interval of each day was generated from a negative binomial distribution using the RAND function in SAS (SAS, 2002). The distribution of daily counts was parameterized with the daily mean count and *k* parameter approximated by

$$k = \frac{\hat{\bar{y}}^2}{s^2 - \hat{\bar{y}}}$$

where *y* is the daily mean count per 5-minute interval and s^2 is the sample variance of the mean count (Krebs, 1989). The negative binomial distribution was used because fish count data are often characterized as having many zero counts and highly skewed non-zero counts (Figure 8). The daily mean count of the first 5-minute intervals collected in 1999 was assumed to represent each day's average. Because of sampling deficiencies, s^2 for each daily mean was estimated from a linear equation that related s^2 to mean counts derived using daily data with complete counts from 1999 and 2000 (Figure 9). Depending on the value of the mean (*m*), s^2 was estimated as

$$s^2(m) = \begin{cases} m = 0, & 0 \\ m > 0, & \exp^{1.103+1.556 * \ln(m)} \end{cases}$$

The daily trends in numbers passing followed the actual estimates made for the Parker River in 1999 (Figure 10). The daily passage of herring was simulated as a bimodal event with a minor peak occurring in the mid-morning hours and a major peak occurring in the afternoon hours (Jessop and Harvie, 1990; Figure 6C this document). The actual total run size was 12,442 herring. In the simulations, a St1WRS design was used to select randomly 10-minute intervals each day of the 33-day run from which the daily total was calculated. The sum of all daily totals was the estimate of total run size.

To demonstrate the effect of sample size on precision, 1 to 13 ten-minute intervals were randomly selected each day. The simulation was repeated 500 times for each sample size to generate a distribution of total run size values that could be obtained if sampling was repeated. The results of the simulations are shown in Figure 11. The mean of the 500 simulations for each sample size is indicated by a circle and the 95% percentiles are represented by the lower and upper whiskers. Notice that the mean for a given sample size equals the known run size of 12,442 fish, indicating that the sampling design and random selection process produces unbiased estimates of the total run size. Also, notice that the range of the whiskers decreases rapidly as sample size increases, indicating precision increases rapidly with sample size.

To demonstrate the effect of patterned sampling, similar simulations and analyses were conducted except that only afternoon intervals were used in the random selection process to simulate the volunteers' availability for only afternoon counting. The results of the simulations are shown in Figure 11B. Notice that the mean of all simulations for a given sample size is well above the known run size of 12,442 fish. This indicates that patterned sampling produces biased estimates of the total run size, and even though precision of the estimates increases with increasing sample size, the accuracy does not improve (Figure 11B).

To demonstrate the effect of using linear interpolation to estimate missing daily totals, simulations were made in which actual daily totals for the 33-day run were randomly coded as missing, and then linear interpolation was used to estimate the total for the missing day. Because the accuracy of linear interpolation is likely to diminish if days with missing data occur sequentially in time, the simulations examined the effect of increasing the sequential number of days with missing data, and also the effect of increasing the number of missing day sequences. That is, the simulations examined the impact of not making counts for several days in a row and making those sequential non-counts many times throughout the duration of the run. The results are presented in Figure 12. Notice that when the number of sequential missing days was ≤ 4 and the number of sequences increased, the means were close to the actual run size, indicating that linear interpolation produces fairly unbiased estimates of the total run size.

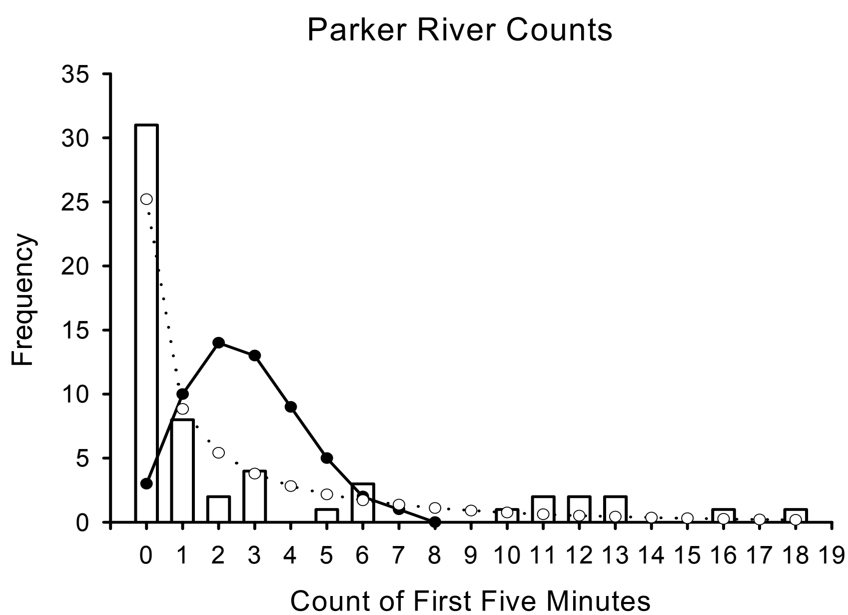

Figure 8. Comparison of observed herring counts from the Parker River (bars) to predicted values generated from the negative binomial (hollow circles) and Poisson (solid circles) distributions.

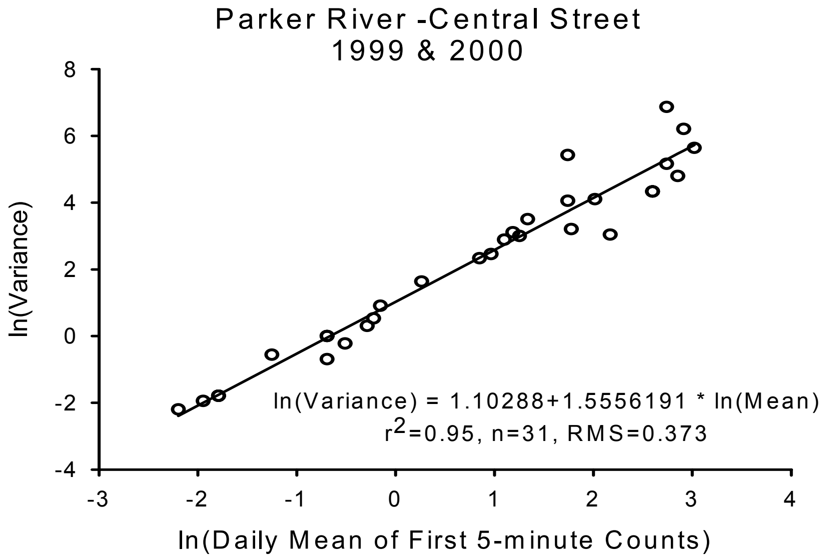

Figure 9. Relationship between ln(sample variance) and ln(mean count per 5-minutes) from the Parker River—Central Street ladder, 1999-2000.

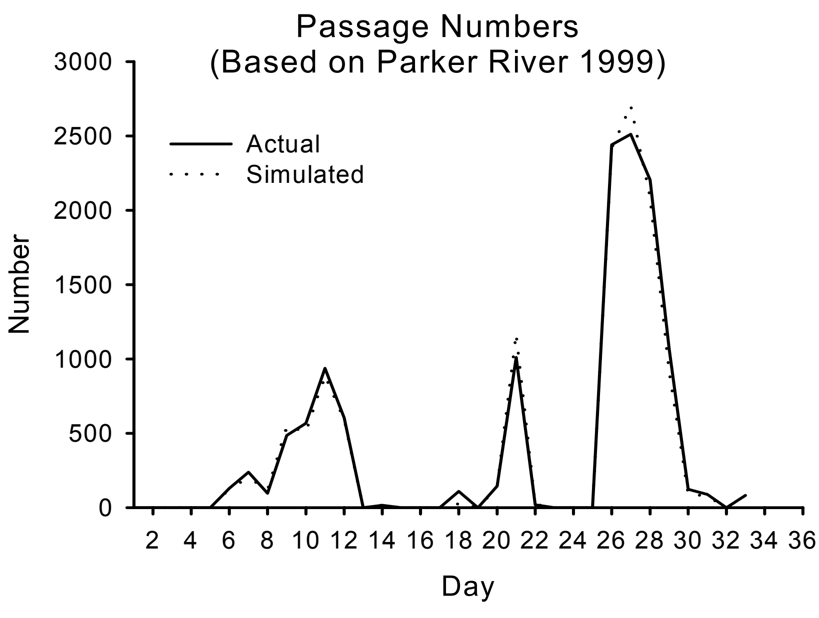

Figure 10. Comparison of estimated and simulated daily run sizes using data from the Parker River-Central Street ladder, 1999.

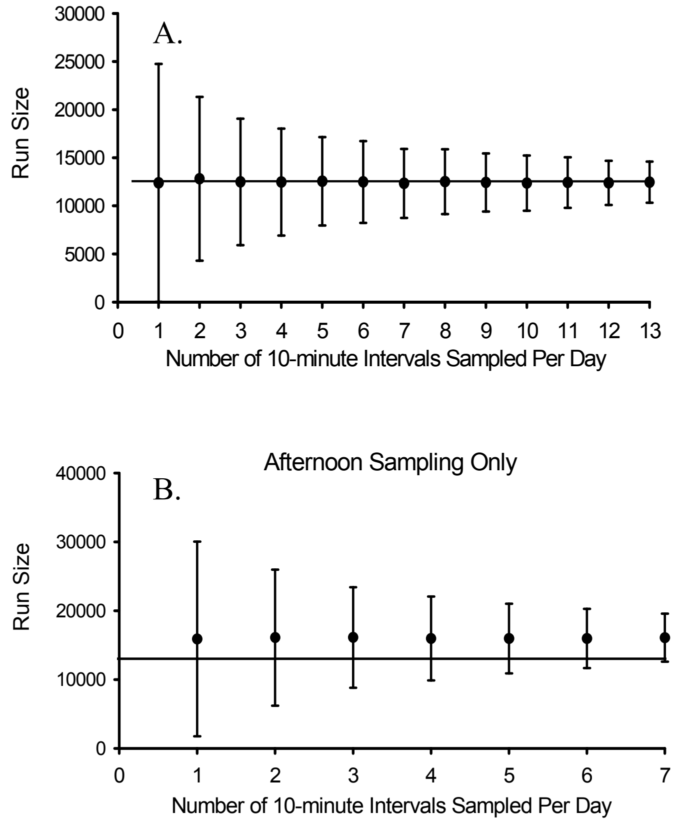

Figure 11. Mean and 95% percentiles of the run size distribution generated from the simulations, examining A) effects of differing sample size, and B) effects of selection of only afternoon time intervals.

However, precision decreases as the number of missing days and the number of sequences of missing days increases (Figure 12). Bias in the estimates is produced when the number of sequential missing days exceeds 5, and the number of sequences increases (Figure 12). From these results, as long as the number of days in a row with missing data does not exceed 5, then linear interpolation should only affect the precision of the estimate of total run, not the accuracy.

How to Proceed With Estimation of Herring Run Size

The steps to a statistically-sound survey design used to estimate herring run sizes are as follows:

1. Determine the effort (number of volunteers and availability) that can be realistically sustained for the expected duration of the run.

2. Select the most appropriate stratified random design for the effort available.

3. Determine the size of the counting interval (can use results from other studies).

4. Obtain estimates of mean and sample variance over several days from a pilot study, from neighboring watershed groups, or from scientific literature.

5. For the chosen survey design, estimate the appropriate daily sample size given the desired precision and error rate.

6. Calculate the proportional standard error for the daily sample size, average coefficient of variation, and estimated length of the run.

7. Determine the percent change in run size that can be detected over time given the propor-

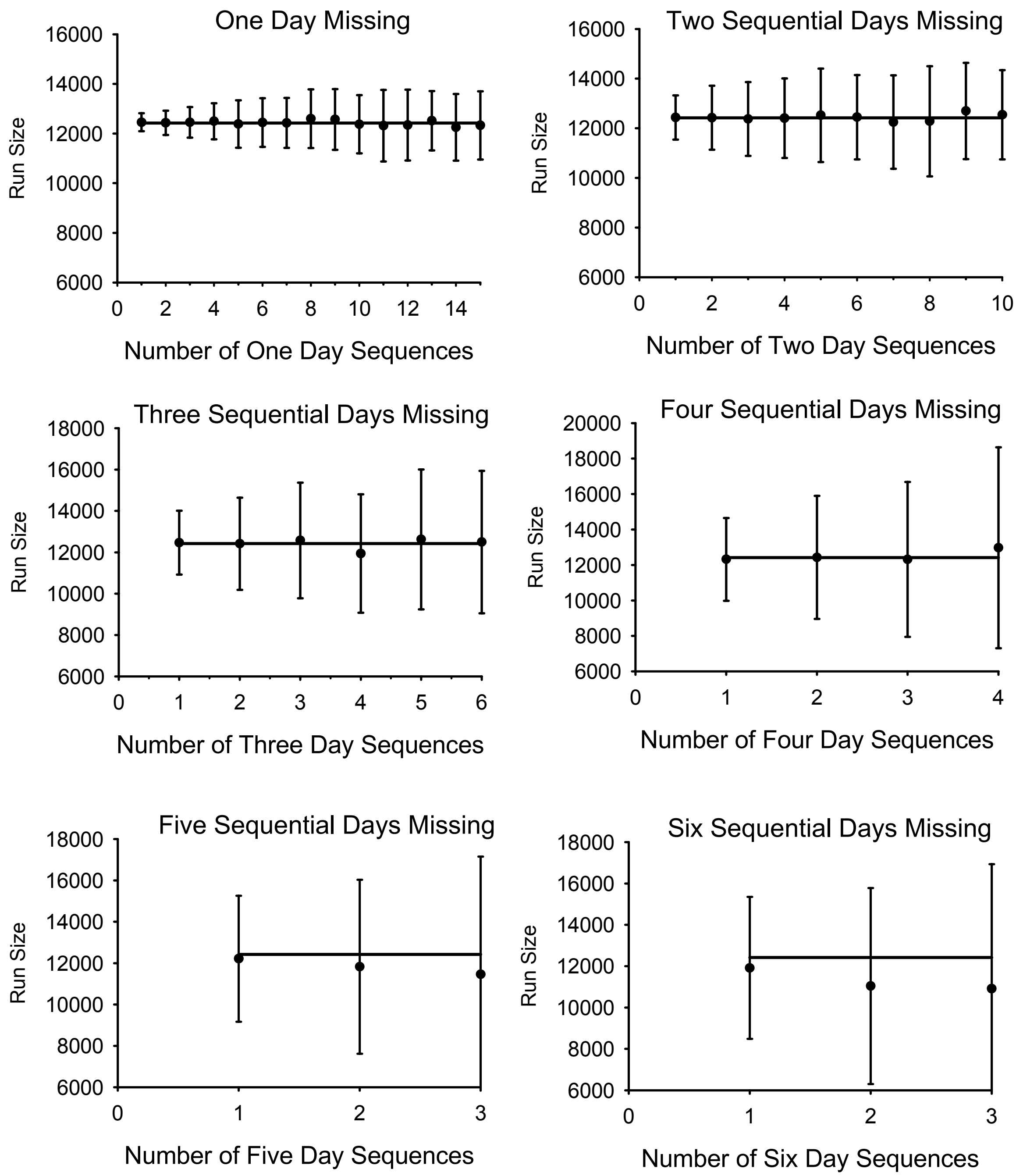

Figure 12. Mean and 95% percentiles of the run size distribution generated from the simulations examining the effects of linear interpolation on accuracy and precision. The horizontal line is the true run size.

tional standard error, desired power and error rate, and number of years over which a decline or increase would be important.

8. If detecting smaller changes in run size are required, then repeat steps 5-7, increasing precision (r) until the desired detectability is reached.

9. Conduct the survey and make adjustments to design for the following year if required.

The following hypothetical example will show the steps that are needed to conduct a statistically-sound and useful survey design for the estimation of herring run size.

Suppose the Ipswich Watershed Association wants to conduct a survey to estimate the size of the Ipswich River herring run passing over the Sylvania Dam in 2006. The group believes that they will have no trouble soliciting the help of at least 5-10 volunteers who will make one 5 minute count per day for the entire 33 day duration of the run. Based on this information, it is decided that a St1WRS design is most appropriate for their needs and a 13 hour day will be assumed. Because of lack of data for the Ipswich River, estimates of mean counts and sample variances from Rideout et al. (1979) were used by the group to estimate total and daily sample size. The group is interested in obtaining an estimate within 0.2 (r) of the true total with 95% confidence. These values are entered in Table 1, along with the total number (N) of 5-minute intervals per 13 hour day (156=13 hours*60 minutes /5 minutes) and the t-value for the associated confidence level (=1.960). After calculating the daily coefficients of variation, the average CV and selected parameters are inserted into equation 30 and 31 to derive daily sample sizes. In this example, daily sample size is 12 (Table 1). If 12 counts per day are too high for the organization to support, the only options are to reduce the precision by reducing *r* in equation 30, or increase the duration of the interval count which will lower the sample variance.

The next step in the process is to determine the change in percent run size that can be detected over time. The PSE derived using equation 38 is 0.10. Since the group believes it is more important to know that the herring abundance is declining exponentially, a one-tailed test is assumed. They want the power to be fairly high, so a value of 0.80 is selected. Using Figure 7M, the minimum percent decline in the abundance that can be detected with a power of 0.80 is 45% over 4 years, 32% over 5 years, and 28% over 6 years. Notice

Table 1. Example calculations for determining sample size and power of the herring run size estimates.

Fishway	Mean number of herring per 5-minutes	s^2	CV
1	4.377	111.59	2.413
2	3.773	46.43	1.806
4	2.571	30.11	2.134
		Average CV	2.118

CV	N	L	r	t	Total n	Daily n	PSE of Total
2.118	156	33	0.2	1.96	397	12.04	0.102

that there is not enough power to detect a 100% decline in abundance over three survey years.

Recommendations to Community Groups

There are many choices of statistical designs offered in this document to suit each program and volunteer behavior. However, it may be difficult deciding which design and level of counting are right for your group. At the herring counting workshop held at the Massachusetts Division of Marine Fisheries' Annisquam River Field Station on February 2, 2005, it was recommended that counting groups use the two-way stratified random sampling design to help overcome problems discussed earlier and to produce reasonably accurate and precise estimates of run size. The detailed recommendations are below:

<u>Counting Season</u>:	April 1st to mid-June.
<u>Counting Day</u>:	7 am to 7 pm (12 hrs.)
<u>Counting Periods</u>:	7-11 am, 11-3 pm, and 3-7 pm.
<u>Counting Interval</u>:	10 minutes.
<u>Counting Coverage</u>:	3 counts per period

There is no need to strictly coordinate the arrival times of volunteers by picking random intervals each day and then assigning each to a volunteer. As long as they arrive in somewhat of a random fashion and effort occurs in each period, the estimate should be reasonably accurate. What volunteers should avoid is setting arrivals at the same time each day. This would produce a systematic-like sampling design that has not been formulated correctly and it will ultimately produce biased estimates.

Literature Cited

Cochran, W. G. 1963. Sampling Techniques. John Wiley & Sons, Inc., New York, NY. 413 p.

Gerrodette, T. 1987. A power analysis for detecting trends. Ecology 68(5): 1364-1372.

Gerrodette, T. 1991. Models for power of detecting trends- a reply to Link and Hatfield. Ecology 72(5): 1889-1892.

Jessop, B. M. and C. J. Harvie. 1990. Evaluation of designs of periodic count surveys for the estimation of escapement at a fishway. N. Am. J. Fish. Manage., 10:39-45.

Krebs, C. J. 1989. Ecological Methodology. Harper & Row, Publishers. New York, NY. 654 pp.

Loesch. 1987. Overview of life history aspects of anadromous alewife and blueback herring in freshwater habitats. *In* M. J. Dadswell, R. J. Klauda, C. M. Moffitt, R. L. Saunders, R. A. Rulifson, and J. E. Copper (eds.), Common strategies of anadromous and catadromous fishes. American Fisheries Society Symposium 1. pp 89-103.

Levy, P. and S. Lemeshow. 1999. Sampling of Populations: Methods and Applications. John Wiley and Sons, Inc. New York, NY. 528 p.

Mullen, D. M., C. W. Fay, and J. R. Moring. 1986. Species profile:life histories and environmental requirements of coastal fishes and invertebrates (North Atlantic). U. S. Fish Wildl. Serv. Biol. Rep. 82(11.56). U. S. Army Corps of Engineers, TR EL-8204. 21 pp.

Rideout, S. G., J. E. Johnson, and C. F. Cole. 1979. Periodic counts for estimating the size of spawning population of alewives, *Alosa pseudoharengus* (Wilson). Estuaries 2(2): 119-123.

SAS Institute. 2004. SAS OnlineDoc 9.1.2, SAS Language Reference Dictionary.

Sokal, R. R. and F. J. Rohlf, 1981. Biometry. W. H. Freeman and Company, New York, NY. 859 p.

Thompson, S. K. 2002. Sampling. John Wiley & Sons, Inc., New York, NY. 367 p.

Appendix A. Critical values of Student's t-distribution (two-tails).

df	α / Confidence	0.2 / 80%	0.15 / 85%	0.1 / 90%	0.05 / 95%	0.01 / 99%	0.001 / 99.9%
1		3.078	4.165	6.314	12.706	63.657	636.619
2		1.886	2.282	2.920	4.303	9.925	31.599
3		1.638	1.924	2.353	3.182	5.841	12.924
4		1.533	1.778	2.132	2.776	4.604	8.610
5		1.476	1.699	2.015	2.571	4.032	6.869
6		1.440	1.650	1.943	2.447	3.707	5.959
7		1.415	1.617	1.895	2.365	3.499	5.408
8		1.397	1.592	1.860	2.306	3.355	5.041
9		1.383	1.574	1.833	2.262	3.250	4.781
10		1.372	1.559	1.812	2.228	3.169	4.587
11		1.363	1.548	1.796	2.201	3.106	4.437
12		1.356	1.538	1.782	2.179	3.055	4.318
13		1.350	1.530	1.771	2.160	3.012	4.221
14		1.345	1.523	1.761	2.145	2.977	4.140
15		1.341	1.517	1.753	2.131	2.947	4.073
16		1.337	1.512	1.746	2.120	2.921	4.015
17		1.333	1.508	1.740	2.110	2.898	3.965
18		1.330	1.504	1.734	2.101	2.878	3.922
19		1.328	1.500	1.729	2.093	2.861	3.883
20		1.325	1.497	1.725	2.086	2.845	3.850
21		1.323	1.494	1.721	2.080	2.831	3.819
22		1.321	1.492	1.717	2.074	2.819	3.792
23		1.319	1.489	1.714	2.069	2.807	3.768
24		1.318	1.487	1.711	2.064	2.797	3.745
25		1.316	1.485	1.708	2.060	2.787	3.725
26		1.315	1.483	1.706	2.056	2.779	3.707
27		1.314	1.482	1.703	2.052	2.771	3.690
28		1.313	1.480	1.701	2.048	2.763	3.674
29		1.311	1.479	1.699	2.045	2.756	3.659
30		1.310	1.477	1.697	2.042	2.750	3.646
31		1.309	1.476	1.696	2.040	2.744	3.633
32		1.309	1.475	1.694	2.037	2.738	3.622
33		1.308	1.474	1.692	2.035	2.733	3.611
34		1.307	1.473	1.691	2.032	2.728	3.601
35		1.306	1.472	1.690	2.030	2.724	3.591
36		1.306	1.471	1.688	2.028	2.719	3.582
37		1.305	1.470	1.687	2.026	2.715	3.574
38		1.304	1.469	1.686	2.024	2.712	3.566
39		1.304	1.468	1.685	2.023	2.708	3.558
40		1.303	1.468	1.684	2.021	2.704	3.551
50		1.299	1.462	1.676	2.009	2.678	3.496
60		1.296	1.458	1.671	2.000	2.660	3.460
70		1.294	1.456	1.667	1.994	2.648	3.435
80		1.292	1.453	1.664	1.990	2.639	3.416
90		1.291	1.452	1.662	1.987	2.632	3.402
100		1.290	1.451	1.660	1.984	2.626	3.390
110		1.289	1.450	1.659	1.982	2.621	3.381
120		1.289	1.449	1.658	1.980	2.617	3.373
130		1.288	1.448	1.657	1.978	2.614	3.367
140		1.288	1.447	1.656	1.977	2.611	3.361
200		1.286	1.445	1.653	1.972	2.601	3.340
500		1.283	1.442	1.648	1.965	2.586	3.310
∞		1.282	1.440	1.645	1.960	2.576	3.291

Appendix B. R code for power analysis. R is free and can be downloaded at *www.r-project.org*.

```
################################################################
#### Power Analysis for Linear and Exponential Trends  ########
#### Based on Gerrodette 1987, 1991                    ########
#### Tested Against Gerrodette's TREND program         ########
################################################################

 Trend<-2    # 1= Linear, 2=Exponential
A1<-100     # Start Year Abundance
CV<-0.30    # Proportional Standard Error (CV in Gerrodette)
cvrel<-1    # PSE of A: 1=1/sqrt(A),2=constant, 3=sqrt(A)
maxyrs<-6   # Maximum number of Years
pR<-100 # Highest Positive R in percent
step<-5 # Increment of R
alpha<-0.05 # Alpha Level (Type I error)
tail<-2    # 1=one-tail, 2=two-tail test


 ######################  Program   ####################
tlen<-length(seq(-100,pR,by=step))
results<-as.data.frame(array(rep(NA,tlen*(maxyrs-3+1)*8),dim=c(tlen*(maxyrs-3+1),8)))
names(results)<-c("Years","Trend","CV","alpha","side","R","r","Power")
place<-0

 for (i in 3:maxyrs){ #Loop number of years
        nyr<-i
      R<--100
   for (j in 0:(tlen-1)){
    s2x<-((nyr+1)*(nyr-1))/12 ###Variance of X

  ####### Linear Model #########
   if(Trend==1){
             r<-(R/100/(nyr-1))
                b<-A1*r
            s2res<-ifelse(cvrel==1,((CV*A1)^2)*(1+(r/2)*(nyr-1)),
              ifelse(cvrel==2,((CV*A1)^2)*(1+r*(nyr-1)*(1+(r/6)*(2*nyr-1))),
              ((CV*A1)^2)*(1+(3*r/2)*(nyr-1)*(1+(r/3)*(2*nyr-1)+(r^2/6)*nyr*(nyr-1)))))
   }
 ###### Exponential Model ##############
   if (Trend==2){
        r<-((R/100+1)^(1/(nyr-1)))-1
        b<-log(1+r)
         if(cvrel==1){
                sum<-0
             for(k in 1:i){
                 sum<-sum+log((CV^2/((1+r)^(k-1)))+1)
             }
             s2res<-sum/i
         }
         if(cvrel==2){
             s2res<-log(1+CV^2)
         }
         if(cvrel==3){
                sum<-0
             for(k in 1:i){
               sum<-sum+log((CV^2*(1+r)^(k-1))+1)
             }
             s2res<-sum/i
         }
   }#If  close

       sb<-sqrt(s2res/(nyr*s2x))
      delta<-abs(b/sb)
      v<-nyr-2
```

Appendix B cont'd.

```
      tdist<-abs(qt(alpha/tail,v))
      powert<-round(1-pt(tdist,df=nyr-2,ncp=delta),2)
      place<-place+1
       results[place,1]<-i;results[place,2]<-Trend;results[place,3]<-CV;results[place,4]<-
alpha
      results[place,5]<-tail;results[place,6]<-R;results[place,7]<-r
      results[place,8]<-powert
        R<-R+step
    }
}

    t1<-ifelse (Trend==1,"Linear","Exponential")
   #t2<-ifelse (cvrel==1,"1/SQRT(A)",ifelse(cvrel==2,"Constant","SQRT(A)"))
   tit<-c(paste("Trend=",t1,sep=""),paste("PSE=",CV,sep="")
      ,paste("Alpha=",alpha,sep=""),paste("Test=",tail,"-tailed",sep=""))
   tit<-c(paste(tit,sep="",collapse=", "))

 #########Make Legend Labels##########
par(mfrow=c(1,1),adj=0.5, mgp=c(1.5,0.6,0))
leg.txt<-rep(NA,maxyrs-2)
for (i in 1:maxyrs-2){
leg.txt[i]<-c(i+2)
}
legtype<-seq(1:(maxyrs-2))
for (i in 3:maxyrs){

    if (i==3){
       data<-subset(results,results$Years==i)
       plot(data$R,data$Power,type="l",col="black",lty=c(1),ylim=c(0,1),
        ylab="",xaxt="n",yaxt="n",
        xlab="",main=tit,cex.main=1.2)
        axis(1,seq(-100,pR,10),tcl=-0.3,cex.axis=0.9,mgp=c(1.5,0.6,0))
         axis(2,seq(0,1,0.1),las=1)
         mtext("Power",at=-135,cex=1.0,line=-12,las=2)
           mtext("Percent Change Over Years",at=0.5,line=-25,cex=1.0)
    }
    if (i>3){
       par(new=T)
       data<-subset(results,results$Years==i)
       lines(data$R,data$Power,type="l",lty=c(i-2),col="black")
    }
}
 legend(x=-20,1,legend=leg.txt,ncol=maxyrs/2,lty=legtype,col="black",
   bty="n",title="Years         ",cex=0.8)
```