LRH: R. M. Dorazio

RRH: Inference in surveys of mollusks

Design-based and model-based inference in surveys of freshwater mollusks

Robert M. Dorazio

United States Geological Survey, Biological Resources Division, 7920 NW 71 Street,

Gainesville, Florida 32653 USA

Abstract. Well-known concepts in statistical inference and sampling theory are used to develop recommendations for planning and analyzing the results of quantitative surveys of freshwater mollusks. Two methods of inference commonly used in survey sampling (design-based and model-based) are described and illustrated using examples relevant in surveys of freshwater mollusks. The particular objectives of a survey and the type of information observed in each unit of sampling can be used to help select the sampling design and the method of inference. For example, the mean density of a sparsely distributed population of mollusks can be estimated with higher precision by using model-based inference or by using design-based inference with adaptive cluster sampling than by using design-based inference with conventional sampling. More experience with quantitative surveys of natural freshwater mollusk assemblages is needed to determine the actual benefits of different sampling designs and inferential procedures.

Key words: survey, sampling design, statistics, inference, model, mollusk, benthic.

Surveys of assemblages of freshwater mollusks are undertaken for many reasons. Estimation of the composition and number of species is often the primary objective of faunal surveys (Watters 1992). In ecological surveys, mollusks and their environment are typically sampled to examine associations between the 2. For example, ecological surveys are used to identify mussel habitat and to detect temporal or spatial trends in mussel density with respect to environmental gradients (Strayer and Ralley 1993, Layzer and Madison 1995). The detection of temporal trends is particularly important in longitudinal studies of the sensitivity of mollusks to various forms of environmental degradation, such as loss of habitat or reduction in water quality.

Collection methods and sampling designs used in surveys of freshwater mollusks vary; however, malacologists distinguish their approaches to sampling as either **qualitative** or **quantitative**. In qualitative surveys, visual or tactile searches for mollusks at or near the sediment surface are completed while wading or diving (using snorkel or SCUBA). The searches may be timed (all mollusks encountered within a fixed period of time are included in a sample) or based on a fixed count of animals (search continues for an unspecified duration until a fixed number of mollusks is encountered). Timed searches produce observations of catch per unit of effort (CPUE), the number of mollusks encountered in the search divided by the search time. Fixed-count searches, which are used primarily in faunal surveys (Barbour and Gerritsen 1996, Courtemanch 1996, Vinson and Hawkins 1996), identify the cumulative number of species included among all mollusks encountered in the search. Quantitative sampling differs from qualitative sampling primarily in the unit of sampling. Units of quantitative samples are obtained by dividing the entire area to be sampled into non-overlapping subareas. This procedure is often assisted by physical survey instruments, such as rectangular quadrats or transect lines. Mollusks within sample units are detected by searching (visual or tactile) the sediment surface or, more commonly, by excavating the sediment to a prescribed depth and passing mollusks and sediment through sieves to help separate animals from particles of sediment. Therefore, each unit in a quantitative sample contains an observed density of mollusks, which equals the number of mollusks divided by the surface area of the unit.

Hornbach and Deneka (1996), Strayer et al. (1997), and Vaughn et al. (1997) recently identified various strengths and weaknesses of qualitative and quantitative sampling by comparing mussels collected with both timed searches and quadrats at identical locations. Estimates of species richness and diversity did not vary with method of collection when sufficient effort was used to collect similar numbers of mussels; however, the composition and relative abundances of species differed significantly with collection method. Mussels that were small, smooth-shelled, or deeply buried were difficult to locate or to distinguish from sediments in the visual or tactile searches. Consequently these mussels either were missed completely or were undersampled relative to the mussels collected in excavated quadrats. Large mussels with highly sculptured shells, on the other hand, tended to be oversampled relative to the mussels collected in quadrats. Another shortcoming of timed searches is that search efficiency may vary with observer, because of differences in training or fatigue, or with physical conditions at the study site, such as fallen logs, dense vegetation, or poor water clarity, which impede the search effort (Strayer et al. 1997). Miller and Payne (1993) compared mussels located by another form of qualitative sampling, fixed-count searches, with mussels found in quadrats at the same study sites. As with timed searches, estimates of species richness and diversity were similar in samples collected by either method, and several species of mussels were either undersampled or oversampled in the fixed-count searches relative to the quadrats. The differences were attributed to the same kinds of species-specific differences in detection that occurred with timed searches.

Quantitative sampling that includes excavation and sieving of sediments appears to be an effective but costly (in terms of time and effort) method of collecting freshwater mollusks. The primary advantages are significant improvements in detection of individuals and elimination of errors caused by variations in search efficiency. Quantitative sampling also exploits the relatively sedentary nature of freshwater mollusks. Although some mollusks are capable of limited horizontal movements, animals are unlikely to move beyond the region of sampling during the time required to complete a typical survey. Therefore, the statistical population may be defined in terms of a region of interest that is subdivided into non-overlapping, spatial units of sampling. Spatial definitions are useful in faunal surveys designed to estimate species richness because the estimated number of species is referenced to a finite region of measurable area. The importance of relating species richness to area is well known (Connor and McCoy 1979).

Despite the advantages of quantitative sampling methods, their relatively high cost has discouraged their use in surveys of mollusks. Qualitative searches generally reveal a much higher number of individuals than quadrat samples taken in the same amount of time and are therefore more likely to encounter rare species when sample size is limited by search time. Consequently, Miller and Payne (1988, 1993) recommended qualitative sampling for estimation of community-level features, such as richness or diversity of species, and quantitative sampling for estimation of mussel density, size composition, and recruitment.

A conceptual framework is clearly needed to help develop recommendations for planning and implementing surveys of natural mollusk assemblages. In this paper I use well-known concepts in statistical inference and sampling theory to develop recommendations for designing quantitative surveys of freshwater mollusks and their environment. I begin with a general introduction of the 2 philosophically distinct forms of statistical inference commonly used in survey sampling (design-based and model-based). More detailed descriptions of these 2 approaches follow, drawing on examples that are relevant in surveys of freshwater mollusks. I conclude with a list of recommendations for surveys of different objectives.

Statistical inference in survey sampling

Design-based and model-based approaches to inference are generally described in textbooks of survey sampling (Hedayat and Sinha 1991, Thompson 1992), but 1 is often emphasized in favor of the other because of differences in authors' perceptions of the relative strengths and weaknesses of the 2 approaches (Cassel et al. 1977, Särndal 1978, Hansen et al. 1983). Here, I discuss the 2 approaches in the context of their applicability to surveys of freshwater mollusks. My notation follows that of Thompson and Seber (1996).

Suppose the statistical population is composed of *N* individual units. In surveys of mollusks *N* might be the total number of quadrats or strip transects needed to define the region being surveyed. Each unit is distinct and can be identified by a label, say *i*, where $i \in \{1,2,...,N\}$. Associated with the *i*th unit is an observable characteristic or measurement y_i, which, for example, might be the number of mollusks per quadrat. In survey sampling we are interested in making an inference about the population of *y*-values, summarized by the vector $\mathbf{y} = (y_1,...,y_N)^T$. Inferences are made from information contained in a sample of *n* (< *N*) units of the population. The sample is represented by an ordered sequence of unit labels $s_O = (i_1,...,i_n)$. The sample data d_O comprise both the *y*-values and the labels of the sample units; therefore, d_O is represented by the following sequence of ordered pairs: $d_O = ((i_1, y_{i_1}),(i_2, y_{i_2}),...,(i_n, y_{i_n}))$. To avoid double subscripts, a useful shorthand notation for this sequence is $d_O = (s_O, \mathbf{y}_O)$, where $\mathbf{y}_O$ represents the sample *y*-values associated with the ordered sequence of unit labels. In descriptions of model-based inference, it is also useful to consider the sample *y*-values associated with a reduced set of unit labels that are distinct and uniquely ordered from smallest to largest label. Denote this set and its corresponding *y*-values as *s* and $\mathbf{y}_s$, respectively. All the information in d_O is contained in the pair $(s_O, \mathbf{y}_s)$ (Thompson and Seber 1996); therefore, an equivalent definition of the sample data is $d_O = (s_O, \mathbf{y}_s)$.

Given this notation, we now introduce the differences between design-based inference and model-based inference. In the design-based view the population *y*-values are regarded as a fixed set of unknown constants, say $\theta = \mathbf{y}$. Consequently, any parameter defined as a function of **y** is also regarded as a fixed constant. The population mean, defined as $\mu = \frac{1}{N}\sum_{i=1}^{N} y_i$, is a common example. In contrast to this approach, model-based inference considers **y** to be a single realization of 1 or more stochastic processes. The vector **y** is modeled in terms of a random vector $\mathbf{Y} = (Y_1, \ldots, Y_N)^T$, whose joint distribution $f_{\mathbf{Y}}(\mathbf{y};\theta)$ is characterized by 1 or more unknown, fixed parameters θ. In the model-based view the observable population mean, like **y**, is a chance outcome; therefore the population mean is formulated as a random variable $\mu = \frac{1}{N}\sum_{i=1}^{N} Y_i$. Let $E_f(\mu)$ denote the value of μ that is expected under the model $f_{\mathbf{Y}}(\mathbf{y};\theta)$. The subscript on the $E(\cdot)$ operator emphasizes that the expectation is taken with respect to the model that is assumed to have generated **y**, the particular realization of **Y**. For many models $E_f(\mu)$ is a simple function of one or more model parameters (e.g., $E_f(\mu) = \theta$ for scalar θ), and inferences about model parameters provide inferences about the population mean μ.

To illustrate the difference between design-based inference and model-based inference, consider a population of N = 100 coins that is found on the ground with only 1 side (heads or tails) visible. Suppose 55 heads and 45 tails are visible among the 100 coins. If we let y = 1 for each head and y = 0 for each tail, then in the design-based view the proportion of heads in the population is a fixed constant, which equals the population mean $\mu = [(55)(1) + (45)(0)]/100 = 0.55$. The goal of design-based inference is to estimate this fixed constant from a sample of n coins. In model-based inference the observed set of y-values is assumed to be a random outcome of 1 or more underlying processes. As an example, suppose we assume that N independent coin tosses have produced the observed numbers of heads and tails in the population. Given this assumption, each y-value is modeled as a random outcome of a Bernoulli process, which has the discrete density function $f_Y(y;\pi) = \pi^y (1-\pi)^{(1-y)}$. The unknown parameter π of the Bernoulli distribution corresponds to the probability that each tossed coin has landed heads side up. Under the Bernoulli model, $E_f\left(\frac{1}{N}\sum_{i=1}^{N} Y_i\right) = E_f(\mu) = \pi$ and inferences about π and μ are equivalent. The goal of model-based inference is to estimate the fixed parameter π from a sample of n coins.

The design-based and model-based approaches have somewhat different reference populations. In the former the scope of inference is limited to the surveyed population. Model-based inference, on the other hand, applies to a broader set of populations whose observable characteristics are statistically similar to the surveyed population. The model-based approach is sometimes called a superpopulation view because the surveyed population is regarded as only 1 of several possible finite populations that could have been encountered. The advantages and shortcomings of the 2 approaches are described in the following sections.

Design-based inference

In this view of survey sampling, the observable characteristics of a population, such as **y** or μ, are regarded as fixed constants. Not surprisingly, the choice of sampling design is often motivated by differences in *y*-values anticipated from previous experience. Spatial stratification (shoal vs. channel) and clustering (bedding) are typical examples in surveys of freshwater mussels. The idea is to choose a sampling design that will improve the precision of parameter estimates if differences in *y*-values anticipated in the population are actually realized in the sample.

The sampling designs most commonly used may be classified as conventional or adaptive (Thompson and Seber 1996). In a conventional design the method of sample selection does not depend on any of the observed *y*-values in the population; therefore, all units included in a sample can be selected prior to the survey. Examples include simple random sampling, stratified random sampling, and cluster sampling. Suppose we represent a sampling design mathematically as the conditional probability $P(s_O;\mathbf{y})$ of drawing a sample s_O, given the population vector **y**. In a conventional design $P(s_O;\mathbf{y}) = P(s_O)$ because the sample is selected independent of **y**. Unequal-probability designs, including those in which sample units are selected using auxiliary information known in advance of the survey (e.g., physical size of units), also are conventional.

In an adaptive sampling design, sample selection is based on the sequence of *y*-values observed in the sample $\mathbf{y}_s$, and units are selected for inclusion in s_O during the survey. Examples of adaptive designs are simple random sampling with a sequential stopping rule and adaptive cluster sampling. Adaptive designs are distinguished mathematically as those in which $P(s_O;\mathbf{y}) = P(s_O;\mathbf{y}_s)$.

Textbooks of survey sampling, such as Levy and Lemeshow (1991) and Thompson (1992), describe the practical details of implementing conventional or adaptive designs in surveys. The details include step-by-step procedures for selecting samples and formulas for computing estimates of common population parameters and their variances.

A distinguishing feature of design-based inference is that the estimators (= formulas) used to compute estimates of population parameters, such as μ, are closely linked to the sampling design. A sampling design specifies the method of randomization used to select individual units of the population; therefore, the design induces a probability of selecting one sample over another. An estimator of a population parameter commonly is derived to ensure that the estimator's expected value, taken over all possible samples that might be selected under the design, exactly equals the value of the population parameter. Such estimators are said to be design-unbiased for the parameter. An example is $\hat{\mu} = \frac{1}{n}\sum_{i \in s_O} y_i$, the estimator of the population mean μ based on simple random sampling. It is easily shown that $E_P(\hat{\mu}) = \mu$, where the subscript on the $E(\cdot)$ operator emphasizes that the expectation is taken with respect to all samples that can be selected under the design $P(s_O; \mathbf{y})$. The estimator $\hat{\mu}$ is said to be design-unbiased for μ.

Some estimators used in design-based inference are biased. For example, ratio estimators of the mean, which are intended to exploit an underlying association between the population's *y*-values and 1 or more auxiliary variables, are biased for μ under simple random sampling and unequal probability designs (Thompson 1992). However, many commonly used estimators of population means and totals, which often are the parameters of interest in surveys of mollusks, are design-unbiased.

An important consequence of the linkage between sampling and estimation in design-based inference is that the estimators are unbiased, regardless of the nature of the population. No assumptions about the vector **y** are needed to guarantee the unbiasedness of the estimators. This situation is very different from the assumptions used to derive model-based estimators (see below).

For design-unbiased estimators of μ, inference is commonly based on a version of the Central Limit Theorem that applies to simple random samples selected without replacement from finite populations (Thompson 1992). According to this theorem, the distribution of $(\hat{\mu}-\mu)/\sqrt{\hat{\mathrm{v}}\mathrm{ar}(\hat{\mu})}$, where $\hat{\mu}$ and $\hat{\mathrm{v}}\mathrm{ar}(\hat{\mu})$ are unbiased estimates of the population mean and of the variance of the sample mean, respectively, approaches the standard normal distribution as the numbers of observed (n) and unobserved $(N-n)$ units become large. In other words, the theorem states that the sampling distribution of the estimate is approximately normal, irrespective of the population y-values, if the population mean is estimated from a sufficiently large sample taken from a sufficiently large population. This finite-population version of the Central Limit Theorem applies strictly to simple random samples, but the normal approximation is often used to make inferences in surveys that involve more complicated designs and estimators (Levy and Lemeshow 1991).

The assurance of unbiasedness and of approximate, large-sample normality of design-based estimators is the principal advantage (and attraction) of design-based inference. It is comforting to know that the design-based approach may be used for any population of y-values when little or nothing is known about a population prior to completing a survey.

Design-based estimators have been used only recently in quantitative surveys of freshwater mollusks and with mixed results. Strayer et al. (1996, 1997) used adaptive cluster sampling to compare differences in density of mussels in 13 North American streams. Mussel density was estimated with low precision for all but the most abundant species, even though 30-100 random quadrats were deployed initially at each site and supplemented with additional quadrats to satisfy the requirements of the adaptive design. Other studies are needed to evaluate designs that may lead to improvements in precision, but the results of Strayer's studies illustrate an important shortcoming of design-based inference. For many types of observations, such as counts of sparsely distributed mollusks, the design-unbiased estimates of population parameters may lack precision, irrespective of the type of design (conventional or adaptive). Furthermore, the number of units sampled from these populations may need to be very large before approximate normality, guaranteed by the Central Limit Theorem, is actually achieved. For example, consider a hypothetical population of rare mollusks that is randomly distributed along a stream bottom with a mean density of 0.01 mollusks/m^2. Complete spatial randomness implies that the number of mollusks observed in a 1-m^2 quadrat is expected to have a Poisson distribution with mean 0.01. Now suppose the mollusks in n = 200 randomly selected quadrats are counted to estimate the mean density of the population. A sum of independent Poisson random variables is also Poisson; therefore, the number of mollusks in 200 quadrats is expected to have a Poisson distribution with mean 2.0, which is still highly nonnormal (Jolliffe 1995). Many more quadrats are required for the sampling distribution of the mean density to approximate normality. This hypothetical example is admittedly contrived, but it makes the point that in some populations the advantages of design-based inference may only be realized with infeasibly large samples. If the number of units included in a sample is necessarily limited, which is often the case in surveys of mollusks, other approaches may be needed to make inferences from survey data. An alternative approach that is often useful in small samples is model-based inference.

Model-based inference

In this view of survey sampling, the population vector **y** is considered to be a realization of 1 or more stochastic processes and is therefore modeled as a random vector **Y**. A family of density functions $\{f_{\mathbf{Y}}(\mathbf{y};\theta):\theta\in\Theta\}$, characterized by 1 or more unknown parameters θ in the parameter space Θ, is used to approximate the joint distribution of **Y**. The density functions may be discrete or continuous, depending on the nature of the *y*-values in the population.

Any dependence between sample selection and **y** is specified in the conditional probability $P(s_O;\mathbf{y})$ induced by the sampling design. Because **y** is not fixed in the model-based view, the sample itself can be regarded a chance outcome of the underlying processes assumed to have produced **y**, provided, of course, that **y** is used to select s_O in the design. Therefore, it is possible to model the joint distribution of **Y** and S_O, a random variable used to indicate a sequence of ordered units in the sample, in terms of the probability density function

$$f_{S_O,\mathbf{Y}}(s_O,\mathbf{y};\theta)=P(s_O;\mathbf{y})\cdot f_{\mathbf{Y}}(\mathbf{y};\theta)$$

(Thompson and Seber 1996). However, for conventional and adaptive designs, $P(s_O;\mathbf{y})$ can depend on **y** only through $\mathbf{y}_s$, the vector of *y*-values observed in the sample. Therefore, the joint density of the data $d_O=(s_O,\mathbf{y}_s)$ observed under these designs may be modeled as follows:

$$f_{D_O}(s_O,\mathbf{y}_s;\theta)=P(s_O;\mathbf{y}_s)\cdot f_s(\mathbf{y}_s;\theta),$$

where $f_s(\mathbf{y}_s;\theta)$ is the marginal density of $\mathbf{y}_s$. For other sampling designs, which are called nonstandard by Thompson and Seber (1996), selection of sample units may depend on unobserved *y*-values in the population or on assumed, but unknown, values of model parameters. Nonstandard designs can lead to complicated density functions and are not considered further (see Thompson and Seber 1996 for additional references).

Estimation of the unknown parameters θ is a common objective in model-based inference. Given a well-fitting model, estimates of θ are used to make conclusions about the entire population of *y*-values, including those not observed in the sample. In model-based inference, θ and other population parameters are often estimated by the method of maximum likelihood (Mood et al. 1974), which defines the value of θ that is most likely, given the data and the model, as the maximum-likelihood estimate $\hat{\theta}$. It is computed by maximizing the likelihood function $L(\theta;d_O)$, which equals the joint density of the *n* random variables observed in the sample s_O:

$$L(\theta;d_O)=P(s_O;\mathbf{y}_s)\cdot f_s(\mathbf{y}_s;\theta).$$

In this equation notice that $L(\theta;d_O)$ is a function of θ for the realized set of data $d_O(=(s_O,\mathbf{y}_s))$. This notation emphasizes that once the sample is selected the data are regarded as fixed and that different parameter values provide different levels of support for the *y*-values in the sample. The greatest level of support is provided by $\hat{\theta}$.

This method of estimation highlights an important difference between design-based and model-based approaches. In the former, estimators of population parameters are closely linked to sampling design. In the latter, estimation is based on the likelihood function, which is invariant to conventional and adaptive sampling designs because the sample selection probabilities $P(s_O;\mathbf{y}_s)$ convey no information about θ. Sampling design is also irrelevant in likelihood-based procedures for testing hypotheses. Once a sample of units thought to be representative of the population is obtained using a particular conventional or adaptive design, the probabilities of sample selection become irrelevant in matters of inference. This statement is easily proved with an example. Suppose some data d_O are collected with a particular (conventional or adaptive) design to test the null hypothesis H_0 that the model parameters equal a particular value θ_0 (i.e., H_0: $\theta = \theta_0$). The data in the sample may either support or contradict H_0. The decision to accept or reject H_0 is made by comparing the likelihood that $\theta = \theta_0$ with the likelihood that $\theta = \hat{\theta}$. These likelihoods are represented mathematically as

$$L(\theta_0;d_O) = P(s_O;\mathbf{y}_s)\cdot f_s(\mathbf{y}_s;\theta_0)$$

and

$$L(\hat{\theta};d_O) = P(s_O;\mathbf{y}_s)\cdot f_s(\mathbf{y}_s;\hat{\theta}),$$

respectively. A formal comparison of the 2 likelihoods is made with a likelihood-ratio test statistic (Mood et al. 1974), which is computed as follows:

$$W(d_O) = -2\log\left\{\frac{L(\theta_0;d_O)}{L(\hat{\theta};d_O)}\right\}$$

$$= -2\log\left\{\frac{P(s_O;\mathbf{y}_s)\cdot f_s(\mathbf{y}_s;\theta_0)}{P(s_O;\mathbf{y}_s)\cdot f_s(\mathbf{y}_s;\hat{\theta})}\right\}$$

$$= -2\log\left\{\frac{f_s(\mathbf{y}_s;\theta_0)}{f_s(\mathbf{y}_s;\hat{\theta})}\right\}.$$

If H_0 is correct and the statistical model satisfies the usual regularity conditions (Azzalini 1996), the distribution of $W(d_O)$ converges to a chi-squared distribution as the number of units in the sample increases. The degrees of freedom of the chi-squared distribution equal the difference in degrees of freedom of null and alternative models; therefore, in sufficiently large samples H_0 is rejected if $W(d_O)$ exceeds a pre-determined critical value of the chi-squared distribution. The important thing to notice in this example is that the probabilities of sample selection induced by the sampling design are irrelevant in the test of H_0 because $W(d_O)$ does not depend on $P(s_O;\mathbf{y}_s)$.

Model-based estimators and inferential procedures offer many advantages in the analysis and design of surveys. By taking the model-based view that **y** is a single realization of 1 or more stochastic processes, a researcher has considerable flexibility in identifying and selecting classes of models for approximating the true, underlying processes believed to have generated **y**. Model specification may be guided initially by the nature of the observable *y*-values in the population. For example, consider surveys undertaken to estimate the density of mollusk species from a sample of quadrats. The number of mollusks of each species is observed in each quadrat, and *y*-values are often dominated by counts of 0, 1, or 2 mollusks per quadrat (Green and Young 1993, Strayer and Ralley 1993). Such observations are consequences of using relatively small units (0.25 - 1.0 m^2 /quadrat) to sample an assemblage of sparsely distributed animals. In these cases the sampling distribution of the mean density of mollusks (no./m^2) is almost certainly nonnormal, given the small number of quadrats typically used in surveys of freshwater mollusks. It makes more sense, at least initially, to regard the counts of mollusks observed in such surveys as independent outcomes of a Poisson process.

It is not always easy to find a model that adequately fits the data in surveys of freshwater mollusks. Simple density functions, such as the 1-parameter Poisson, are unsatisfactory when goodness-of-fit statistics reveal that the observed variation in the data exceeds that which is expected under the model. This excess variation or overdispersion is often attributed to spatial clustering or aggregation of mollusks in freshwater assemblages (Elliott 1977). If overdispersion is detected, alternative classes of models must be considered to reallocate the excess variation among systematic and random components of a model. An example is the negative binomial, which can be derived by mixing the Poisson and gamma density functions (Mood et al. 1974). The resulting mixture is a 2-parameter density function that consigns more of the variation in *y*-values to random error and less to the systematic components of the model. Accommodating overdispersion in this manner is generally successful when data contain modest levels of overdispersion. Fortunately, the excess variation in mussel counts that is assumed to be induced by spatial aggregation appears to be low in many natural populations of these animals (Downing and Downing 1992, Strayer et al. 1997).

Another option for dealing with overdispersion is to model the *y*-values in terms of auxiliary information that is measured in the sample and thought to be important in predicting $\mathbf{y}$. Here, the mean response (or a 1:1 function of the mean response) is specified as a function of predictor variables. For example, the logarithm of the mean density of mollusks can be expressed as a linear function of habitat-related measurements in a Poisson regression model. In this way the 1-

parameter Poisson is extended to a family of density functions, providing a framework for exploring relationships between features of mollusk assemblages and their environment. Poisson regression models are included in a larger class of generalized linear models (McCullagh and Nelder 1989), which have a common systematic component (a linear combination of parameters involving predictor variables) but different random components for specifying different forms of error. The random components of generalized linear models comprise the Bernoulli, binomial, Poisson, gamma, normal, and inverse Gaussian density functions; therefore, these models may be widely applicable in analyses of data collected in surveys of freshwater mollusks.

The model-based approach to estimation and inference has an important shortcoming. If the model of the data is not correct, estimates of population parameters and their variances may be biased. This criticism, though technically correct and often cited by advocates of the design-based view, fails to recognize that models are only intended to be approximations of truth that contain both errors of approximation (bias) and errors of estimation (variability). The model-based approach does not seek the true or correct model of the data, only a model that is supported by the level of information contained in the data. To be useful in matters of inference, models should provide a parsimonious balance between errors of approximation and errors of estimation (Linhart and Zucchini 1986), not an overfitting of the data that is unlikely to approximate novel observations made in other samples or in other populations.

The contrasting opinions of design-based and model-based advocates can be traced to inherent differences in the statistical foundations of the 2 approaches. Advocates of the design-based view adopt a frequentist approach to inference, which emphasizes the performance of an estimator or inferential procedure in hypothetically repeated samples of the population and which highly values the property of unbiasedness. In contrast, proponents of the model-based view develop estimators and make inferences based on the likelihood principle (Berger and Wolpert 1984), which requires all evidentiary conclusions about the population to be based on the data observed in the sample. Average performance of an estimator in hypothetically repeated samples is irrelevant in likelihood-based inference (Berger and Wolpert 1984, Edwards 1992).

Design-based vs. model-based inference: an example with sparsely distributed mussels

An example with 2 hypothetical populations of sparsely distributed mussels illustrates some of the differences that may result from applications of design-based and model-based inference. Suppose a survey is undertaken to test whether 2 stream reaches of equal area (100 m long by 20 m wide) contain the same density (no./m^2) of a rare mussel species. In 1 reach the locations of 200 mussels were generated as a realization of a Poisson cluster process (Cressie 1993) parameterized by an average of 20 parent locations and an average radius of dispersion of 2 m between parents and offspring (Fig. 1A). The realized mean density of mussels in this reach was 0.1 mussels/m^2 (= 200/2000). In the 2nd reach the locations of 100 mussels were generated similarly except that the average number of parent locations was reduced to 10 to produce the same average level of small-scale variation (10 offspring/parent) as in the 1st reach (Fig. 1B). In the 2nd reach the realized mean density of mussels was 0.05 mussels/m^2 (= 100/2000). The densities of mussels in these 2 reaches are in the range of densities estimated for threatened and endangered species of unionids (Strayer et al. 1996).

Each reach was partitioned for sampling into 2000 square (1 m x 1 m) quadrats. This procedure resulted in sample units that primarily contained 0, 1, or 2 mussels each, similar to the counts observed in quadrat samples of natural mussel populations (Green and Young 1993, Strayer and Ralley 1993). Two sampling designs were used to compare the density of mussels in the 2 reaches. The 1st was simple random sampling in which 20, 40, 80, 160, or 320 quadrats were randomly selected from each reach. Although commonly used, simple random samples can be impractical in some streams where it is difficult or expensive to relocate people and equipment to different quadrats that are randomly selected but spaced far apart. A 2-stage design was used to collect a random sample of quadrats from each simulated population as an alternative. In this design, each reach 1st was divided into 100 transects that ran across the stream and therefore included 20 quadrats each. In the 1st stage of sampling, 2, 4, 8, 16, or 32 transects were randomly selected from each reach. In the 2nd stage, 10 quadrats were randomly selected from each transect selected in the 1st stage, producing the same set of sample sizes obtained with the simple random sampling design but potentially at lower cost.

The 2 approaches to inference were compared by drawing 5000 independent samples from each reach with simple random sampling or with 2-stage sampling. Two criteria were used to compare the 2 approaches: 1) the statistical power to detect the 2-fold difference in density of mussels in the 2 reaches, and 2) the variance of the estimated mean density of mussels in each reach. Procedures for estimating these criteria are summarized in the appendix.

Model-based and design-based methods of inference provided estimates of mean mussel density with similar variances when quadrats were selected from the 2 reaches by simple random sampling (Table 1). Estimates of the statistical power to detect the difference in mussel density in the 2 reaches also were similar between methods of inference (Fig. 2A). However, power was relatively low (0.07 - 0.41) over the range of sample sizes used. Different results were obtained when quadrats were selected from the 2 reaches by 2-stage sampling. Model-based estimates of mean mussel density were much more precise than the design-based estimates (Table 2). Efficiencies of the model-based estimates relative to the design-based estimates were highest

(2.99 and 3.01) at the lowest sample size (20 quadrats) and declined with increases in sample size. Surprisingly, higher precision of the model-based estimates of mean mussel density at low sample sizes was associated with only minor differences in the statistical power to detect the 2-fold difference in density of mussels in the 2 reaches (Fig. 2B). Apparently, the sampling distribution of the F statistic used in the model-based approach converged more slowly to its asymptotic distribution under H_0 than the sampling distribution of the z statistic. Although needed to ensure that model-based and design-based tests of H_0 were compared at the same level of significance, higher variability of the F values simulated under H_0 was associated with higher critical values of F, producing fewer rejections of H_0 than would have occurred if critical values had been selected from the asymptotic distribution of F.

This example illustrates that detection of differences in density of sparsely distributed mussels may be difficult unless samples include a large number of quadrats. The statistical power of detection was relatively low in samples of 20-320 quadrats, regardless of sampling design (simple-random or 2-stage) or method of inference (model-based or design-based). Samples of <40 quadrats, though common in benthic surveys of macroinvertebrates (Downing and Downing 1992, Resh and McElravy 1993), were inadequate for detecting the 2-fold difference in density of mussels. However, the precision of estimates of mean density was improved in small samples collected by 2-stage sampling when model-based estimators of mean density were used.

Recommendations

Estimation of mean density or abundance of mollusks

Optimal choice of sampling design and method of inference depends on the particular objectives of a survey. An estimate of the mean density of each species is all that is desired in some surveys of freshwater mollusks. In these surveys, the combination of conventional sampling designs and design-based estimators yields estimates that are free of bias but are almost surely imprecise for many species. As described earlier, this result is a consequence of the relatively small size (<1 m^2) of typical units of sampling and of the sparseness that is characteristic of the spatial distribution of many species. In these surveys, model-based estimators can sometimes provide improvements in precision without requiring dramatic increases in sample size. The key is to identify a model that adequately approximates the *y*-values (i.e., counts) observed in the sample. The Poisson density function should be tried initially. If this model is inadequate, parameters may be added to model the extra-Poisson variation.

Models also can be used to select sample size. For example, Green and Young (1993) illustrated how the Poisson and negative binomial density functions may be used to compute the number of sample units needed to detect at least 1 animal in a survey, given prescribed levels of animal abundance and variation. Jackson and Resh (1988, 1989) also used Poisson and negative binomial models to develop sequential sampling designs for assessing environmental impacts. They showed that the expected number of sample units required to detect prescribed differences in species richness, species diversity, or density (i.e., to detect an environmental impact) by sequential sampling was lower than the number of units required by conventional sampling designs.

Adaptive cluster sampling designs and estimators would seem to be ideal for obtaining relatively precise estimates of mean densities of mollusk species because the method was developed for estimating the characteristics of rare, spatially clustered populations of plants and animals (Thompson and Seber 1996). Adaptive cluster sampling has been used in only 1 survey of freshwater mollusks (Strayer et al. 1996, 1997), in which mean density was estimated with low precision for all but the most abundant species. Failure of the adaptive design to yield precise estimates of mollusk density in this instance should not vitiate the approach without further evaluation. For example, a unit is added to an adaptive cluster sample only if the unit is physically adjacent to one that satisfies an arbitrary condition established in advance of the survey. In a survey designed to estimate the mean density of all mollusks, the condition might be that the number of mollusks of all species exceeds some lower limit, say *c*; therefore, if mollusks are spatially clustered, the adaptive procedure selects units with greater numbers of mollusks more efficiently than simple random sampling. The adaptive cluster design is intuitively appealing for the collection of rare animals, but additional research is needed to determine the best condition for extra sampling when estimates of mean density of more than 1 species are required. For example, in a survey of several species of mollusks an adaptive procedure may require extra sampling if 1) the number of mollusks of a particular species exceeds a lower limit *c*, 2) the number of mollusks of any 1 or more species exceeds *c*, 3) the number of mollusks of every species exceeds *c*, or 4) the total number of mollusks of all species exceeds *c*. Other conditions for extra sampling are possible, of course. Thompson and Seber (1996) have shown that design-unbiased estimators of mean density and estimators of their covariances exist regardless of the nature of the requirements for extra sampling. However, malacologists must determine through experience the kinds of conditions for extra sampling that improve the precision of estimates of mean density for most, if not all, species encountered in the survey, including rare species that are often of greatest interest.

Detection of associations between mollusks and their environment

Some surveys of freshwater mollusks are undertaken to determine whether the density or spatial distribution of mollusks is associated with physical or chemical features of their environment. Sampling along an environmental gradient created by a point-source discharge of pollutants in a stream is an example. Model-based inference can be especially helpful in both analysis and design of these surveys. In terms of analysis, a likelihood-ratio comparison of models with and without environmental effects provides a direct test of the strength of association between mussels and their environment. In terms of survey design, models may be used to select the number of sample units needed to detect a mollusk-environmental association with a prescribed level of statistical power. It is important to remember, however, that the survey must still include a random selection of sample units. Randomization helps to ensure that observations are statistically independent and that personal bias is avoided.

Model-based inference is essential in 1 type of sampling design that has never been applied in a survey of freshwater mollusks but may be useful in quantifying the physical habitat of sparsely distributed species. In this design the finite region of measurable area that defines the statistical population is not divided into discrete spatial units. Instead the entire region is searched noting each of the spatial locations where a mollusk is found. The observed set of locations is called a spatial point pattern (Cressie 1993) and may be regarded as a realization of 1 or more spatial point processes. For example, a homogeneous Poisson process corresponds to complete spatial randomness and is generally used as a null model in analyses of spatial point patterns. Models of inhomogeneous Poisson processes may include spatially varying covariates and are used to specify spatial variation in mean density. These models can be used to test whether spatial distributions of mollusks are correlated with the spatial distributions of 1 or more habitat-related variables. Spatial point patterns exploit the relatively sedentary nature of mollusks; however, they are not easily observed if visibility is poor or if detection of mollusks is otherwise impaired. An obvious advantage of spatial point patterns is that they do not require the size, number, or location of sample units to be specified. In addition, the potential covariates of a spatial point pattern do not have to be measured at every location where mollusks are found. Associations between the mean density of mollusks and 1 or more covariates can be tested using interpolated values of the covariates at the locations of the spatial point pattern (Rathbun 1996). Therefore, habitat-related variables may be measured at locations other than those occupied by mollusks, and analyses can still be completed to estimate the level of association between mollusk density and measures of habitat.

Evaluation of survey design appears to be a fertile area for new research with freshwater mollusks. The design-based and model-based approaches described in this paper are not new; however, they have been applied only recently in surveys of mollusks. Rigorous assessments with natural assemblages of mollusks are needed to determine whether the potential benefits of these approaches can be realized in practice.

Acknowledgements

Previous drafts of this paper were improved by the comments of Robert DiStefano, Ananda Ranasinghe, David Strayer, James D. Williams, and an anonymous referee. I also profited from stimulating discussions with presenters and participants at the workshop, Freshwater Mollusks as Indicators of Water Quality, which was held in Atlanta, Georgia, on 4 March 1997.

Literature Cited

Azzalini, A. 1996. Statistical inference based on the likelihood. Chapman and Hall, London, UK.

Barbour, M. T., and J. Gerritsen. 1996. Subsampling benthic samples: a defense of the fixed-count method. Journal of the North American Benthological Society 15:386-391.

Berger, J. O., and R. L. Wolpert. 1984. The likelihood principle. Institute of Mathematical Statistics, Hayward, California.

Cassel, C. M., C. E. Särndal, and J. H. Wretman. 1977. Foundations of inference in survey sampling. Wiley, New York.

Connor, E. F., and E. D. McCoy. 1979. The statistics and biology of the species-area relationship. American Naturalist 113:791-833.

Courtemanch, D. L. 1996. Commentary on the subsampling procedures used for rapid bioassessments. Journal of the North American Benthological Society 15:381-385.

Cressie, N. A. C. 1993. Statistics for spatial data, revised edition. Wiley, New York.

Downing, J. A., and W. L. Downing. 1992. Spatial aggregation, precision, and power in surveys of freshwater mussel populations. Canadian Journal of Fisheries and Aquatic Sciences 49:985-991.

Edwards, A. W. F. 1992. Likelihood. The Johns Hopkins University Press, Baltimore, Maryland.

Elliott, J. M. 1977. Some methods for the statistical analysis of samples of benthic invertebrates. 2nd edition. Freshwater Biological Association Scientific Publication No. 25.

Green, R. H., and R. C. Young. 1993. Sampling to detect rare species. Ecological Applications 3:351-356.

Hansen, M. H., W. G. Madow, and B. J. Tepping. 1983. An evaluation of model-dependent and probability-sampling inferences in sample surveys. Journal of the American Statistical Association 78:776-793.

Hedayat, A. S., and B. K. Sinha. 1991. Design and inference in finite population sampling. Wiley, New York.

Hornbach, D. J., and T. Deneka. 1996. A comparison of a qualitative and a quantitative collection method for examining freshwater mussel assemblages. Journal of the North American Benthological Society 15:587-596.

Jackson, J. K., and V. H. Resh. 1988. Sequential decision plans in monitoring benthic macroinvertebrates: cost savings, classification accuracy, and development of plans. Canadian Journal of Fisheries and Aquatic Sciences 45:280-286.

Jackson, J. K., and V. H. Resh. 1989. Sequential decision plans, benthic macroinvertebrates, and biological monitoring programs. Environmental Management 13:455-468.

Jolliffe, I. T. 1995. Sample sizes and the Central Limit Theorem: the Poisson distribution as an illustration. American Statistician 49:269.

Layzer, J. B., and L. M. Madison. 1995. Microhabitat use by freshwater mussels and recommendations for determining their instream flow needs. Regulated Rivers: Research and Management 10:329-345.

Levy, P. S., and S. Lemeshow. 1991. Sampling of populations: methods and applications. Wiley, New York.

Linhart, H., and W. Zucchini. 1986. Model selection. Wiley, New York.

McCullagh, P., and J. A. Nelder. 1989. Generalized linear models. 2nd edition. Chapman and Hall, London, UK.

Miller, A. C., and B. S. Payne. 1988. The need for quantitative sampling to characterize size demography and density of freshwater mussel communities. American Malacological Bulletin 6:49-54.

Miller, A. C., and B. S. Payne. 1993. Qualitative versus quantitative sampling to evaluate population and community characteristics at a large-river mussel bed. American Midland Naturalist 130:133-145.

Mood, A. M., F. A. Graybill, and D. C. Boes. 1974. Introduction to the theory of statistics. 3rd edition. McGraw-Hill, New York.

Rathbun, S. L. 1996. Estimation of Poisson intensity using partially observed concomitant variables. Biometrics 52:226-242.

Resh, V. H., and E. P. McElravy. 1993. Contemporary quantitative approaches to biomonitoring using benthic macroinvertebrates. Pages 159-194 *in* D. M. Rosenberg and V. H. Resh (editors). Freshwater biomonitoring and benthic macroinvertebrates. Chapman and Hall, New York.

Särndal, C. E. 1978. Design-based and model-based inference in survey sampling. Scandanavian Journal of Statistics 5:27-52.

Strayer, D. L., S. Claypool, and S. J. Sprague. 1997. Assessing unionid populations with quadrats and timed searches. Pages 163-169 *in* K. S. Cummings, A. C. Buchanan, C. A. Mayer, and T. J. Naimo (editors). Conservation and management of freshwater mussels II: initiatives for the future. Upper Mississippi River Conservation Committee, Rock Island, Illinois.

Strayer, D. L., and J. Ralley. 1993. Microhabitat use by an assemblage of stream-dwelling unionaceans (Bivalvia), including two rare species of *Alasmidonta*. Journal of the North American Benthological Society 12:247-258.

Strayer, D. L., S. J. Sprague, and S. Claypool. 1996. A range-wide assessment of populations of *Alasmidonta heterodon*, an endangered freshwater mussel (Bivalvia:Unionidae). Journal of the North American Benthological Society 15:308-317.

Thompson, S. K. 1992. Sampling. Wiley, New York.

Thompson, S. K., and G. A. F. Seber. 1996. Adaptive sampling. Wiley, New York.

Vaughn, C. C., C. M. Taylor, and K. J. Eberhard. 1997. A comparison of the effectiveness of timed searches vs. quadrat sampling in mussel surveys. Pages 157-162 *in* K. S. Cummings, A. C. Buchanan, C. A. Mayer, and T. J. Naimo (editors). Conservation and management of freshwater mussels II: initiatives for the future. Upper Mississippi River Conservation Committee, Rock Island, Illinois.

Vinson, M. R., and C. P. Hawkins. 1996. Effects of sampling area and subsampling procedure on comparisons of taxa richness among streams. Journal of the North American Benthological Society 15:392-399.

Watters, G. T. 1992. Unionids, fishes, and the species-area curve. Journal of Biogeography 19:481-490.

Appendix: Simulation-based estimation of statistical power and variance

Design-unbiased estimates of the mean density of mussels in the *j*th stream reach ($\hat{\mu}_j$, j = 1,2) and of the variance of this estimate ($\hat{\mathrm{v}}\mathrm{ar}(\hat{\mu}_j)$) were computed in each sample to test the null hypothesis H_0 of equal densities of mussels in the 2 reaches. Tests of H_0 were based on the test statistic

$$z = \frac{(\hat{\mu}_1 - \hat{\mu}_2)}{\sqrt{\hat{\mathrm{v}}\mathrm{ar}(\hat{\mu}_1) + \hat{\mathrm{v}}\mathrm{ar}(\hat{\mu}_2)}}.$$

When samples contain a sufficiently large number of quadrats, the distribution of z approximates a standard normal distribution $N(0,1)$ if H_0 is true, allowing quantiles of $N(0,1)$ to be used as critical values for testing H_0. However, not all samples selected from the 2 reaches were sufficiently large to satisfy the standard normal approximation; therefore, the sampling distribution of z under H_0 was simulated by combining the units from both populations into 1 population, drawing a sample of twice the required size from the combined units, and then randomly allocating the units in this combined sample into 2 samples of equal size. This procedure was repeated 5000 times, computing z for each pair of samples. The $\alpha/2$ and $1-\alpha/2$ quantiles of these z values (denoted $z(\alpha/2)$ and $z(1-\alpha/2)$, respectively) were used as critical values to test H_0 at the α level of significance; therefore, H_0 was rejected at the 0.05 significance level if $z < z(0.025)$ or $z > z(0.975)$. The statistical power of detecting the difference in mean density of mussels in the 2 reaches was estimated as the proportion of the 5000 samples in which H_0 was rejected. The variance of the estimated mean density of mussels in each reach, $\mathrm{var}(\hat{\mu}_j)$, was estimated by averaging the variance estimates $\hat{\mathrm{v}}\mathrm{ar}(\hat{\mu}_j)$ computed for each of the 5000 samples.

Because the entire population of y-values is known, $\text{var}(\hat{\mu}_j)$ could have been computed exactly; however, sample estimates of this variance were averaged to provide a direct comparison with the average of the model-based variance estimates calculated for each sample. By using 5000 samples, the difference between $\text{var}(\hat{\mu}_j)$ and the average of $\hat{\text{v}}\text{ar}(\hat{\mu}_j)$ was negligible for all sample sizes examined.

To apply model-based inference in the test of H_0, the number of mussels per quadrat was modeled with a Poisson density function, modified for overdispersion. Specifically, I assumed that $E_f(Y_{ji}) = \lambda_j$, where Y_{ji} denotes a random variable for the number of mussels in the ith quadrat (i = 1, …, N_j) of the jth reach (j = 1, 2) and $E_f(\cdot)$ indicates expectation with respect to the Poisson model

$$f_{Y_{ji}}(y_{ji};\lambda_j) = \frac{\exp(-\lambda_j)\cdot(\lambda_j)^{y_{ji}}}{y_{ji}!}.$$

To account for overdispersion, I assumed that the variance of Y_{ji} expected under the model is $\text{var}_f(Y_{ji}) = \phi\cdot\lambda_j$, where ϕ is a positive constant to be estimated from the data. Values of ϕ that exceed unity indicate overdispersion, whereas $\phi = 1$ indicates that the distribution of Y_{ji} follows a regular Poisson density function. Modeling overdispersion in this way is called quasi-likelihood (McCullagh and Nelder 1989) and is often useful when the exact mechanism generating overdispersion is unknown.

Given the overdispersed Poisson model, inferences about the mean density of mussels in the 2 reaches are equivalent to inferences about the Poisson parameters λ_j because $E_f(\mu_j) = E_f\left(\frac{1}{N_j}\sum_{i=1}^{N_j} Y_{ji}\right) = \left(\frac{1}{N_j}\right) N_j\lambda_j = \lambda_j$. Therefore, the null hypothesis H_0 of equal densities of mussels in the 2 reaches can be restated in terms of the model's parameters:

H_0: $\lambda_1 = \lambda_2$.

Each of the 5000 samples was used to test H_0 by computing likelihood-based estimates of model parameters and test statistics. The common Poisson mean λ, which parameterizes the distribution of Y_{ji} under H_0, was computed using the maximum-likelihood estimator

$$\hat{\lambda} = \frac{1}{(n_1 + n_2)} \sum_{j=1}^{2} \sum_{i=1}^{n_j} y_{ji}$$

where y_{ji} is the number of mussels observed in the *i*th quadrat ($i = 1, \ldots, n_j$) sampled from the *j*th reach. Distinct Poisson means, which hold when H_0 is false, were estimated by

$$\hat{\lambda}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ji} .$$

Overdispersion in the mussel counts was estimated using the conventional quasi-likelihood estimator of ϕ (McCullagh and Nelder 1989):

$$\hat{\phi} = \frac{1}{(n_1 + n_2 - 2)} \sum_{j=1}^{2} \sum_{i=1}^{n_j} \frac{\left(y_{ji} - \hat{\lambda}_j\right)^2}{\hat{\lambda}_j} .$$

Tests of H_0 were based on the statistic $F = W / \hat{\phi}$, where W is the likelihood-ratio statistic

$$W = 2\left[n_1 \hat{\lambda}_1 \log\left(\frac{\hat{\lambda}_1}{\hat{\lambda}} \right) + n_2 \hat{\lambda}_2 \log\left(\frac{\hat{\lambda}_2}{\hat{\lambda}} \right) \right].$$

Under H_0 the distributions of *W* and $\hat{\phi} \cdot (n_1 + n_2 - 2)$ converge asymptotically to chi-squared distributions based on 1 and $n_1 + n_2 - 2$ degrees of freedom, respectively (McCullagh and Nelder 1989). Therefore, when H_0 is true and samples are sufficiently large, the test statistic $F = W/\hat{\phi}$ approximates an *F* distribution parameterized by 1 and $n_1 + n_2 - 2$ degrees of freedom, and quantiles of $F(1, n_1 + n_2 - 2)$ may be used as critical values for testing H_0. However, not all samples selected from the 2 reaches were sufficiently large to satisfy this approximation; therefore, the sampling distribution of *F* under H_0 was simulated by drawing samples from the combined units of both populations as was done for *z* (see above). The 1-α quantile of the simulated distribution of *F* values, denoted *F*(1-α), was used to test H_0 at the α level of significance; therefore, H_0 was rejected at the 0.05 significance level if $F > F(0.95)$. The statistical power of detecting the difference in mean density of mussels in the 2 reaches was estimated as the proportion of the 5000 samples in which H_0 was rejected.

The variance of the model-based estimates of mean density of mussels in each reach, $\mathrm{var}(\hat{\lambda}_j)$, was estimated by averaging the variance estimates $\hat{\mathrm{v}}\mathrm{ar}(\hat{\lambda}_j)$ computed for each of the 5000 samples. These variance estimates were based on the quasi-likelihood assumptions and were computed as follows: $\hat{\mathrm{v}}\mathrm{ar}(\hat{\lambda}_j) = \hat{\phi} \cdot \hat{\lambda}_j / n_j$.

Table 1. Variances of model-based and design-based estimates of mean density of mussels based on simple random samples of quadrats from each stream reach shown in Fig. 1. Efficiency of the model-based estimator $\hat{\lambda}_j$ relative to the design-based estimator $\hat{\mu}_j$ is indicated by the ratio of their variances: eff($\hat{\lambda}_j$) = $\mathrm{var}(\hat{\mu}_j)/\mathrm{var}(\hat{\lambda}_j)$.

No. quadrats	$\mathrm{var}(\hat{\lambda}_1)$	$\mathrm{var}(\hat{\mu}_1)$	eff($\hat{\lambda}_1$)	$\mathrm{var}(\hat{\lambda}_2)$	$\mathrm{var}(\hat{\mu}_2)$	eff($\hat{\lambda}_2$)
20	0.00648	0.00900	1.39	0.00379	0.00469	1.24
40	0.00372	0.00439	1.18	0.00216	0.00237	1.10
80	0.00210	0.00216	1.03	0.00112	0.00115	1.03
160	0.00112	0.00103	0.92	0.00057	0.00055	0.96
320	0.00057	0.00047	0.82	0.00029	0.00025	0.86

Table 2. Variances of model-based and design-based estimates of mean density of mussels based on 2-stage samples of quadrats from each stream reach shown in Fig. 1. Efficiency of the model-based estimator $\hat{\lambda}_j$ relative to the design-based estimator $\hat{\mu}_j$ is indicated by the ratio of their variances: eff($\hat{\lambda}_j$) = $\mathrm{var}(\hat{\mu}_j)/\mathrm{var}(\hat{\lambda}_j)$.

No. quadrats	$\mathrm{var}(\hat{\lambda}_1)$	$\mathrm{var}(\hat{\mu}_1)$	eff($\hat{\lambda}_1$)	$\mathrm{var}(\hat{\lambda}_2)$	$\mathrm{var}(\hat{\mu}_2)$	eff($\hat{\lambda}_2$)
20	0.00529	0.01583	2.99	0.00322	0.00970	3.01
40	0.00315	0.00774	2.46	0.00200	0.00513	2.57
80	0.00188	0.00381	2.03	0.00110	0.00248	2.25
160	0.00106	0.00176	1.66	0.00056	0.00112	2.00
320	0.00056	0.00076	1.36	0.00029	0.00048	1.66

Figure Legends

Fig. 1. Simulated spatial distributions of a rare mussel species in 2 stream reaches of identical size and shape (plotted with a 1:1 aspect ratio). Locations of mussels in each reach were generated as a realization of a Poisson cluster process (see text). Each reach contains a different mean density of mussels: 0.1 mussels/m^2 (A) and 0.05 mussels/m^2 (B).

Fig. 2. Estimates of statistical power of detecting a difference in mean density of mussels (at $\alpha = 0.05$) using samples selected from the 2 stream reaches shown in Fig. 1. Quadrats were selected by simple random sampling (A) or by 2-stage sampling (B). Symbols indicate whether power calculations were based on design-based inference (Δ) or model-based inference (○).