Environmental and Ecological Statistics **12,** 55–94, 2005

A review of adaptive cluster sampling: 1990–2003

PHILIP TURK and JOHN J. BORKOWSKI

Department of Mathematical Sciences, Montana State University, Bozeman, 59717, MT, USA
E-mail: dfletcher@maths.otago.ac.nz

Received April 2002; Revised July 2004

Adaptive cluster sampling (ACS) is an adaptive sampling scheme which operates under the rule that when the observed value of an initially selected sampling unit satisfies some condition of interest, C, other additional units in some pre-defined accompanying neighborhood are also added to the sample. In turn, if any of these additional units satisfy C, then their corresponding unit neighborhoods are added to the sample as well, and so on. This process stops when no additional units satisfying C are encountered. This paper will provide a review of the major developments and issues in ACS since its introduction by Thompson (1990) [*Journal of the American Statistical Association*, **85**, 1050–1059].

Keywords: Bootstrapping, cluster sampling, detectability, double sampling, Rao–Blackwell estimator, stratified sampling

1. Introduction

With an adaptive sampling scheme the procedure for selecting units to include in the sample may depend on values of the variable of interest observed during the survey, i.e. the sampling is "adapted" to the data (Thompson and Seber, 1996). Adaptive cluster sampling (ACS) is a specific type of adaptive scheme. With ACS, we assume an initial sample of size n_1 is taken from the population. ACS then operates under the rule that when the observed value of an initially sampled unit satisfies some condition of interest, C, additional units in some pre-defined neighborhood will be added to the sample. Then, if any of these additional units satisfy C, the units in their neighborhoods are added to the sample as well, and so on. This process is iterated until no units satisfying C are encountered. "In general terms, it means that if you find what you are looking for at a particular location you sample in the vicinity of

1352-8505 © 2005 Springer Science + Business Media, Inc.

that location with the hope of obtaining even more information.'' (Salehi and Seber, 1997a, p. 959).

At this point, we have clusters of units. Each *cluster* contains units satisfying C and a boundary of edge units. An *edge unit* is any unit not satisfying C but in the neighborhood of one that does. The final sample consists of n_1 clusters which are not necessarily unique because two units satisfying C could have been selected from the same cluster in the initial sample. If a unit in the initial sample does not satisfy C, then, by definition, it is considered to be a cluster of size 1.

Neighborhoods can be defined in a variety of ways although the *first-order neighborhood* consisting of the unit itself and the four adjacent units sharing a common boundary (denoted as north, south, east and west) is by far the most prevalent. The *second-order neighborhood* contains eight units consisting of the first-order neighborhood plus the northwest, northeast, southwest, and southeast units. These two neighborhood types would be appropriate for studies in which the y-values satisfying C tend to cluster and are not oriented in any particular direction (Christman, 2000). However, the neighborhood relationship is assumed to be symmetric. Also, the neighborhoods do not depend on the data. Technically, the units in the neighborhood don't have to be physically contiguous. For example, the neighborhood may be defined through a social relationship between units, i.e., people. A person's neighborhood could be all people sexually linked people to the individual with the network units being the subset of people (neighbors) that are HIV positive and edges units being the subset of neighbors that are HIV negative (Dryver and Thompson, 1998).

Let A_i be the network generated by unit i where a *network* is defined to be a cluster generated by that unit but with its edge units removed. Note that any unit in A_i leads to the selection of all units in A_i (the network symmetry assumption). By definition, any unit that does not satisfy C is, by definition, a network of size 1. Thus, any multi-unit cluster can be decomposed into one network whose units satisfy C and individual networks (edge units) of size 1 that do not satisfy C. It is important to note the distinction between clusters and networks. Clusters are not necessarily disjoint because they may have overlapping edge units. On the other hand, the entire population of units can be partitioned into a set of disjoint and exhaustive networks.

Consider a region divided into N rectangular units, called quadrats, of equal area that can be arranged in a grid. Assume the units $(u_1, u_2, \ldots, u_N)$ form a mutually exclusive and exhaustive partition of the region such that the unit labels $(1, 2, \ldots, N)$ identify the locations of N quadrats. Unit i will be associated with a response of interest y_i, $i = 1, 2, \ldots, N$. Thus, we have a population vector of y-values, $\mathbf{y} = (y_1, y_2, \ldots, y_N)$. In the fixed population view, $\mathbf{y}$ is considered to be a fixed set of unknown constants. Thus, alternatively, we may write $\mathbf{y} = \theta$ to represent the values as a vector of unknown parameters. Most of the following theory and notation in this review is attributed to Thompson and Seber (1996). Throughout this review we will consider the population in terms of quadrats. The applications will be primarily ecological given this is the focus of the vast majority of the research cited in this review. We do note, however, that ACS can be applied to other types of sampling units (e.g., homes in an urban area) with an appropriate definition of a neighborhood.

For example, Fig. 1 describes the blue-winged teal population given in the often-cited Smith *et al.* (1995) paper. The 200 quadrats represent 25 km^2 regions with counts given for quadrats having nonzero teal counts. A simple random sample of 10 quadrats was selected. These 10 quadrats are indicated with $\otimes$ in Figs. 2 and 3. The quadrats with $\bigcirc$ in Figs 2 and 3 indicate the additional units that would be sampled

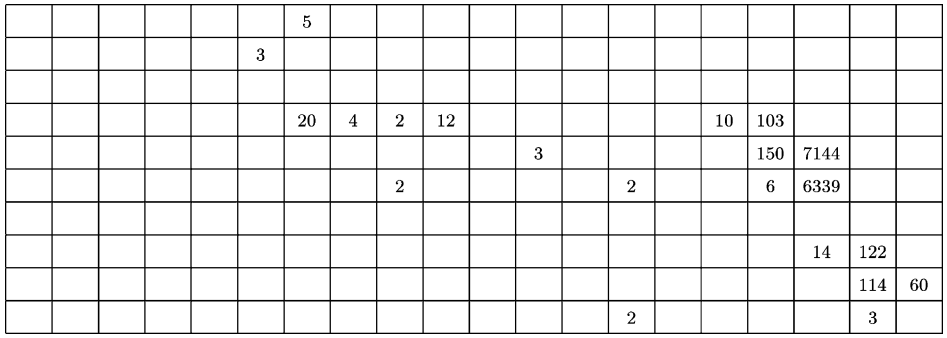

Figure 1. Blue-winged teal population from Smit *et al.* (1995).

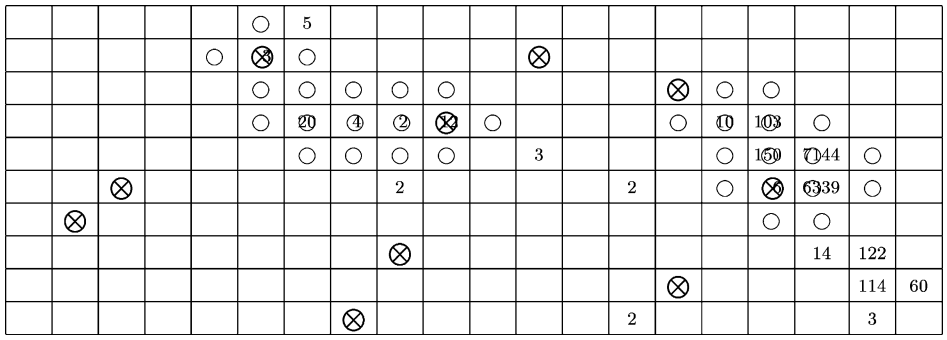

Figure 2. An adaptive cluster sample (C: $y_i \geq 1$).

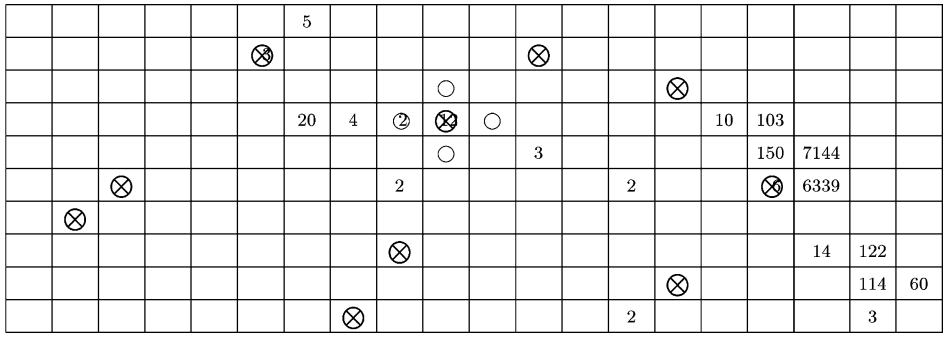

Figure 3. An adaptive cluster sample (C: $y_i \geq 10$).

with ACS when C is $y_i \geq 1$ and when C is $y_i \geq 10$, respectively. Quadrats with ○ but teal counts not satisfying C are edge units. A first-order neighborhood is assumed.

The primary application of ACS is with populations whose individuals tend to cluster and the number of individuals is relatively small. In these cases, if a classical design (e.g., simple random sampling) were to be used, the majority of measurements will be 0 and many clusters will be missed. Consequently, variances associated with estimation of the population mean or total will be large and resulting confidence intervals for corresponding parameters will be wide (Seber and Thompson, 1994).

Fields of study where ACS has been effectively and successfully used and where such populations naturally occur include, but are not confined to, ecology, biology, epidemiology, environmental sciences, demography and geology (Thompson, 1990; Brown, 1994; Seber and Thompson, 1994; Thompson and Seber, 1996; Dryver and Thompson, 1998). ACS may be applied to populations of animals and plants that have aggregation tendencies due to such factors as schooling, flocking, dispersal patterns, and environmental patchiness (Thompson, 1991a). For example, Smit *et al.* (2003) applied ACS techniques to a survey of freshwater mussels which tend to be spatially clustered and found at low density.

2. ACS versus classical designs

In this section, many of the differences between ACS and the classical designs will be discussed. Advantages and disadvantages of both designs will be reviewed.

With the classical designs, the entire selection of sample units in the sample is made prior to observing any y_i response. With ACS, however, the selection of units in the sample depends on observed y_i values. Despite these differences, the objective is the same: to select a sample of units, observe the corresponding y-values, and then estimate some function $z(\mathbf{y})$ (or, equivalently, $z(\theta)$) such as the population total $z(\mathbf{y}) = \sum_{i=1}^{N} y_i = \tau$, the population mean $z(\mathbf{y}) = \frac{1}{N}\sum_{i=1}^{N} y_i = \mu$, or the finite-population variance $z(\mathbf{y}) = \sum_{i=1}^{N} \frac{(y_i-\mu)^2}{N-1} = \sigma^2$.

For many classical sampling plans, the sample selection probability function $P_\theta(s)$ is nonzero and constant where the sample s is a sequence of labels identifying the units to be sampled. Thus, $P_\theta(s)$ does not depend on θ. On the other hand, for ACS, $P_\theta(s)$ is sometimes zero or nonconstant, i.e., it does depend on θ. Moreover, ACS assigns higher sample selection probabilities to samples that include units where y_i satisfies C. A similar situation exists regarding marginal inclusion probabilities (π_i) for the units where π_i is the probability that unit i is included in the sample. Many classical plans are self-weighting, i.e., $\pi_i = \pi$ for $i = 1, 2, ..., N$. However, for ACS, this is no longer the case because the π_i depend on θ (through the choice of C) and, therefore, are no longer constant (Seber and Thompson, 1994). Another important difference is that classical estimators used in conjunction with ACS will be biased. However, modified estimators exist for ACS that are design-unbiased for the population mean along with design-unbiased estimators of their variances. That is, the unbiasedness depends on the way the sample is selected and does not depend on any assumptions about the population. Also, the Rao–Blackwell method has been used to obtain smaller variance design-unbiased estimators for ACS. This will be dis-

cussed in Section 5. It should also be noted that although design-unbiased estimators are the ones generally used when conducting ACS, estimators based on models are used on occasion. A discussion of frequentist model-based and likelihood model-based estimation along with stochastic population theory and examples can be found in Chapter 3 of Thompson and Seber (1996).

There are numerous advantages that ACS may have over the classical designs. In many instances ACS can be a more efficient design, i.e. the variances of estimators will be smaller for an equivalent amount of sampling effort. Estimators derived via ACS make use of unit labels which can result in lower variance than estimators that do not make use of labels (Seber and Thompson, 1994; Thompson and Seber, 1996). Virtually every ACS paper cited in this review deals to some extent with conditions under which ACS is a more efficient design for an equivalent amount of sampling effort. Because the location and shape of clusters of individuals are oftentimes not known prior to the survey, ACS can be used in situations where stratification may not be possible (Thompson, 1991b). ACS can also identify local maxima or areas of high abundance, thus increasing the information provided by the sample. This attribute can be of particular importance, for example, when sampling rare or endangered species where gathering as much information as possible about individuals is desired (Seber and Thompson, 1994). Another major advantage of the ACS design is its flexibility in terms of its construction. The researcher determines the initial sample size, the unit size, the configuration of the neighborhood, and the sampling condition C relative to the situation to achieve the most efficient design. Furthermore a variety of methods for choosing the initial sample exist, such as simple random sampling with replacement (SRSWR) and simple random sampling without replacement (SRSWOR) (Thompson, 1990), strip sampling, systematic sampling (Thompson, 1991a), stratified sampling (Thompson, 1991b), sampling via probability proportional to size (Roesch, 1993; Smith *et al.*, 1995; Pontius, 1997), and simple Latin square sampling (Borkowski, 1999). Finally, two other potential advantages from a cost standpoint are that the average distances between sampled units are smaller and the quadrat locations are easier to find (Salehi and Seber, 1997a; Brown and Manly, 1998).

There are, however, some disadvantages to ACS relative to the classical designs. In particular, the final number of distinct sampled units and both initial and final number of distinct sampled networks are random. Consequently, it can be difficult to determine and control the total sampling effort and cost of the survey in advance, which is oftentimes necessary for some surveys. Several recent papers have dealt with this issue and they will be discussed in Section 6. The theory for using a pilot study to design an ACS survey with a given efficiency or expected cost has been addressed by (Salehi and Seber, 1997b). Ironically, it is the very flexibility of ACS that makes finding variance optimal designs complicated. For example, the selection of C can be critical to the efficiency of the ACS design. As with the classical designs, there exists a positive probability that a sample will yield a zero abundance estimate. This has been addressed by Christman and Lan (1998). Another important distinction between the two types of designs is that not all the information from sampled units is used in ACS. Specifically, edge units are only used if they are part of the initial sample. This concern has been addressed with the development of Rao–Blackwell estimators that incorporate information from edge units (see Section 5).

2.1 *ACS with an initial SRSWOR*

Suppose that an initial sample of n_1 units is taken via SRSWOR. Recall that if the y-value of a sampled unit satisfies a certain condition C, for example $y_i \geq c_a$ where c_a is a constant, then the rest of the unit's neighborhood is added to the sample. Therefore, unit i will be included in the final sample either if any unit of network A_i, the network containing unit i, is selected as part of the initial simple random sample or if any unit of a network for which unit i is an edge unit is selected. Unfortunately, π_i, the inclusion probability for unit i, cannot be computed from the sample data. Instead, we consider the empirically derived *partial marginal inclusion probability*

$$\pi_i' = 1 - \left[\binom{N - m_i}{n_1} \Big/ \binom{N}{n_1} \right] \tag{1}$$

where m_i is the number of units in A_i, the network containing unit i. Consequently, we are now dealing with a sample of n_1 networks with replication of networks being possible. π_i' can be interpreted as being the probability that at least one unit in the initial sample intersects network A_i.

Thompson (1990) developed an estimator $\widehat{\mu}_{HT}$ based on a modification of the estimator of Horvitz and Thompson (1952) using the partial inclusion probabilities:

$$\widehat{\mu}_{HT} = \frac{1}{N} \sum_{i=1}^{N} \frac{y_i I_i}{\pi_i'} \tag{2}$$

where $I_i = 1$ with probability π_i' if the initial sample intersects A_i, and $I_i = 0$ otherwise. It can also be used when initial sampling is SRSWR but with the appropriate adjustment to (1) for sampling with replacement (Thompson and Seber, 1996, p. 100). Importantly, note that observations that do not satisfy C are ignored if they are not included in the initial sample. Hence, π_i' can also be interpreted as being the probability unit i is used in the estimator.

Another way of expressing (2) is in terms of networks (instead of units). That is, let K be the total number of distinct networks in the population, and $I_k = 1$ with probability α_k if the initial sample intersects the kth network, and $I_k = 0$ otherwise. Then

$$\widehat{\mu}_{HT} = \frac{1}{N} \sum_{k=1}^{K} \frac{y_k^* I_k}{\alpha_k} \tag{3}$$

where y_k^* is the sum of the y-values for the kth network. For each unit i in the kth network, π_i' will be the same, i.e. equal to α_k. If there are x_k units in the kth network, then

$$\alpha_k = 1 - \left[\binom{N - x_k}{n_1} \Big/ \binom{N}{n_1} \right] \tag{4}$$

It should be emphasized that α_k is not the actual network inclusion probability. α_k is referred to as the marginal *initial intersection probability*, or the probability that at least one unit in the initial sample intersects the kth network. If there are x_k units in the kth network and x_j units in the jth network, then the *joint initial intersection*

probability, i.e., the probability that networks j and k are both intersected by at least one unit each in the initial sample, is:

$$\alpha_{jk} = 1 - \left[\binom{N - x_j}{n_1} + \binom{N - x_k}{n_1} - \binom{N - x_j - x_k}{n_1}\right] \Big/ \binom{N}{n_1} \tag{5}$$

Thompson (1990) showed that $\widehat{\mu}_{HT}$ is unbiased and

$$\mathrm{Var}[\widehat{\mu}_{HT}] = \frac{1}{N^2}\left[\sum_{j=1}^{K}\sum_{k=1}^{K} y_j^* y_k^* \left(\frac{\alpha_{jk} - \alpha_j\alpha_k}{\alpha_j\alpha_k}\right)\right] \tag{6}$$

with unbiased estimator of the variance

$$\widehat{\mathrm{Var}}[\widehat{\mu}_{HT}] = \frac{1}{N^2}\left[\sum_{j=1}^{K}\sum_{k=1}^{K} y_j^* y_k^* \left(\frac{\alpha_{jk} - \alpha_j\alpha_k}{\alpha_{jk}\alpha_j\alpha_k}\right) I_j I_k\right] \tag{7}$$

where, by definition, $\alpha_{jj} = \alpha_j$. Note that if $\alpha_k \propto y_k^*$, $\mathrm{Var}[\widehat{\mu}_{HT}]$ will be small because $\widehat{\mu}_{HT}$ will be relatively constant (Thompson and Seber, 1996, pp. 95–97). Also note that necessarily $\alpha_{jk} > 0$ for all pairs of networks $j \neq k$. With only minor modifications, the preceding Horvitz–Thompson estimators can be used when the initial sample is conducted with replacement. The reader is referred to Thompson (1990) for further details.

Thompson (1990) also developed an estimator $\widehat{\mu}_{HH}$ based on a modification of the Hansen and Hurwitz (1943) estimator which can be used when sampling is either WR or WOR. As was the case with $\widehat{\mu}_{HT}$, units that do not satisfy C are ignored if they are not in the initial sample. In terms of the n_1 networks (not necessarily unique) intersected by the initial sample:

$$\widehat{\mu}_{HH} = \frac{1}{n_1}\sum_{i=1}^{n_1} w_i = \overline{w} \tag{8}$$

where $w_i = \frac{1}{m_i}\sum_{j\in A_i}(y_j)$ is the mean of the m_i observations in A_i. Thompson (1990) showed that $\widehat{\mu}_{HH}$ is unbiased and

$$\mathrm{Var}[\widehat{\mu}_{HH}] = \frac{N - n_1}{Nn_1(N-1)}\sum_{i=1}^{N}(w_i - \mu)^2 \;=\; \frac{N - n_1}{Nn_1}\sigma_B^2 \tag{9}$$

where σ_B^2 is the between-network variance. An unbiased estimator of $\mathrm{Var}[\widehat{\mu}_{HH}]$ is

$$\widehat{\mathrm{Var}}[\widehat{\mu}_{HH}] = \frac{N - n_1}{Nn_1(n_1 - 1)}\sum_{i=1}^{n_1}(w_i - \widehat{\mu}_{HH})^2. \tag{10}$$

The asymptotic distributions of $\widehat{\mu}_{HT}$ and $\widehat{\mu}_{HH}$ were studied by DiConsigilio and Scanu (2001). Sufficient (but not necessary) conditions for asymptotic normality of $\widehat{\mu}_{HT}$ and $\widehat{\mu}_{HH}$ are given. Their simulation results confirm that the distributions of $\widehat{\mu}_{HT}$ and $\widehat{\mu}_{HH}$ are approximately normal when the population is composed by a large number of well-segregated small clusters. Thus, if there is a large dominating network with a high probability of being observed (which we will refer to as a *hyper-network*), then it is not reasonable to assume normality for the distribution of the

estimators. For the populations considered, the simulation results indicated that the coverage percentages of confidence intervals were closer to the nominal 95% level using Hansen–Hurwitz estimation than for Horvitz–Thompson estimation.

Felix-Medina (2003) also studied the asymptotic properties of $\widehat{\mu}_{HT}$ and $\widehat{\mu}_{HH}$. He developed an asymptotic framework based on four assumptions applied to a sequence of populations $\{U_t\}_{t=1}^{\infty}$ and samples. The essence of these four assumptions is that the number of units in the initial sample and the number of units and networks in the population tend to infinity while restricting the network sizes to be bounded. Felix-Medina proved that for the given assumptions, the variances of $\widehat{\mu}_{HT}$ and $\widehat{\mu}_{HH}$ converge to 0 as $t \to \infty$. Using this result and the unbiasedness of the estimators, $\widehat{\mu}_{HT}$ and $\widehat{\mu}_{HH}$ are, therefore, design-consistent. The estimators were also shown to be asymptotically normally distributed. He also includes a warning that despite their consistency, the estimators may behave poorly in the finite-sample setting. For Horvitz–Thompson estimation, if the y-values and intersection probabilities are not correlated, the variance estimators are unstable and can yield negative variance estimates.

Because it is unlikely that prior knowledge exists regarding the strength of the correlation and because it will be impossible to verify the set of assumptions, Felix-Medina (2003) suggests the set of conditions be used only as general guidelines. If the finite population to be sampled has a large number of units, is partitioned into a large number of networks, and the initial sample is large, then this situation is in reasonable agreement with the assumptions. A major problem occurs for the common ACS criterion C of adaptively sampling if $y > 0$. Typically, both the number of networks satisfying C and the initial sample size will be small. When this occurs, it is not recommended to assume the estimators are approximately normally distributed. Because the results of Felix-Medina (2003) are strictly valid only when both the initial sample size and population size tend to infinity, additional research is required to study the behavior of the estimators for typical sampling scenarios.

2.2 *Comparisons to SRS*

Because each unit has a different selection probability, the sample mean $\overline{y}$ corresponding to the *final* ACS sample (the initial sample plus the adaptively sampled units) as well as the average of the ACS sample cluster means in the final ACS sample, are both biased estimators of μ (Thompson 1990). This is unlike the *initial* sample mean based only on units initially selected via SRSWOR or SRSWR which is unbiased. Analogously, the sample variance s^2 corresponding to the *final* ACS sample is biased for estimating σ^2.

Using an example with an initial sample of fixed size n_1 selected via SRSWOR from a small population, Thompson (1990) demonstrated that an ACS strategy can be more efficient than a SRS strategy. In other words, the variance of the estimators $\widehat{\mu}_{HT}$ in (6) and $\widehat{\mu}_{HH}$ in (9) can both be less than the variance of $\overline{y}$ from SRSWOR. For comparative purposes, the expected effective sample size $E[\nu]$ under ACS was used in lieu of n in the formula for $\mathrm{Var}[\overline{y}]$.

In another example, Thompson (1990) used a rare, clustered population produced via simulation and having C as $y_i \geq 1$. He again compared the ACS and SRS design

strategies for a variety of initial sample sizes n_1. The relative efficiency of estimator $\widehat{\mu}_i$ ($i = HT, HH$) was calculated as

$$\text{efficiency}(\widehat{\mu}_i) = \text{Var}[\overline{y}]/\text{Var}[\widehat{\mu}_i]. \tag{11}$$

Thus, $\widehat{\mu}_i$ is considered a more efficient estimator when efficiency$(\widehat{\mu}_i) > 1$. Thompson's results showed that both $\widehat{\mu}_{HT}$ and $\widehat{\mu}_{HH}$ became increasingly more efficient as n_1 increased, with the increases being far greater for $\widehat{\mu}_{HT}$ than for $\widehat{\mu}_{HH}$. Next, the y-values were set to indicate the presence ($y_i = 1$) or absence ($y_i = 0$) of point objects in unit i. Of interest is the efficiency of $\widehat{\mu}_{HH}$ which was worse (< 1) regardless of the initial sample size. The efficiency of $\widehat{\mu}_{HT}$, however, was < 1 only for smaller n_1, while for larger n_1 the efficiency of $\widehat{\mu}_{HT}$ was > 1 once again.

Although the design-unbiasedness of ACS does not depend on the type of population being sampled, the relative efficiency of ACS does. By definition, the within-network variance component of $\text{Var}[\widehat{\mu}_{HH}]$ is

$$\sigma_W^2 = \left[\sum_{k=1}^{K}\sum_{i\in B_k}(y_i - w_{k(i)})^2\right]/(N-1), \tag{12}$$

where $k(i)$ is the kth network that includes unit i, B_k denotes the set of units that comprise the kth network, and $w_{k(i)}$ is the average of the y-values of the units in the kth network that includes unit i. Thompson (1990) showed that $\widehat{\mu}_{HH}$ from ACS using initial SRSWOR of samples of size n_1 will have lower variance than the SRS sample mean of size n if and only if:

$$\left(\frac{1}{n_1} - \frac{1}{n}\right)\sigma^2 < \frac{N - n_1}{Nn_1}\sigma_W^2. \tag{13}$$

Thus, (13) says that ACS using $\widehat{\mu}_{HH}$ will be more efficient than SRS if the within-network variance of the population is "high". This is similar to conventional cluster sampling, i.e., it is desirable to form clusters having y-values that are as variable as possible within each cluster while having the cluster totals as similar as possible.

In light of (13), we now have an explanation for the result that $\widehat{\mu}_{HH}$ is less efficient than $\overline{y}$ for all initial sample sizes for the presence/absence case when the population consisted of 0 or 1 y-values. For this population, the within-network variance $\sigma_W^2 = 0$ because every network would consist either of a single unit with $y_i = 0$ or of a group of one or more units each with $y_i = 1$. Thus, for this type of population, the ACS design using $\widehat{\mu}_{HH}$ cannot be more efficient than SRS using $\overline{y}$. Similarly, Christman (1996b) demonstrated that, using an initial SRSWOR, $\widehat{\mu}_{HT}$ under ACS cannot be more efficient than $\overline{y}$ when all the networks are of size 1 or when C is so restrictive that no unit satisfies the condition.

2.3 *ACS with an initial sample of primary units*

Two ACS designs utilizing both primary and secondary units were described by Thompson (1991a), each using modifications of the Horvitz–Thompson and Hansen–Hurwitz estimators. The first ACS design involves an initial SRSWOR of n_1 primary units where the primary units are clusters in the form of strips containing an

equal number of secondary units. All secondary units are sampled with additional adaptive sampling performed at the secondary level.

The second ACS design involves initial SRSWOR of n_1 primary units where the primary units are formed from k systematic selections. These primary units contain an equal number of noncontiguous equispaced secondary units, all of which are sampled. Further adaptive sampling is again done at the secondary unit level.

For these two designs, let N be the number of primary units in the population with each primary unit consisting of M secondary units. Unit (i, j) represents the jth secondary unit in the ith primary unit with associated y-value, y_{ij}. The modified Horvitz–Thompson ACS estimator is

$$\widehat{\mu}_{HT} = \frac{1}{MN} \sum_{k=1}^{K} \frac{y_k^* I_k}{\alpha_k}. \tag{14}$$

where $I_k = 1$ with probability α_k if at least one primary unit of the initial sample intersects the kth network, and $I_k = 0$ otherwise. Like (4), the marginal initial intersection probability is

$$\alpha_k = 1 - \left[\binom{N - x_k}{n_1} \Big/ \binom{N}{n_1} \right]. \tag{15}$$

In (15), however, x_k represents the number of *primary* units in the population that intersect the kth network. The probability that at least one *primary* unit of the initial sample intersects both networks j and k is

$$\alpha_{jk} = 1 - \left[\binom{N - x_j}{n_1} + \binom{N - x_k}{n_1} - \binom{N - x_j - x_k + x_{jk}}{n_1} \right] \Big/ \binom{N}{n_1}. \tag{16}$$

where x_{jk} is defined to be the number of primary units that intersect both networks, Thompson (1991a) showed that $\widehat{\mu}_{HT}$ is unbiased and

$$\mathrm{Var}[\widehat{\mu}_{HT}] = \frac{1}{M^2 N^2} \left[\sum_{j=1}^{K} \sum_{k=1}^{K} y_j^* y_k^* \left(\frac{\alpha_{jk} - \alpha_j \alpha_k}{\alpha_j \alpha_k} \right) \right] \tag{17}$$

with unbiased variance estimator

$$\widehat{\mathrm{Var}}[\widehat{\mu}_{HT}] = \frac{1}{M^2 N^2} \left[\sum_{j=1}^{K} \sum_{k=1}^{K} y_j^* y_k^* \left(\frac{\alpha_{jk} - \alpha_j \alpha_k}{\alpha_{jk} \alpha_j \alpha_k} \right) I_j I_k \right]. \tag{18}$$

For an initial systematic sample where $n_1 = 1$, $\alpha_{jk} = 0$ for some j and k, underscoring the fact that an unbiased estimator of the variance is not available for such a design (Thompson, 1991a). For ACS with an initial sample of primary units, the modified Hansen–Hurwitz estimator is:

$$\widehat{\mu}_{HH} = \frac{1}{Mn_1} \sum_{k=1}^{K} \frac{b_k y_k^*}{x_k} \tag{19}$$

where b_k is the number of times network k is intersected by the initial sample of primary units. Thompson (1991a) showed $\widehat{\mu}_{HH}$ is unbiased and

$$\text{Var}[\widehat{\mu}_{HH}] = \frac{\sigma_W^2}{n_1}\left(1 - \frac{n_1}{N}\right) \tag{20}$$

where $\sigma_W^2 = \frac{1}{N-1}\sum_{i=1}^{N}(w_i - \mu)^2$, $w_i = \frac{1}{M}\sum_{k=1}^{K}\frac{I_{ik}y_k^*}{x_k}$, and $I_{ik} = 1$ if the ith primary unit intersects the kth network, and $I_{ik} = 0$ otherwise. An unbiased estimator of $\text{Var}[\widehat{\mu}_{HH}]$ is:

$$\widehat{\text{Var}}[\widehat{\mu}_{HH}] = \frac{s_W^2}{n_1}\left(1 - \frac{n_1}{N}\right) \tag{21}$$

where $s_W^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(w_i - \widehat{\mu}_{HH})^2$.

2.4 *Comparisons to conventional cluster/systematic sampling*

Thompson (1991a) used a rare, clustered population produced via simulation to examine two primary unit structures. Whenever one or more point objects were encountered in the primary unit, further adaptive sampling was conducted at the secondary level. The first example used initial long, thin strip plots while the second example used a spatially patterned systematic initial sample with two randomized starting points. Once again for comparative purposes, $\text{Var}(\overline{y})$ was computed for SRSWOR of primary units with sample size $E(\nu)$, the expected effective primary unit sample size under ACS, instead of n_1. Comparisons were done for various initial sample sizes of primary units ranging from 1 up to a sampling fraction of .5.

The results from the first example were similar to those from the initial SRS results. For all initial sample sizes, both $\widehat{\mu}_{HT}$ and $\widehat{\mu}_{HH}$ were relatively more efficient than $\overline{y}$, and this disparity became more pronounced as n_1 increased. $\widehat{\mu}_{HT}$ was more efficient than $\widehat{\mu}_{HH}$ for all initial sample sizes $n_1 > 1$, and this disparity also became more pronounced as n_1 increased.

The second example yielded dramatic results. For initial systematic samples, variances were much lower than those obtained when the primary units in the initial sample were long, thin strips. Thompson (1991a) attributed this result to the spatial process used to simulate the population data. Adaptive strategies using $\widehat{\mu}_{HT}$ and $\widehat{\mu}_{HH}$ were far more efficient relative to SRSWOR for this population with the disparity became more pronounced with increasing n_1. In fact, efficiency $(\widehat{\mu}_{HT}) = 1.52$ for $n_1 = 1$ and efficiency $(\widehat{\mu}_{HT}) \rightarrow \infty$ as $n_1 \rightarrow 8$.

2.5 *Stratified ACS*

When a known source of variability can be exploited so that the population can be stratified by grouping relatively homogeneous secondary units into heterogeneous strata, a reduction in the variance of an estimator can be achieved. We will assume that SRSWOR of secondary units occurs in each stratum. Although one can then adaptively sample at the secondary level, a problem arises when a cluster contains secondary units that lie in more than one stratum. One approach is to truncate such clusters at the stratum boundaries yielding stratum estimators that are independent

and can be combined to provide an overall weighted estimator of μ. However, it will be shown that truncating clusters leads to a loss of efficiency. Thus, it is preferable to use an estimator that allows for clusters to overlap stratum boundaries (Thompson, 1991b).

Suppose that the population has been partitioned into H strata with N_h being the size of stratum h, n_h is the size of the simple random samples taken in stratum h, and $n_0 = \sum_{h=1}^{H} n_h$ is the total sample size. Thompson (1991b) proposed several estimators and derived a variety of results to be used in conjunction with stratified ACS. The first unbiased ACS estimator of μ is:

$$\widehat{\mu}_{HT} = \frac{1}{N}\sum_{k=1}^{K}\frac{y_k^* I_k}{\alpha_k} \tag{22}$$

where the K distinct networks are indexed *without* regard to stratum boundaries and $I_k = 1$ with probability α_k if the initial sample of size n_0 intersects network k, and $I_k = 0$ otherwise. The derivation of α_k is more complex than (4) because we now must consider probabilities of intersecting network k with initial samples in *each* stratum. If we define x_{hk} to be the number of units in stratum h that lie in network k,

$$\alpha_k = 1 - \left[\prod_{h=1}^{H}\binom{N_h - x_{hk}}{n_h}\Big/\binom{N_h}{n_h}\right] \tag{23}$$

and the joint initial intersection probability for networks j and k is

$$\alpha_{jk} = 1 - (1-\alpha_j) - (1-\alpha_k) + \left[\prod_{h=1}^{H}\binom{N_h - x_{hj} - x_{hk}}{n_h}\Big/\binom{N_h}{n_h}\right]. \tag{24}$$

$\mathrm{Var}[\widehat{\mu}_{HT}]$ and its unbiased estimator are the same as (6) and (7), respectively.

For the modified Hansen–Hurwitz estimators, three things need to be considered: (i) A_{hi}, the network containing the i^{th} unit in stratum h, (ii) A_{ghi}, the part of A_{hi} in stratum g, and (iii) m_{ghi}, the number of units in A_{ghi} (Thompson, 1991b). The first of three unbiased estimators of μ is

$$\widehat{\mu}_{HH} = \sum_{h=1}^{H}\frac{N_h}{N}\overline{w}_h \tag{25}$$

where $\overline{w}_h = \frac{\sum_{i=1}^{n_h} w_{hi}}{n_h}$, $w_{hi} = \frac{n_h}{N_h}\frac{Y_{hi}}{\sum_{g=1}^{H}\frac{n_g}{N_g}m_{ghi}}$, and Y_{hi} is the sum of the y-observations in A_{hi}. Its variance is

$$\mathrm{Var}[\widehat{\mu}_{HH}] = \frac{1}{N^2}\sum_{h=1}^{H} N_h(N_h - n_h)\frac{\sigma_h^2}{n_h} \tag{26}$$

where $\sigma_h^2 = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}(w_{hi} - \overline{W}_h)^2$ and $\overline{W}_h = \frac{\sum_{i=1}^{N_h} w_{hi}}{N_h}$. An unbiased estimator of $\mathrm{Var}[\widehat{\mu}_{HH}]$ is obtained by replacing σ_h^2 in (26) with $s_h^2 = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}(w_{hi} - \overline{w}_h)^2$.

A second estimator based on the stratified "multiplicity" estimator of network sampling is

$$\widehat{\mu}'_{HH} = \sum_{h=1}^{H} \frac{N_h}{N} \overline{w}'_h \tag{27}$$

where $\overline{w}'_h = \frac{\sum_{i=1}^{n_h} w'_{hi}}{n_h}$ and $w'_{hi} = \frac{y_{hi}}{\sum_{g=1}^{H} m_{ghi}}$. $\mathrm{Var}[\widehat{\mu}'_{HH}]$ and its unbiased estimator, $\widehat{\mathrm{Var}}[\widehat{\mu}'_{HH}]$, are calculated via (26) using w'_{hi} instead of w_{hi}. Thompson (1992) provides a variety of references on the history of this type of estimator.

The third estimator requires truncating clusters, i.e., it ignores any units added from crossing stratum boundaries:

$$\widehat{\mu}''_{HH} = \sum_{h=1}^{H} \frac{N_h}{N} \widehat{\mu}_h \tag{28}$$

where $\widehat{\mu}_h = \frac{\sum_{i=1}^{n_h} w''_{hi}}{n_h}$ and w''_{hi} is the total of the y-values in the intersection of stratum h with A_{hi} divided by the number of units in the intersection. $\mathrm{Var}[\widehat{\mu}''_{HH}]$ and its unbiased estimator, $\widehat{\mathrm{Var}}[\widehat{\mu}''_{HH}]$, are calculated via (26) using w''_{hi} instead of w_{hi}.

2.6 *Comparisons to conventional stratified random sampling*

Thompson (1991b) once again used a rare, clustered population produced via simulation to compare ACS and conventional sampling under stratification. The population was partitioned into two strata with one network lying on their boundary. A variety of sample sizes were examined while keeping the initial within-stratum simple random sample sizes (n_h) equal. A unit satisfied the condition C if $y_i \geq 1$. For comparative purposes, the variance was calculated for the conventional stratified sample mean having total sample size equal to $E[\nu]$, the expected effective sample size under ACS. The following relationship held for all n_0:

$$\mathrm{Var}[\overline{y}_{st}] > \mathrm{Var}[\widehat{\mu}''_{HH}] > \mathrm{Var}[\widehat{\mu}''_{HH}] = \mathrm{Var}[\widehat{\mu}_{HH}] > \mathrm{Var}[\widehat{\mu}_{HT}] \tag{29}$$

with the inequalities becoming more pronounced with increasing n_0. In this case, $\mathrm{Var}[\widehat{\mu}'_{HH}] = \mathrm{Var}[\widehat{\mu}_{HH}]$ because $\widehat{\mu}'_{HH}$ and $\widehat{\mu}_{HH}$ are identical when the stratum sizes and sample sizes are equal. Thompson (1991b) showed that the variances of $\widehat{\mu}_{HH}$, μ'_{HH}, and μ''_{HH} were fairly similar. This result was not surprising because a couple of the point object aggregations were far away from the stratum boundary. He also noted that substantial variance reduction can be achieved relative to SRS, with the reduction oftentimes being dramatic, when using ACS.

Using the same population, the estimates were compared when the initial stratum sample sizes were in a 2:1 ratio. The conclusions were identical to the equal initial sample size case except that $\mathrm{Var}[\widehat{\mu}'_{HH}] > \mathrm{Var}[\widehat{\mu}_{HH}]$. Finally, the population was partitioned into 16 strata where the initial samples were of equal size n_h and a range of n_0 values was considered. The conclusions were in agreement with (29). However, the gain in relative efficiency was larger than in the 2 strata case. Thompson (1994) speculated that this was because the initial stratified sampling design with many small strata is efficient for spatially aggregated populations and the efficiency improved even further with ACS.

3. Practical issues affecting the relative efficiency of ACS

In general, neither ACS nor conventional sampling designs are uniformly better than the other over all possible population configurations. Smith *et al*. 1995 warn that blind application of ACS may lead to a sampling design no better or worse than SRS in terms of precision. Christman (2000) also noted this in her research. We now take an in-depth look at a variety of factors affecting the efficiency of ACS relative to SRS.

3.1 *Choice of sampling design*

Choosing an appropriate sampling design can be complicated when the population is rare and clustered. In the review of quadrat-based sampling of rare and clustered populations, Christman (2000) presented a simulation study comparing different classes of sampling designs across three rare and clustered populations with varying degrees of within-network variance. Furthermore, different schemes for allocating the initial sample size were considered for both conventional and ACS stratified sampling designs. When the population was stratified such that one stratum contained all of the rare elements and the other contained none, the designs that had the lowest variances of the estimators of τ were those using optimal or disproportional allocation of stratum SRS sizes. If stratification was done arbitrarily, as is oftentimes the necessary case, the use of stratified ACS resulted in a dramatic reduction of variance relative to stratified SRSWOR, regardless of the method of stratum sample size allocation. However, lower variance was achieved along with a smaller expected sample size using a conventional systematic sampling design rather than one using stratified ACS.

When systematic ACS was conducted, the variance was even lower although the expected sample size reached its highest level. Consequently, the major result was that if sampling cost is not an issue and stratification based on location of the rare elements is not possible a priori, systematic ACS sampling may provide the best option from a variance efficiency standpoint.

3.2 *Within-network variability*

By partitioning the total sum of squares into between-network and within-network components, it can be shown that relative efficiency can also be expressed as:

$$\frac{\mathrm{Var}[\widehat{\mu}_{HH}]}{\mathrm{Var}[\overline{y}]} = \left[\frac{m}{n_1}\left(\frac{N-n_1}{N-m}\right)\right]\left[1 - \frac{\sum_{k=1}^{K}\sum_{i\in B_k}(y_i - w_{k(i)})^2}{\sum_{i=1}^{N}(y_i-\mu)^2}\right] \tag{30}$$

where m is the size of the SRS, B_k is the set of units that comprise the kth network, and $w_{k(i)}$ is the average of the y-values of the units in the network that includes unit i (Thompson and Seber, 1996, p. 151). It should be noted that the relative efficiency $\mathrm{Var}[\widehat{\mu}_{HT}]/\mathrm{Var}[\overline{y}]$, although straightforward to derive, involves more computation and does not lend itself to easy interpretation. The relative efficiency in general, however, can be improved through the use of $\widehat{\mu}_{HT}$.

The second term in the rightmost factor in (30) is the proportion of the population variance that is contained within networks. For given sample sizes m and n_1, ACS will be a more efficient design when the within-network variation is a high proportion of the total variation. Such a population would be described as *clustered* or as having aggregation tendencies. Thus, the relative efficiency depends on the particular partition of the population into K networks, which depends on the spatial distribution of the population, choice of C, the neighborhood structure, the quadrat size, and the size and the selection of the initial sample. Unfortunately, many of these factors are interrelated which complicates finding "the most efficient" design. (Smith *et al*. 1995; Brown 1996; Thompson and Seber 1996, Chap. 5; Christman 1996a, b). The results of Smith *et al*. (1995) suggest that the magnitude of the within-network variance relative to the total population variance may be the most important factor determining efficiency.

Christman (1996b, 1997) verified computationally that $\widehat{\mu}_{HH}$ performs well relative to $\bar{y}$ from an SRS when the within-network variance is a high proportion of the total variance and a first-order neighborhood is used. Similar results were noted for $\widehat{\mu}_{HT}$.

3.3 *Number and size of networks*

In a simulation study, Brown (2003) utilized a variant of a Poisson cluster process to generate 200 populations with varying rates λ_1 for the number of clusters ($\lambda_1 = 5$, 10, ... , 100) and λ_2 for the number of individuals per cluster ($\lambda_1 = 10$, 20). Individuals were exponentially distributed about the cluster centers at varying distances. As the population total τ increases, the number of networks increases, reaches a maximum, and then decreases. This occurs because of the hypernetwork phenomenon. That is, at some point, for increasing λ_1, adjacent networks begin to merge to form larger and, hence, fewer population networks. In addition to the spatial structure of the population, the actual number of networks also depends on the critical value (see Section 3.4) and on the neighborhood definition (see Section 3.5).

Brown (2003) showed that as the number of networks increases, the relative efficiency of the ACS Hansen–Hurwitz estimator decreased to below 1. The greatest reduction in relative efficiency occurred when there was a rapid increase in the number of networks (which, implicitly, depends on the population's spatial pattern, the neighborhood definition, and the critical value). For populations with compact networks with many individuals, the maximum average number of networks occurred for larger τ than for populations with less compact networks with few individuals. Thus, subsequent changes in the number and size of networks also affected relative efficiency in that for populations with compact networks with many individuals, the relative efficiency was always higher for all τ. Brown raises the issue that sample size and within-network variance can be conflicting factors affecting relative efficiency. Moreover, she claims that if the networks are large with respect to the number of units, then the ratio of the final to initial sample size will also be large, thereby decreasing relative efficiency. Large networks, however, can also have large within-network variances relative to the population variance whose effect should be to increase relative efficiency. Brown concluded that populations with compact

networks with more individuals have networks with large within-network variance σ_w^2 relative to σ^2 and ν is not much larger than n. In the study, the highest relative efficiency was attained for networks of around 2–4 units.

3.4 *Critical value*

Suppose additional units are included with an initial sample when condition $C : y_i \geq c_a$ is satisfied, where c_a is the *critical value* for condition C. Christman (1996b, 1997) demonstrated that $\widehat{\mu}_{HT}$ is an efficient estimator relative to the SRS estimator when $c_a = 1$ for an array of populations. For first-order neighborhoods, the efficiency performance of $\widehat{\mu}_{HT}$ became worse when $c_a = 5$. Christman (1996b) showed that $\widehat{\mu}_{HH}$ performs well relative to $\overline{y}$ from an SRS when the within-network variance is a high proportion of the total variance and there exists a first-order neighborhood. This result was less pronounced, however, with $c_a = 5$ than with $c_a = 1$. It was speculated that this was in part due to a decrease in within-network variability imposed by the more restrictive condition.

Brown (1996) reported an improvement in relative efficiencies for ACS for high density populations with $c_a = 2$ over $c_a = 1$ because increasing c_a had the effect of reducing the size of networks and thus reducing the number of edge units. However, for low density populations, relative efficiencies for ACS became worse because few networks were formed and few units were selected adaptively.

Christman (2000) also noted that the efficiency of ACS relative to SRS with an equivalent sample size is inversely related to the choice of c_a. Moreover, as c_a increases, the within-network variance decreases. Hence, choosing a relatively high critical value to control the overall sample size has the undesired effect of reducing the efficiency of ACS. Furthermore, edge units will likely contain nonzero counts but will not be in the estimation procedure, which leads to a loss of information. Smith *et al.* (2003) indicate that they do not believe that increasing the value of c_a is a practical method to control the overall sample size because of the unpredictable consequences on efficiency.

A small critical value can lead to the formation of larger (but a smaller number of) clusters while a large critical value can lead to the formation of smaller (but a larger number of) clusters. The choice of the critical value will therefore depend on whether sampling effort should be concentrated on sampling larger individual clusters or on sampling many smaller clusters which ultimately depends on where the largest source of variation is, within or between clusters (Brown, 1994).

Brown (2003) considered two critical values ($c_a = 1, 2$) in a simulation study. For the populations considered (see Section 3.3), the ACS Hansen–Hurwitz estimator of τ using either critical value was more efficient than the SRS estimator when the population total τ was small, but had a relative efficiency < 1 for larger τ. For smaller τ (smaller number of networks), the ACS Hansen–Hurwitz estimator using $c_a = 1$ was more efficient than when $c_a = 2$ was used, and it was less efficient for larger τ (greater number of networks). Thus, an interaction between the critical value and the number of networks is indicated.

3.5 *Neighborhood definition*

Altering the definition of the neighborhood (like the selection of critical value) can change the focus of sampling to either within or between clusters (Brown, 1994). Christman (1996b) examined ACS for several populations using three types of neighborhoods. The result was that the most efficient ACS designs were those that utilized neighborhoods based on units that were physically contiguous. In fact, Christman (1996b, 1997) proposed a noncontiguous, "four-away" neighborhood that was constructed such that if unit(i, j) was selected and $y_{ij} \in C$, then units lying two over to the north, south, east and west were sampled. For this type of neighborhood, both $\widehat{\mu}_{HH}$ and $\widehat{\mu}_{HT}$ had a higher variance than $\overline{y}$ from SRSWOR with sample size $E[\nu]$ for almost every initial sample size and population simulated.

There is also some evidence suggesting that a truncated first-order neighborhood, where only two adjacent units are sampled, may be as or more efficient than the usual first-order neighborhood. This result might not be surprising because, in general, a neighborhood that samples only 2 neighbors rather than 4 neighbors lowers the effective sample size. Also, for this reason, the truncated first-order neighborhood might be an attractive consideration when cost is a concern Christman (1996b, 1997).

If clusters tend to form in a specific direction, then the neighborhood definition should reflect the expected orientation (Christman, 2000).

Improved relative efficiencies for ACS were achieved by Brown (1996) when a smaller first-order neighborhood was used instead of a second-order neighborhood. The proposed reason this occurred was that relatively more units were actually used in the estimator $\widehat{\mu}_{HT}$ when there was a smaller neighborhood definition because the number of edge units was reduced.

In a later simulation study (see Section 3.3), Brown (2003) found that the actual number of networks depends on the neighborhood definition. Brown (2003) considered three neighborhood definitions: 8-unit second-order, 4-unit first-order, and 2-unit first-order. Specifically, the number of networks increases as the size of the neighborhood definition decreases for all τ. The results also indicate an interaction, i.e., the use of small neighborhood definitions improved the relative efficiency for larger values of the population total τ (greater number of networks) while the relative efficiency was smaller for very low values of τ (smaller number of networks).

3.6 *Estimator type*

Because $\widehat{\mu}_{HT}$ is a more efficient estimator than $\widehat{\mu}_{HH}$ under a variety of ACS designs, $\widehat{\mu}_{HT}$ should also be more efficient than $\widehat{\mu}_{HH}$ relative to $\overline{y}$ from an SRSWOR of size $E[\nu]$. Furthermore, greater efficiency in ACS may be achieved through use of the Rao–Blackwell improvement of either $\widehat{\mu}_{HH}$ or $\widehat{\mu}_{HT}$ (see Section 5).

Christman (1996b, 1997) compared the variances and efficiencies of $\widehat{\mu}_{HT}$ and $\widehat{\mu}_{HH}$ in a computer simulation study. The results suggested that although $\widehat{\mu}_{HT}$ had the lowest variance for most of the simulation situations explored, it was much more

sensitive to changes in sampling conditions and population structure. Thus, the sampling design should be carefully planned when using $\widehat{\mu}_{HT}$ and potential modification of the sampling effort anticipated. Unfortunately, the efficiency of $\widehat{\mu}_{HT}$ often comes at the cost of a significantly large effective sample size.

Despite the fact that the variances of Horvitz-Thompson ACS estimators have been smaller in applications, the Hansen–Hurwitz estimators have been used more often because of the ease of calculation. Salehi (2003) compared the properties of the two estimators. In practice, it is expected that larger network totals (y_k^* values) will be associated with larger initial intersection probabilities (α_k values) which Salehi refers to as the proportionality property. If this condition is true, then Salehi showed that we can expect $\widehat{\mu}_{HT}$ to perform better than $\widehat{\mu}_{HH}$. Also, if the y-values of the units satisfying C are relatively larger than the y-values of the units not satisfying C, then it was also shown that $\mathrm{Var}[\widehat{\mu}_{HH}]$ is much greater than $\mathrm{Var}[\widehat{\mu}_{HT}]$ for larger initial samples. Felix-Medina (2003) recommends limiting use of $\widehat{\mu}_{HT}$ to situations for which the network totals and intersection probabilities are correlated. Otherwise, $\widehat{\mu}_{HH}$ should be considered.

3.7 *Unit size*

To some extent, the choice of unit size determines the size of networks. As with the choice of a critical value, if the surveyor had prior knowledge of the population structure, then the efficiency of ACS using $\widehat{\mu}_{HH}$ can be increased by selecting the unit size accordingly Christman (1996b, 1997). These results were corroborated by the simulation study of Smith *et al*. (1995) that compared the efficiency of several designs using Horvitz–Thompson estimation to estimate τ for several species of wintering waterfowl. Their results varied according to species, C, and unit size with some combinations leading to increased ACS precision compared to SRSWOR and some not. The size of the unit affected the population structure to the extent that size of the unit affects the number and size of networks and the within-network variability.

3.8 *Geographic rarity and clustering*

For two populations, suppose the total τ and the number of clusters are the same. Then one population has more *geographic rarity* if its clusters are scattered over a larger study region than the other. If this is the case, most of the units will contain no point objects ($y_i = 0$) while only a few units will contain point objects ($y_i = 0$). For a population of fixed size and spatial aggregation pattern:

$$\lim_{N\to\infty} E[v] = n_1 \tag{31}$$

In other words, the expected effective sample size approaches the initial sample size with increasing geographic rarity (Thompson and Seber, 1996, p. 156). This was demonstrated in simulation studies done by Brown (1994) and Brown and Manly (1998). Because of this result, the right side of (34) approaches an uncorrected version of the rightmost factor in (30) as geographic rarity increases (Thompson and

Seber, 1996, pp. 156–157). Thus, if there is any within-network variation whatsoever, many aggregated populations, if geographically rare, will be more efficiently sampled with ACS than with conventional designs. For Horvitz-Thompson estimation, Smith *et al*. (1995) showed that the relative efficiency of ACS was highest when $E[\nu]$ was "close" to n_1. They determined that $E[\nu]$ depended on the same factors that were previously mentioned to have affected within-network variance. Smith *et al*. (1995) also noted that although the uncertainty of the final sample size is a disadvantage of ACS, a small difference between the final and initial sample sizes is characteristic of an efficient design. However, while agreeing that the claim was valid for Hansen–Hurwitz estimation, it has been shown that this characteristic does not necessarily hold for $\widehat{\mu}_{HT}$ (Christman, 1996b).

For Horvitz–Thompson estimation, in a simulation study using a fixed C and n_1, Christman (1996a) found that only those populations where the number of clusters was small would yield relative efficiencies favorable for ACS. As the number of clusters (and consequently, $E[\nu]$) was increased, the distributions of relative efficiencies became very skewed with the majority of simulated populations yielding relative efficiencies favorable for SRSWOR. Occasional relative efficiencies favorable for ACS, were attributed to high within-network variability.

A population is considered to be *clustered* or *patchy* if those units that do contain elements are spatially contiguous. Brown (1996), working with Horvitz–Thompson estimation, stated that ACS "is not a suitable technique for sampling populations that do not have a very patchy spatial pattern". To discern differences in patchiness, indices of aggregation were used to discriminate among the populations according to the appropriateness of using ACS and to construct classification functions for applied use.

DiBattista (2003) considered the situation that the researcher is interested in using dispersion indices to evaluate the structure of a clustered population. Examples include indices that represent clumping, cluster frequency, and patchiness. All of the dispersion indices considered are smooth functions of the first two moments of a point-count variable. The quality of inference regarding these indices will therefore depend on the quality of estimation of μ and σ^2. Although the Hansen–Hurwitz estimators of μ and σ^2 are unbiased, the plug-in estimators for the dispersion indices are not. To address this problem, DiBattista (2003) proposes the use of resampling methods to reduce the bias. In a simulation study of the Fisher dispersion index (σ^2/μ) for one artificial population, DiBattista concluded that (i) the relative bias and root mean squared error (RMSE) are better using ACS than SRS, (ii) ACS estimators using resampling methods had a smaller RMSE than the standard ACS estimators, and (iii) the jackknife procedure is recommended for bias reduction while the bootstrap procedure is recommended for increased estimator precision and confidence interval construction.

3.9 *Sample size and cost*

From (30), it is clear that ACS will be more efficient than SRS when

$$b(n_1, m, N) = \frac{m}{n_1}\left(\frac{N - n_1}{N - m}\right) \tag{32}$$

is small, and the efficiency of ACS improves as b decreases (and vice versa). For fixed n_1, b is an increasing function of m. For most comparisons of interest, $m \geq n_1$. ACS will be more efficient than SRS if m is not much bigger than n_1. If $m \leq n_1$, ACS will always be more efficient than SRS. If m and n_1 increase at a constant ratio, that is, $m = a_1 k$ and $n_1 = a_2 k$, then b is an increasing function of k. Finally, with n_1 and m fixed, b is a decreasing function of N if $m > n_1$, and is increasing if $m < n_1$. Let $g(n_1)$ be $b(n_1, E[\nu], N)$. Then, $g(n_1)$ may be either increasing or decreasing in n_1 (Thompson and Seber, 1996, pp. 154–155). As a result, $g(n_1)$, and ultimately the upper bound on the relative efficiency, are similarly influenced by the choice of n_1 (as well as the population structure).

For Horvitz–Thompson estimation, Brown (1996) showed a general improvement in relative efficiencies for ACS when n_1 was increased, particularly in high density populations. It was speculated that increasing n_1 would also increase the expected number of selected networks. This would then have the effect of increasing the proportion of within-network variation. Brown (2003) also demonstrated an improvement in relative efficiency for Hansen–Hurwitz estimation for larger n_1.

Christman (1996b, 1997) found that in order for $\widehat{\mu}_{HT}$ to be an efficient estimator for populations with "large" numbers of clusters, a larger n_1 is needed. Unfortunately, as a result, there can be a dramatic increase in $E[\nu]$. Across an array of populations and design strategies, the efficiency of $\widehat{\mu}_{HH}$ tended to improve with increasing n_1 but was less efficient than $\overline{y}$ under SRSWOR with sample size $E[\nu]$.

Directly related to sample size is the cost of sampling. Thompson (1994) proposed the following cost function for ACS:

$$c_T = c_0 + (c_1 - c_2)n_1 + c_2 E[\nu] \tag{33}$$

where c_T is the total cost, c_0 is a fixed cost, c_1 is the marginal cost for each unit obtained via SRS, c_2 is the marginal cost for each adaptively added unit, and $E[\nu]$ is the expected final sample size when the initial sample size is n_1. In some studies sampling an edge unit may take less time, and therefore be less costly than sampling a network unit. The cost function would then include separate components for both unit types (Thompson, 1994; Brown, 1996).

When $m_c = \left(1 - \frac{c_2}{c_1}\right) n_1 + \frac{c_2}{c_1} E[\nu]$ is the sample size of a conventional SRS design, its total cost $c_T = c_0 + c_1 m_c$ is equal to the total cost of ACS in (33). Note that $n_1 \leq m_c \leq E[\nu]$. In general, $c_2 < c_1$ because there is less cost associated with sampling units within clusters due to distance and time considerations. Intuitively, this is not an unreasonable assumption. For example, Rocco (2003) noted the existence of such an assumption in her work. Brown (1996) showed in a simulation study that the average distance traveled for ACS was less than the average distance for SRSWOR across an array of population models. Because the variances under SRS decrease as m increases, we have for a fixed expected cost:

$$\frac{\mathrm{Var}[\widehat{\mu}_{HH}; n_1]}{\mathrm{Var}[\overline{y}; n_1]} \leq \frac{\mathrm{Var}[\widehat{\mu}_{HH}; n_1]}{\mathrm{Var}[\overline{y}; m_c]} \leq \frac{\mathrm{Var}[\widehat{\mu}_{HH}; n_1]}{\mathrm{Var}[\overline{y}; E[\nu]]} \tag{34}$$

(Thompson and Seber, 1996, pp. 110–111) showed that the left side of (34) is always ≤ 1. The right side of (34) can serve as a conservative assessment of relative efficiency.

In summary, we have seen that ACS would be least efficient relative to SRS when a population's units that satisfy C are not contiguous, i.e., the within-network variance is low. Thus, ACS would add only edge units which increase the sample size but do not contribute to the precision of estimators (Thompson, 1994). Much attention has been given to devising ways to limit the (random) size of the final sample in ACS. This can be particularly important from a cost standpoint. Although the final sample size can be controlled to a limited extent by the appropriate selection of design factors, prior knowledge of the population structure is required. Sometimes this knowledge is not available making prior planning of resources necessary.

On occasion, when the critical value is set too low, continued sampling would lead to the addition of more units than time or cost allow. If sampling were curtailed prematurely, biases may arise. (One possible remedy, the use of order statistics, will be discussed in Section 8.) Thompson (1994) offered the following guidelines when the need arises to limit sampling effort.

The study region can be stratified a priori. Then the decision of whether or not to use ACS, as well as the choice of C, for any subsequent stratum should be based on the results for the strata already surveyed. Depending on the design used, design-unbiased estimators of τ_h and its variance can be calculated for each stratum. The stratified design-unbiased estimate of τ will be the sum of the individual τ_h with an analogous result holding for the variance assuming independence between strata.

Poststratification could also be used (although Christman (2000) discouraged this practice). For example, one stratum could represent the part of the population where the conventional initial sample was completed, while a second stratum could represent the part of the population where enough resources existed such that ACS was able to be conducted. Perhaps C might have been modified several times during the course of the study as more results became available. Then, the region would be poststratified into strata based on the modification of C.

The choice of strata sample sizes is itself a substantive and important topic. The reader is encouraged to consult Christman (2000) for an overview and (Thompson and Seber, 1996, Chap. 5) for a detailed discussion of adaptive allocation strategies in stratified sampling.

Another guideline would be to modify the neighborhood definition. Neighborhoods could be truncated for a pre-defined region or block boundary to limit the size of networks. Also, defining the neighborhood to be a noncontiguous systematic grid will constrain networks to be spatially sparse. A final guideline would be to use a restricted ACS design or inverse sampling. Section 6 is devoted to this topic.

3.10 *Further developments*

Because a careful and thoughtful preliminary investigation of the population structure should be conducted before utilizing ACS, research is needed for developing guidelines in setting the levels of the factors that influence ACS surveys. It has been suggested that this goal may be accomplished via a pilot study (Smith *et al.*, 1995; Brown, 1996). For example, in their study of freshwater mussels, Smith *et al.* (2003) did use a pilot study to provide an estimate of the population variance. An

approach suggested by Brown (1996) would be to use a small unit size, a small critical value, and a large neighborhood definition in a pilot survey. The data could then be reanalyzed using various rescaled levels of these factors to help define an appropriate design for the population.

Another approach suggested by Brown (1996) would be to sample the population in phases. At the initial phase, educated guesses are made for the levels of the sample design factors, and then an initial area is surveyed. The results from the initial phase are used to assess whether or not the levels of the design factors need to be changed. Sampling would then continue in the next phase with the revised sample design. Again, the sample design would be re-evaluated and, if necessary, revised again for the subsequent phase. Separate estimates from each phase would be combined providing a single estimate for the entire area.

Christman (1997) also suggested adding a sequential component to the sample design. Statistics that describe the aggregation and rarity of the clusters would be calculated from the sample data and then would be used to modify the subsequent sampling effort.

4. Initial unequal probability sampling

There are some circumstances in designing a survey where each unit will not have the same probability of inclusion in the initial sample. Perhaps quadrats are of unequal size, or the usual rectangular grid of secondary units is not adequate to cover the study site in an ecological study (Pontius, 1996, 1997). Roesch (1993) used PPS, or probability proportional to size selection, with replacement to select given points at random in the population study area in a simulation of ACS in a forest inventory. The auxiliary variable was the basal area of trees. He found this strategy to be reasonable when sampling for a rare characteristic that tended to be clustered in groups of trees. Both the theory and methodology of the Roesch paper are well described in (Thompson and Seber, 1996, pp. 101–108).

Smith *et al.* (1995) used PPS in their simulation study for estimating the number of several species of wintering waterfowl. They compared four designs: ACS with initial SRSWOR, ACS with PPS with replacement, conventional SRSWOR, and conventional PPS with replacement. For ACS with PPS, the Horvitz–Thompson estimator was used with modified initial intersection probabilities that were based on the area of available habitat in a sampling unit. The results indicated that ACS with PPS was more efficient than ACS with SRSWOR for certain conditions (unit size and C) when the auxiliary variable habitat was appropriately defined.

ACS with an initial sample of primary units has been extended to situations where the primary units have unequal numbers of secondary units and consequently falls under PPS (Pontius, 1996, 1997). This research centered on selection of primary units with replacement in conjunction with the Horvitz–Thompson estimator. Two cases were considered: the strip ACS design (see Section 2.3) and strip ACS under stratification of primary units. In the latter case, selection of primary units was done independently across strata. Modifications of the marginal and joint initial intersection probabilities were derived for both cases and functions to calculate them in S+ were presented (Pontius, 1996).

Felix-Medina (2003) studied the asymptotic properties of $\widehat{\mu}_{HT}$ and $\widehat{\mu}_{HH}$ under initial unequal probability sampling with replacement. Based on five assumptions applied to a sequence of populations $\{U_t\}_{t=1}^{\infty}$ and samples that were analogous to the SRS without replacement cases discussed in Section 2.1, Felix-Medina proved that $\widehat{\mu}_{HT}$ and $\widehat{\mu}_{HH}$ are design-consistent and asymptotically normally distributed. The same concerns about quality of ACS estimation in the finite-sample setting when an initial SRS is used also apply when an initial unequal probability sampling with replacement is taken.

5. Rao–Blackwell estimators

An interesting development in ACS has been the application of the Rao–Blackwell method to obtain smaller variance design-unbiased estimators for ACS. To see how this is done we need the following definitions and notation.

A selection-ordered sample of size n is a sequence $s_0 = (i_1, \ldots, i_n)$ of the n sampled-unit labels. We define d_0 to be $d_0 = (s_0, \mathbf{y}_0)$ where $\mathbf{y}_0 = (y_i : i \in s_0)$ is the set of selection-ordered sample y-values. Let $s_R = \{i_1, \ldots, i_v\}$ be the unordered *reduced set* of ν distinct labels in s_0 and $\mathbf{y}_R$ be the corresponding set of y-values in s_R. By definition, $d_R = (s_R, \mathbf{y}_R)$. The *reduction* from d_0 to d_R is achieved by ignoring information about order and the multiplicities of the units in the sample. Let D_R be the random variable or set corresponding to d_R. For the ACS designs considered here, D_R from the final sample is a minimally sufficient statistic for $\boldsymbol{\theta}$ (Thompson and Seber, 1996, p. 38)

In addition to the Rao–Blackwell estimators presented in Thompson and Seber (1996), several other Rao–Blackwell estimators have been proposed over recent years (Salehi and Seber 1997a; Dryver and Thompson, 1998; Salehi 1999; Felix-Medina 2000). Unfortunately, none of these estimators are uniformly better than the others with respect to their variances. For all of the ACS designs considered in this review, a uniform minimum variance unbiased estimator (UMVUE) of parameter $\phi(\boldsymbol{\theta})$ does not exist. In fact, if we consider either adaptive or conventional sampling designs in the fixed-population setting, Lehmann (1983) showed that a UMVUE does not exist because it would have to have zero variance with this being due to the fact that D_R is not complete. The incompleteness of D_R in the finite population sampling situation can be attributed to the presence of the unit labels (Thompson and Seber 1996, pp. 45–46).

The three unbiased estimators $\overline{y}_1$, $\widehat{\mu}_{HT}$, and $\widehat{\mu}_{HH}$ applied to an adaptive cluster sample depend on the order of selection. Additionally, $\widehat{\mu}_{HH}$ depends on repeat selections and, if the initial sample was selected WR, then all three estimators would depend on repeat selections. $\widehat{\mu}_{HT}$ depends on repeat selections to the extent that the initial intersection probability α_k in (4) is modified for WR sampling. For an initial SRSWOR, the number of possible combinations of n_1 distinct units from the ν in the sample is $G = \binom{\nu}{n_1}$. Because all three estimators are not functions of D_R, the Rao–Blackwell theorem can be applied to each. (Thompson and Seber, 1996, pp. 108–109) presented the Rao–Blackwell estimator:

$$T_{RB} = \frac{1}{\xi} \sum_{g=1}^{G} t_g I_g. \tag{35}$$

where t_g denotes the value of any one of the three estimators when the initial sample consists of combination g, $I_g = 1$ when the gth combination is *compatible* with d_R (i.e., could give rise to d_R), and $I_g = 0$ otherwise. Thus, $\xi = \sum_{g=1}^{G} I_g$ is the number of compatible combinations. Analogously, for SRSWR, the summation in (35) will be over the $G = v^{n_1}$ possible order-preserving permutations of n_1 distinct units from the v in the sample. Unlike $\widehat{\mu}_{HH}$ and $\widehat{\mu}_{HT}$ which utilize observations not satisfying C only when they are selected as part of the initial sample, Rao–Blackwell estimators, by construction, *do* utilize edge units even if they are not selected in the initial sample. (Note: although the focus is on estimation of the mean μ, the Rao–Blackwell method can also be applied to estimation of the variance.)

If T is any one of the three estimators:

$$\mathrm{Var}[T_{RB}] = \mathrm{Var}[T] - E[\mathrm{Var}[T|d_R]]. \tag{36}$$

Because $\mathrm{Var}[T]$ is a function of θ and $\widehat{\mathrm{Var}}[T]$ is an unbiased estimator that is not a function of D_R, the Rao–Blackwell theorem yields an unbiased variance estimator with smaller variance:

$$\widehat{\mathrm{Var}}_{RB}[T] = E[\widehat{\mathrm{Var}}[T]|d_R] = \frac{1}{\xi}\sum_{g=1}^{G} \widehat{\mathrm{Var}}_g[T] I_g. \tag{37}$$

Furthermore,

$$\mathrm{Var}[T|d_R] = E[(T - T_{RB})^2|d_R] = \frac{1}{\xi}\sum_{g=1}^{G} (t_g - T_{RB})^2 I_g. \tag{38}$$

Consequently, for any one of the three estimators,

$$\widehat{\mathrm{Var}}[T_{RB}] = \frac{1}{\xi}\sum_{g=1}^{G} [\widehat{\mathrm{Var}}_g[T] - (t_g - T_{RB})^2] I_g. \tag{39}$$

Unfortunately, it is possible for $\widehat{\mathrm{Var}}[T_{RB}] < 0$.

Let $D_f = \{(i, y_i, f_i) : i \in s_R\}$. That is, D_f is D_R along with f_i, the frequency of intersection of the initial sample with the network A_i. For the modified estimator $\widehat{\mu}_{HH}$ and the initial sample mean $\overline{y}_1$, (Thompson and Seber, 1996, pp. 110–111) proved the following:

Theorem 1 Suppose sampling is WR or WOR. Then:

1. D_f is sufficient for $\boldsymbol{\theta}$.
2. $E[\overline{y}_1|D_f = d_f] = \widehat{\mu}_{HH}$.
3. $E[\overline{y}_1|D_R = d_R] = E[\widehat{\mu}_{HH}|D_R = d_R] = \widehat{\mu}_{RBHH}$.

Theorem 1 indicates that the resulting estimator $\widehat{\mu}_{RBHH}$ is the same when Rao–Blackwell theory is applied to both $\overline{y}_1$ and $\widehat{\mu}_{HH}$. Based on statements 1. and 2. of Theorem 1, $\mathrm{Var}[\widehat{\mu}_{HH}] \leq \mathrm{Var}[\overline{y}_1]$ when both have sample size n_1. In the case of $\widehat{\mu}_{HH}$, n_1 is the initial sample size.

Thompson (1990) calculated $\widehat{\mu}_{RBHH}$ and $\widehat{\mu}_{RBHT}$ for all possible samples given an initial sample of fixed size n_1 was selected via SRSWOR from a small population. $\widehat{\mu}_{RBHT}$ had the lowest variance among the five unbiased estimators ($\overline{y}_1$, $\widehat{\mu}_{HT}$, $\widehat{\mu}_{RBHT}$, $\widehat{\mu}_{HH}$, and $\widehat{\mu}_{RBHH}$).

Although the Rao–Blackwell method can also be used in stratified ACS, it is complicated by two facts discussed in Section 2.5: (i) there will be substantially more computation involved to determine all initial compatible stratified samples and (ii) there are four distinct Rao–Blackwell estimators, including $\widehat{\mu}_{RBHT}$, and one for each of the three different modified Hansen–Hurwitz estimators ($\widehat{\mu}_{HH}$, $\widehat{\mu}'_{HH}$, and $\widehat{\mu}''_{HH}$) that exist for stratified ACS. Extending the use of Theorem 1, it can be shown that the Rao–Blackwell version of the initial weighted stratified sample mean $\overline{y}_0 = \sum_{h=1}^{H} \frac{N_h}{N} \overline{y}_h$ is the same as $\widehat{\mu}''_{RBHH}$ (Thompson and Seber, 1996, p. 111). Thompson (1991b) presented an example where initial samples of fixed and equal size for each stratum ($n_1 = n_2$) were selected from a small population with unequal stratum sizes. All possible samples were generated. When the four Rao–Blackwell estimates corresponding to $\widehat{\mu}_{HT}$, $\widehat{\mu}_{HH}$, $\widehat{\mu}'_{HH}$, and $\widehat{\mu}''_{HH}$ were obtained for each sample, $\widehat{\mu}_{RBHT}$ had the lowest variance, and each of these estimators was improved by the Rao–Blackwell method.

In the case where ACS is conducted using an initial sample of primary units, the Rao–Blackwell method can still be used but is unlikely to result in any improvement. The reason is that it is unlikely that more than one initial selection of primary units will lead to the same d_R (Seber and Thompson, 1994). Consequently, $\xi = 1$, $T_{RB} = t_g$, and $\mathrm{Var}[T|d_R] = 0$.

Given the complexity of computation, computational aspects of the Rao–Blackwell estimators deserve further study. Because the summations in (35) and (39) depend on the samples compatible with d_R, the number of estimates (ξ) to evaluate is potentially huge. In an effort to circumvent this problem, Dryver and Thompson (1998) introduced Rao–Blackwell estimators utilizing *sample edge units*. A sample edge unit is an edge unit whose network that classifies it as an edge unit is intersected in the initial sample, even if the sample edge unit itself is in the initial sample. To obtain these estimators, they conditioned $\widehat{\mu}_{HH}$ and $\widehat{\mu}_{HT}$ on

$$D^+ = \{(i, y_i, f_i) : i \in s_c, (j, y_j) : j \in s_{\overline{c}}\}. \tag{40}$$

where s_c is the set of all the distinct units in the final sample satisfying C, and $s_{\overline{c}}$ is the set of all the distinct units in the final sample <u>not</u> satisfying C. D^+ is a sufficient but not minimal statistic. Two new estimators were developed by considering the average of the sample edge units in the final sample. The first estimator is

$$\widehat{\mu}_{RBHH+} = \frac{1}{n} \sum_{i=1}^{n} w'_i \tag{41}$$

where $w'_i = w_i(1 - e_i) + \overline{y}_e e_i$, w_i is the average value of a unit in the network which contains unit i, $\overline{y}_e$ is the average y-value for the sample edge units in the final sample, and $e_i = 1$ if unit i is a sample edge unit and $e_i = 0$ otherwise. The second estimator is:

$$\widehat{\mu}_{RBHT+} = \frac{1}{N} \sum_{k=1}^{K} \frac{y'_k I_k}{\alpha_k} \tag{42}$$

where y'_k is either (i) the sum of the y-values in network k if the network satisfies C or (ii) $\overline{y}_e$ for a sample edge unit, and $I_k = 1$ with probability α_k if the initial sample intersects the k^{th} network, and $I_k = 0$ otherwise . Derivation of variances, unbiased variance estimators, and unbiased Rao–Blackwell variance estimators for $\widehat{\mu}_{RBHH+}$ and $\widehat{\mu}_{RBHT+}$ are given in Dryver and Thompson (1998). Because we are conditioning on a statistic that is sufficient but not minimal, we have:

$$\mathrm{Var}[\widehat{\mu}_{RB.}] \leq \mathrm{Var}[\widehat{\mu}_{RB.+}] \leq \mathrm{Var}[\widehat{\mu}_{.}]. \tag{43}$$

where subscript '.' signifies either HT or HH. Salehi (1999) introduced closed forms for both $\widehat{\mu}_{RBHH}$ and $\widehat{\mu}_{RBHT}$ along with unbiased estimators of their variances which can be more readily computed. He also revealed the following interesting results. It is frequently the case for C to be $y_i \geq 1$. The author showed that if the y-values for the sample edge units are zero, then $\widehat{\mu}_{HT}$ cannot be improved using Rao–Blackwell theory, and $\widehat{\mu}_{RBHT} = \widehat{\mu}_{HT}$. On the other hand, $\widehat{\mu}_{HH}$ can be improved using $\widehat{\mu}_{RBHH}$ when the y-values of sample edge units are zero. He also showed that applying Rao–Blackwell theory to $\widehat{\mu}_{HH}$ will result in a greater relative improvement (variance reduction) than for the case of applying the theory to $\widehat{\mu}_{HT}$.

Salehi illustrated his computations, notation, and conclusions via a small population example and an analysis of a set of data from Smith *et al.* (1995). For the latter, he deliberately selected an initial sample demanding the most computation. In both examples, $\widehat{\mu}_{RBHH}$ was more efficient than $\widehat{\mu}_{RBHT}$. Even though these results suggest otherwise, (Thompson and Seber, 1996, p. 111) caution us that neither $\widehat{\mu}_{RBHH}$ nor $\widehat{\mu}_{RBHT}$ is uniformly better than the other. Salehi (2003), however, derived a Rao–Blackwellized estimator $\tilde{\mu}_{HH}$ that ignores information from the edge units, but has uniformly smaller variance than $\widehat{\mu}_{HH}$. Applying the same Rao–Blackwell conditioning to $\widehat{\mu}_{HT}$ yielded $\widehat{\mu}_{HT}$, i.e., no improvement was realized. Salehi, therefore, initially suggested that a possible competitor for $\widehat{\mu}_{HT}$ is $\tilde{\mu}_{HH}$. When studying a small population and then a population in Thompson (1990), he found that $\widehat{\mu}_{HT}$ was more efficient than $\tilde{\mu}_{HH}$ in both cases. Because of these results, Salehi (2003) strongly encourages the use of $\widehat{\mu}_{HT}$ over $\widehat{\mu}_{HH}$ (although he acknowledges the need for a comprehensive study comparing the two estimators $\widehat{\mu}_{HT}$ and $\tilde{\mu}_{HH}$).

Felix-Medina (2000) addressed the problem that only a few closed-form (analytic) expressions of Rao–Blackwell estimators have been derived. For example, (Dryver and Thompson, 1998) present formulas for ACS Rao–Blackwell estimators and Rao–Blackwell estimators of their respective variances that are not closed-form (given an initial SRSWOR). Felix-Medina (2000) derived closed-form expressions of Rao–Blackwell modified Horvitz–Thompson estimators for ACS with an initial SRSWOR and for ACS with an initial unequal probability sample with replacement. Expressions for the variances of these estimators, as well as closed-forms for unbiased Rao–Blackwell estimators of those variances, are also derived. For the without replacement case, derivations are based on the multivariate hypergeometric distribution. For the with replacement case, derivations are based on the multinomial distribution. These closed-forms allow users of ACS to take advantage of computer software that compute probabilities from hypergeometric or multinomial distributions.

Rocco (2003) utilized Rao–Blackwell theory to develop estimators for μ based on the set of distinct and unordered data from only the initial sample. Because these data are a function of the minimally sufficient statistic D_R, the former are, therefore, a sufficient statistic for $\boldsymbol{\theta}$.

6. Controlling the final sample size

A major concern with ACS is the randomness of the final sample size. To address this problem, inverse sampling and restricted ACS procedures were developed to control the final sample size. For both methods, an initial sample is sequentially selected with a stopping rule applied to indicate termination of sampling.

6.1 *Inverse sampling*

Christman and Lan (1998) considered ACS based on inverse sampling, i.e., sequentially selecting initial sampling units until some prespecified number (k) of nonzero y-values are observed. Hence, the number of networks intersected by the initial sample will then be less than or equal to k. The rationale is that by stopping after we sample k units satisfying C, we control the sample size and avoid estimates of zero abundance. For this form of sequential sampling, known as *inverse adaptive cluster sampling*, the usual modified ACS Hansen–Hurwitz estimator in (8) will be positively biased. Rocco (2003) also notes that ACS estimators will be biased using inverse ACS.

Christman and Lan (1998) introduced $\widehat{\tau}_{IA}$, a sequential unbiased ACS estimator of τ that is based on the use of network means in conjunction with a conventional sequential SRSWOR or SRSWR estimator. For comparative purposes, the ACS Hansen–Hurwitz estimator in (8) was also considered with n_1 set equal to the expected size of the sequential sample and the sample was randomly taken. This nonsequential ACS estimator is unbiased. Simulations of samples from a small population were taken according to the sequential adaptive strategy under both WR and WOR strategies. Results showed that adding an adaptive component to sequential sampling reduced the variance of two different conventional estimators. The variance of the nonsequential ACS estimator was smaller than the variance of $\widehat{\tau}_{IA}$. With the former, however, there is a positive probability of returning an estimated value of 0 for τ. Because the distribution of $\widehat{\tau}_{IA}$ was highly skewed right, Christman compared efficiencies using the median absolute deviations of the estimator, a measure more resistant to extreme values than the variance. Using this criterion, the results showed that $\widehat{\tau}_{IA}$ was more efficient than the comparative nonsequential ACS estimator. The decision regarding which type of ACS design to use, sequential versus nonsequential, amounts to evaluating the tradeoff between two risks: the risk of obtaining a very high outlying estimate and the risk of obtaining a zero estimate. This decision could be facilitated by considering the rarity of the population because, by construction, the more rare the population the lower the probability of obtaining a very high $\widehat{\tau}_{IA}$.

Christman and Lan (2001) extended their prior research by considering two stopping rules and by deriving unbiased nonzero estimators of τ that require an estimate of the proportion of the population that possess the rare characteristic of interest. Because the variance formulas did not permit a direct analytic comparison of efficiencies across ACS and nonACS designs and across stopping rules, they ran a simulation study. Three duck populations in Smit *et al.* (1995) and an artificial (ART) population for which quadrat counts are low and 29% of the quadrat counts are nonzero were considered in the simulation study. The simulation results indicated that the ACS estimators almost always have smaller variability than the nonACS estimators, while for the ART population, the variability was similar across all estimators. For all four populations and any choice of k, the unbiased estimator based on inverse sampling $\widehat{\tau}_{IA}$ was the best because it had the smallest variance. Also, the final sample size is only slightly larger when adaptively sampling. Overall, their simulation study indicated that a modified stopping rule is the best sampling strategy in terms of having the smallest estimator variance. This modified strategy incorporates adaptive sampling and utilizes an initial SRSWR of fixed size such that if at least k units satisfying C are not observed in the initial sample, sequentially sample until k units are observed.

In Salehi and Seber (2001), a new proof which extends the use of Murthy's estimator (Murthy, 1957) to any sequential sampling plan (including inverse sampling), as well as any fixed size sampling design, is given (see Section 7). This proof applies the Rao–Blackwell Theorem to a very simple estimator to obtain the general form of the Rao–Blackwell estimator $\widehat{t}_{RB}$ which is also the same as the unbiased Murthy estimator $\widehat{t}_M$. Explicit simplified forms of the estimator $\widehat{\mu}$ of μ and an unbiased variance estimator $\widehat{\mathrm{Var}}[\widehat{\mu}]$ are provided for inverse sampling.

The unbiased estimator $\widehat{\tau}_I$ for inverse sampling derived in Christman and Lan (2001) is actually Murthy's estimator of τ. This can be seen by multiplying the $\widehat{\mu}$ estimator in Salehi and Seber (2001) by N. Although Christman and Lan (2001) comment that they are pursuing approaches to estimating the variance, an unbiased estimator is provided by Salehi and Seber (2001).

Rocco (2003) introduced a procedure called *constrained inverse adaptive cluster sampling* (CAICS). With CIACS an initial SRS is taken and if at least two units satisfy C then no further sampling is conducted. Otherwise, sampling is continued sequentially until at least two units satisfying C are obtained. Two different initial samples, denoted s^n and s^{n+1}, were considered. When the number of units in the initial sample exceeded n_1, she considered the sample s^n which does not include the last selected unit and the sample s^{n+1} which includes all selected units. Unbiased Rao–Blackwell estimators for μ were developed using Horvitz–Thompson estimation and by conditioning on the set of distinct, unordered data from only the initial sample. This was done for both initial sample types. Rao–Blackwell expressions of the variance of the two estimators and of their unbiased variance estimators were also proposed.

Rocco (2003) conducted a simulation study to assess the relative efficiency of the two new estimators compared to both SRS and the usual inverse sampling. Ten patchy populations were simulated using a Poisson cluster process model. 10,000 CAIC s^n and s^{n+1} samples were taken from each population, and the corresponding estimators were used to estimate μ. Final sample size and the number of those units

satisfying C were also obtained. For both sample types, the expected final sample size was used to subsequently draw (i) 10,000 samples via SRSWOR for comparative purposes and (ii) 10,000 inverse samples containing a number of units satisfying C equal to the expected number of final sampled units satisfying C as observed in the 10,000 CAIC samples. The sample average was calculated ignoring the last selected unit along with the equally weighted mean of all possible values of these sample averages. In general, the CAICS estimator is more relatively efficient when the last unit selected in the initial sample and the corresponding network were not considered than when they are. For both initial sample types, CIACS was more relatively efficient than inverse sampling and SRSWOR for almost all of the populations considered.

6.2 *Restricted ACS designs*

Brown (1994) and Brown and Manly (1998) presented *restricted* ACS designs that allow the final sample size (n_f) to be set prior to sampling. Units in the initial sample are sequentially selected and the cluster associated with a selected unit is sampled if C is met. If the cumulative total sample size is $< n_f$, a unit is added to the initial sample. If this unit and associated units in the cluster result in the cumulative total sample size exceeding n_f, the cluster is included and sampling ceases, otherwise, the initial sample is increased by a unit and adaptive sampling continues. This methodology restricts the selected initial sample to one that produces a final sample size at or just over n_f and has the effect of reducing variation in the final sample size.

However, for ACS, using the $\widehat{\mu}_{HT}$ and $\widehat{\mu}_{HH}$ estimators for this design results in a positive bias with the severity of the bias depending on the population structure. In certain types of populations this bias can be accurately estimated via bootstrapping and adjustments can be made to the estimators. Using $\widehat{\text{Var}}[\widehat{\mu}_{HT}]$ and $\widehat{\text{Var}}[\widehat{\mu}_{HH}]$ variance estimators in the restricted ACS setting yielded biased, but in all cases, ''reasonable'' results (Brown, 1994; Brown and Manly, 1998). Salehi and Seber (2002) state that under restricted ACS, $\widehat{\mu}_{HH}$ will, in general, yield an overestimate. An exception to this is the hypernetwork case which will tend to yield an underestimate.

In a simulation study across an array of population models, Brown (1994) showed results that generally favored restricted over unrestricted ACS for patchy populations in terms of MSE. However, in a similar and more extensive study, Brown and Manly (1998) discovered that restricted ACS was actually worse than unrestricted ACS in terms of relative efficiency for all population models under consideration.

One potential drawback of restricted ACS is that if the critical value is set too low, then most or all the units selected will be from the same cluster because the final sample size will be immediately obtained. Thus, there will be poor coverage of the study area. Again, prior knowledge of the population structure is very helpful. Another concern is that prior development of a travel plan between units in the initial sample cannot be done. If travel between units is difficult, then this might be a significant problem. For some studies, a possible solution might be to stratify the study area and set a desired final sample size within each stratum (Brown and Manly, 1998).

Salehi and Seber (2002) investigated the case when the initial sample is without replacement and restricted ACS is then applied. Salehi and Seber refer to this sampling scheme as restricted *ACS with networks selected with replacement* because a network can be selected more than once. They derived a closed-form unbiased estimator $\widehat{\tau}_{RB}$ of τ by applying the results of Salehi and Seber (2002) on Murthy's estimator for sequential sampling schemes. An unbiased estimator of the variance of $\widehat{\tau}_{RB}$ was also derived. In their simulation, $\widehat{\tau}_{RB}$ outperformed $\widehat{\tau}_{HT}$ and $\widehat{\tau}_{HH}$ with respect to the MSE (particularly for a smaller stopping value M). A problem, however, can arise with a rare population. For example, suppose a sample contains a hypernetwork of size $\geq M$ and all other clusters are of size 1 with $y = 0$. Using the unbiased estimator, we would get a 0 estimate. On the other hand, when a biased estimator is used, the estimate could be unreasonably large when 0 is more realistic. The probability of this occurring increases with decreasing M. Thus, for small M, the use of restricted ACS is questionable.

Salehi and Seber (2002) considered working with cost as a restriction assuming that the cost of sampling edge units differs from the cost of sampling nonedge units. They propose that sampling is to continue until the total cost reaches a prespecified value C_T. They referred to this scheme as restricted *ACS with networks selected without replacement*, and derived unbiased estimators $\tilde{\tau}_{RB}$ and $\mathrm{Var}[\tilde{\tau}_{RB}]$. Salehi and Seber note that (i) when the cost is relatively small for sampling units not satisfying the ACS criterion or when a rapid assessment variable is available, then restricted ACS can be made to be more efficient and (ii) the unbiased estimators $\widehat{\tau}_{RB}$ and $\tilde{\tau}_{RB}$ can be improved by using a Rao–Blackwell approach that incorporates the edge units when computing the estimate.

Salehi and Seber (2002) compare $\widehat{\tau}_{RB}$ and $\tilde{\tau}_{RB}$ in a simulation study using a population from Thompson (1990). They found that for a small stopping value M, $\mathrm{Var}[\widehat{\tau}_{RB}]$ is slightly smaller while for a large M, $\mathrm{Var}[\tilde{\tau}_{RB}]$ is slightly smaller. They cautiously state that if this is the case in general, they would prefer using $\widehat{\tau}_{RB}$ because of the ease of computation.

7. Other estimators

Salehi and Seber (1997a) introduced a modification of ACS in which networks are sampled WOR. They presented an unbiased version of the conventional PPS WOR estimator $\widehat{t}_{Raj}$ of τ due to Raj (1956) along with its variance and an unbiased estimator of its variance. Subsequently, $\widehat{t}_{Raj}$ was modified according to Murtthy (1957) to yield another unbiased estimator $\widehat{t}_M$ of τ along with its variance and an unbiased estimator of its variance. These two estimators were compared to the usual modified ACS Hansen–Hurwitz and Horvitz–Thompson estimators, their Rao–Blackwell estimators, and the SRS estimator via a small population example for a fixed n_1. Results included some interesting phenomena. Using $\widehat{\mu}_{RBHH}$ resulted in a greater relative improvement (variance reduction) than for $\widehat{\mu}_{RBHT}$, which is not a surprising result given the discussion of Salehi's (1999) results in Section 5. Ignoring the Rao–Blackwell estimators, for the design with the networks selected WOR, $\widehat{t}_M$ was more efficient than the other estimators. However, $\widehat{\mu}_{RBHH}$ was the most efficient estimator overall for the example. Next, utilizing computer simulation, the authors analyzed

one of the data sets from Smith *et al.* (1995). Comparisons were made among the estimators except for the Rao–Blackwell versions. Once again, $\widehat{t}_M$ was more efficient than the other estimators and its efficiency gain over the next best estimator $\widehat{\mu}_{HT}$ increased with n (number of sampled networks). Although the authors felt these results would hold for many populations after a review of the literature, they suggested further research. Unfortunately, one disadvantage of $\widehat{t}_M$ is that it requires substantial computation when n_1 is large. Finally, $\widehat{t}_{Raj}$, and consequently $\widehat{t}_M$, can be improved using the Rao–Blackwell method in conjunction with D_R. The reader is referred to (Salehi and Seber, 1997a, pp. 212–214) for derivation of this estimator along with its variance and corresponding unbiased estimator.

8. Order statistics

The selection of C may be difficult or impossible to ascertain in some survey situations. Perhaps the investigator also wants to search for high values of y in addition to estimating μ or τ. In such cases, the criterion for additional sampling may be made relative to the observed sample values by basing it on the sample order statistics. For example, if the initial sample size is n_1, additional sampling would occur in the neighborhoods of a prespecified number (n_1-r) of units where $r \in \{0,1,\ldots,n_1-1\}$. The n_1-r units correspond to the units having the largest n_1-r order statistics, $y_{(r+1)}, y_{(r+2)}, \ldots, y_{(n_1)}$. Thus, C is $y_i \geq y_{(r+1)}$. If any of the additional units satisfy C, then we add their neighborhoods to the sample. This process is iterated until no further units satisfy C. This scheme differs radically from previous schemes in which C is fixed and prespecified. Using order statistics, C, and hence the clusters, are determined by the initial data rather than in advance (Thompson and Seber, 1994). Su and Quinn II (2003) envision that use of order statistics with ACS would only be considered in situations where logistical costs of revisiting sampling units and conducting adaptive sampling are small compared to taking additional samples.

The final sample average is a biased estimator of μ. However, an unbiased estimator $\widehat{\mu}_{ord}$ of μ, used in conjunction with the order statistics of an adaptive cluster sample is the usual Hansen–Hurwitz estimator in (8) (Thompson, 1996; Thompson and Seber, 1996, p. 165). Because α_k in (4) depends on the sample, it can be shown that $\widehat{\mu}_{HT}$ is not unbiased when ACS is based on order statistics (Seber and Thompson, 1994).

Thompson and Seber (1996, pp. 165–166) showed that $\widehat{\mu}_{ord}$ is also the expectation of $\overline{y}_1$, the initial sample mean, conditional on the sufficient statistic $D_f = \{(i, y_i, f_{is}) : i \in s_R\}$ (defined in Section 5 but with subscript s indicating the network configuration is now determined by the sample). Thus, by Rao–Blackwell theory, $\widehat{\mu}_{ord}$ based on an initial sample size n_1 will be more efficient than $\overline{y}_1$ based on a sample size n_1.

Both $\mathrm{Var}[\widehat{\mu}_{ord}]$ and its unbiased estimator have been derived using conditional arguments. Unfortunately, this estimator is computationally complex and can yield negative estimates. Two biased estimators of $\mathrm{Var}[\widehat{\mu}_{ord}]$ exist and guarantee a nonnegative result (Thompson and Seber, 1996, pp. 167).

The Rao–Blackwell method can also be used to improve $\widehat{\mu}_{ord}$, because $\widehat{\mu}_{ord}$ depends on the sample selection order, and therefore it is not a function of D_R. The resulting

Rao–Blackwell estimator, $\widehat{\mu}_{RBord}$, is the expectation of either $\widehat{\mu}_{ord}$ or $\overline{y}_1$ conditional on D_R, which is a minimally sufficient statistic (Thompson and Seber, 1996, pp. 168). Thus, using the Rao–Blackwell theory in Section 5, we have:

$$\widehat{\mu}_{RBord} = \frac{1}{\xi}\sum_{g=1}^{G}\widehat{\mu}_{ord,g}I_g = \frac{1}{\xi}\sum_{g=1}^{G}\overline{y}_{1g}I_g. \tag{44}$$

Explicit formulas for $\mathrm{Var}[\widehat{\mu}_{RBord}]$ and its unbiased estimator are given in (Thompson and Seber, 1996, pp. 167–169).

Thompson and Seber (1996) worked through an example which consisted of drawing all possible samples via SRSWOR from a small population and letting C be $y_i \geq y_{(n)}$. They calculated the initial sample total, $\widehat{\tau}_1$, $\widehat{\tau}_{ord}$, and $\widehat{\tau}_{RBord}$ for each sample and the variance of these estimators. Also, the three different estimators of $\mathrm{Var}[\widehat{\tau}_{ord}]$ were calculated. Upon comparison, $\mathrm{Var}[\widehat{\tau}_{RBord}]<\mathrm{Var}[\widehat{\tau}_1]<\mathrm{Var}[\widehat{\tau}_{ord}]$, where $\mathrm{Var}[\widehat{\tau}_1]$ was calculated using $E[\nu]$. Also, $MSE[\widehat{\mathrm{Var}}[\widehat{\tau}_{ord}]]$ was lower for the two biased estimators than for the unbiased estimator of $\mathrm{Var}[\widehat{\tau}_{ord}]$.

Su and Quinn II (2003) studied properties of ACS designs based on the order statistics (which they call *acsord* designs). Three simulated fish populations having different aggregation levels were generated, and the unbiased Hansen–Hurwitz estimator $\widehat{\mu}_{HH}$ and the biased Horvitz–Thompson estimator $\widehat{\mu}_{HT}$ for *ascord* were compared. The problem of how to handle ties in the y-values without introducing bias in the estimation was also discussed. The simulations indicated that $\widehat{\mu}_{HT}$ was positively biased with the maximum bias increasing as the aggregation level increases. The bias also increased at smaller values of r for more highly aggregated populations. The bias is small when r is small or large for populations with low aggregation. Thus, they recommend a large r for more aggregated populations. They concluded that $\widehat{\mu}_{HH}$ was also more sensitive than $\widehat{\mu}_{HT}$ to the use of order statistics.

With respect to efficiency, Su and Quinn concluded that (i) for a fixed n_1, the efficiencies of $\widehat{\mu}_{HH}$ and $\widehat{\mu}_{HT}$ increase as the populations become more aggregated, (ii) for each population, the efficiencies of $\widehat{\mu}_{HH}$ and $\widehat{\mu}_{HT}$ increase as n_1 increases. For most cases, $\widehat{\mu}_{HT}$ was more efficient than $\widehat{\mu}_{HH}$ with $\widehat{\mu}_{HT}$ being much more efficient than $\widehat{\mu}_{HH}$ as n_1 and r become large, and (iii) for the highly aggregated population, $\widehat{\mu}_{HH}$ and $\widehat{\mu}_{HT}$ are more efficient than SRS, while for the lowly aggregated population, SRS is more efficient.

Su and Quinn II (2003) extended the use of *acsord* by combining the use of order statistics with a stopping rule. They define a stopping level S as the S^{th} iteration of ACS beyond the initial sampling unit after which the sampling process terminates. The $\widehat{\mu}_{HH}$ and $\widehat{\mu}_{HT}$ estimators under *acsord* with a stopping rule are now both biased. Based on simulations applied to the three simulated populations, the bias of $\widehat{\mu}_{HH}$ increases as the aggregation level increases and it does not appear to change with n_1. The bias of $\widehat{\mu}_{HT}$ increases as the aggregation level increases for smaller r values, and increasing n_1 reduces the bias for smaller values of r. Su and Quinn stated that if a stopping rule is used then the bias of $\widehat{\mu}_{HH}$ was more sensitive than $\widehat{\mu}_{HT}$ and that it lessens the bias of $\widehat{\mu}_{HT}$ for smaller values of r.

For the two least aggregated populations, the SRS estimator was almost always more efficient than $\widehat{\mu}_{HH}$. For the lowest level of aggregation and for the mid-level with small n_1, the SRS estimator tended to be more efficient than $\widehat{\mu}_{HT}$. For smaller n_1,

$\widehat{\mu}_{HT}$ with the stopping rule tended to yield a more efficient estimator than for the no stopping rule case. The stopping rule had little impact on the efficiency of $\widehat{\mu}_{HT}$ for smaller r values for the largest n_1, while for larger r values, the relative efficiency decreased using the stopping rule and with smaller S.

Because of the presence of interactions related to the choice of order statistic (r), sample size (n_1) and stopping level (S), Su and Quinn II (2003) admit that it would be difficult to provide general recommendations for designing ACS studies. The design problem in the presence of complex interactions was seldom addressed in the numerous studies cited in this review, and it needs to be studied further.

9. Double sampling and multivariate ACS

In some ACS surveys, there will be $p > 1$ variables of interest. In this review so far, C depended on some function of a single variable. In the multivariate setting, there are many options for C (e.g., C may be a function of any one or all of the p variables, such as their sum). In (Thompson and Seber, 1996, Chap. 8) multivariate aspects of ACS are extensively discussed. The reader is referred to that monograph for theory and details. Per the monograph, unbiased estimates, $\widehat{\mu}_{HH}$ and $\widehat{\mu}_{HT}$ still exist for each of the p variables irrespective of the choice of C. Also, an unbiased estimate of the variance–covariance matrix exists. Univariate results can be applied to each of the p-variables, even though C may depend on other variables. The ACS strategy will be efficient relative to the conventional strategy if the within-network variation is large relative to the total variation. Analogous to the univariate case, the relative efficiency will ultimately depend on the particular partitioning of the population into networks, which, in turn, depends on the myriad of issues already discussed.

For a mean vector $\boldsymbol{\mu}$ (or total vector $\boldsymbol{\tau}$) there exist multivariate estimators based on the Rao–Blackwell method. These are simply extensions of the univariate case discussed in Section 5. The reader is referred to (Thompson and Seber, 1996, pp. 206–207) for a straightforward derivation of the variance-covariance matrix for the Rao–Blackwell estimators of $\boldsymbol{\mu}$ along with the unbiased estimator of this matrix.

Dryver (2003) considered multivariate ACS in the setting of estimation of the population mean μ for a single variable of interest, but the condition to adaptively sample was based on another variable or function of several variables. A simulation was conducted using the blue-winged teal (BWT) and green-winged teal (GWT) data in Smith *et al.* (1995). Adaptive sampling was considered under several conditions including Horvitz–Thompson estimation of μ_{BWT} using ACS based on BWT counts, estimation of μ_{GWT} using ACS based on BWT counts, and estimation of both μ_{BWT} and μ_{GWT} using the sum of BWT and GWT counts. The following conclusions were drawn: (i) When prediction was on one type of duck but the condition was on the other, a SRS can be more efficient that ACS. (ii) When the condition was the count sum, the ACS was more efficient that SRS. Two general conclusions were given. First, ACS will perform well if the variables used in the parameter estimation are highly correlated with one another. Second, if more than one parameter is to be estimated and the variables are not highly correlated, then ACS using some function of the variables as the condition should be used to obtain an optimal sampling strategy. This function, however, must be known prior to data collection.

A special case of multivariate ACS is a survey that make use of a rapid assessment variable. In this case, it may be desirable to base ACS on one variable, such as rapid assessment of abundance, in order to estimate τ for another variable, such as actual abundance. For example, Felix-Melina and Thompson (1999) combined ACS and conventional double sampling to develop a new design called Adaptive Cluster Double Sampling (ACDS) that is appropriate to use when C (or C_x) is based on an easy-to-measure auxiliary variable x that is correlated with the difficult-to-measure (or more expensive-to-measure) survey variable y. They speculated that the final sample size of y-values could be reduced while still obtaining a good estimate of μ_y. First, they defined the auxiliary variable x to be a binary variable such that $x=1$ if it is speculated that $y \geq c_y$ and $x=0$ otherwise. That is, the x_i represent quick estimated indicators of whether or not y_i satisfies C_y having critical value c_y. An ordinary adaptive cluster sample based on C_x is then taken. Because x is easy-to-measure, n_1 can be large, thereby increasing the probability of detecting clusters having units satisfying C_x, and hence, C_y. Next, a conventional subsample of the networks is taken from the initial sample. From each of these subsampled networks, conventional and independent subsamples of units are taken and their y-values are measured. Notice the number of y-values can be fixed beforehand. Also, when $x_i=0$ for any adaptively added unit, the more expensive variable of interest y_i does not have to be measured for that edge unit. Thus, the total cost of sampling is reduced.

The goal is to use regression to model the relationship between x and y with the adaptive cluster sample based on x. Felix-Melina and Thompson (1999) constructed regression estimators $\widehat{\mu}_R$ based on $\widehat{\mu}_{HT}$ and $\widehat{\mu}_R^*$, the solution of a system of estimating equations. They also presented asymptotic variances for these estimators along with estimators of their variances using the Delta Method.

To compare the performance of ACDS with that of ACS, a simulation study was conducted. Samples were selected via SRSWOR from a simulated population and three sampling designs were used: the usual ACS design, ACDS I with a prespecified maximum number of y_i to be recorded, and ACDS II where the number of y_i was not specified in advance. For ACDS I, subsampling was conducted after the adaptive sample based on x had been completed implying that sampled networks would be visited twice. For ACDS II, subsampling was conducted immediately after the network x-values were measured so that sampled networks would be visited only once. Cost functions were constructed and defined for each design and initial sample sizes were determined so that expected costs were equivalent. Results showed that the two ACDS designs yielded lower MSEs than the ACS design.

The results of this study raised two issues that might affect the implementation of ACDS. First, the performance of ACDS could be improved when n_1 was increased because the between-network variance was high for the particular simulated population. Thus, the best strategy was one that emphasized sampling more networks. Second, the performance of the ACDS design was very poor with respect to a data set from Smith *et al.* (1995) because the ACDS estimator is not resistant to outliers. That is, it is seriously and negatively affected by high within-network variation. In both cases, the researcher would need some prior knowledge of the nature of the decomposition of the total variance of the population to set appropriate sample sizes for each phase.

A second case of multivariate ACS called *adjusted two-stage* ACS (ATSACS) (Muttlak and Khan, 2002) makes use of a rapid assessment auxiliary variable when it is expected that some networks are large while others are small. For ATSACS, a second-stage sample of network units is selected from the large networks using SRS, while all units will be sampled within the small networks. The goal is to avoid a large final sample size while still obtaining a good estimate of μ_y. Because only large networks are subsampled in ATSACS, it is necessary to rapidly determine which networks are large without having to observe the y-value for every network unit and then proceed to subsample only the large networks. A network is classified as large if the network size is $\geq k$. The choice of k is selected after accounting for factors such as the amount of time and money it costs to sample a unit. Muttlak and Khan (2002) provide an unbiased estimator $\widehat{\mu}_{mk}$ of μ_y using ATSACS that is a linear combination of an estimator based on the two-stage sample of large networks and a Hansen–Hurwitz estimator from the sample of small networks. They used the blue-winged teal population given in Smith *et al.* (1995) to compare the ATSACS to SRS and ACS. They defined the binary rapid assessment variable x based on the presence ($x=1$) or absence ($x=0$) of blue-winged teals in a sampling unit. A network was classified as large if it had at least 5 units with $x=1$. One initial SRS of size $n=12$ was taken yielding ACS and ATSACS estimates that were close to μ and an SRS estimate that was not. The unbiased variance estimates were a degree of magnitude smaller under ACS or ATSACS than under SRS, but was still 20% larger under ATSACS than under ACS. Because only one sample was examined and because the final sample sizes differed, it is unknown how the variances of the ATSACS and SRS estimators differ for comparable sample sizes (i.e, when the SRS size equals the effective sample size under ATSACS).

Smith *et al.* (2003) introduced another application of double sampling with the goal of avoiding costly oversampling. Based on a qualitative assessment of freshwater mussel density at 30 sites, they classified each site as either low or high density. Then ACS and conventional sampling were used to subsample low and high density sites, respectively.

10. Two-stage ACS designs

Salehi and Seber (1997b) described a two-stage ACS design that used SRSWOR of primary units, and then ACS of secondary units within each of the selected primary units. Two variations on this design were proposed. The first was a scheme where the clusters were allowed to cross primary unit boundaries while in the second scheme, the clusters were truncated at the boundaries. The estimators and their variances in Section 2.3 were modified for both schemes. Their unpublished results suggested that the nonoverlapping scheme appeared to be more efficient in terms of estimation, although the comparison was dependent on the population structure. They also stated that they preferred the use of their Horvitz–Thompson estimator. However, Felix-Medina and Thompson (1999) cautioned that if the clusters are large relative to the size of the primary units, then only their portions that are intersected by the primary units in the initial sample are sampled and a reduction in efficiency of this sampling design would be expected.

When the sampling fraction of primary units was small, (Salehi and Seber, 1997b) constructed an upper bound on $\mathrm{Var}[\widehat{\mu}_{HH}]$. They derived a formula for the number of primary units, m, to be sampled to achieve the desired upper bound on $\mathrm{Var}[\widehat{\mu}_{HH}]$. Input for the formula came from a pilot study. They also developed a cost function for the nonoverlapping scheme. Using m in the cost function would allow the investigator to plan for the total cost of the survey for a given desired level of precision.

11. Detectability in ACS

In this review so far, it has been assumed that y_i is recorded without error for every sample unit. In certain studies, this will not be a realistic assumption. The probability that an object in a selected area is observed is called its *detectability*. Imperfect detectability is a prevalent source of nonsampling error in many surveys which will lead to an underestimate of τ (Thompson, 1992). For example, Smith *et al.* (1995) admitted that detectability was an issue in their research. To estimate τ in a survey with imperfect detectability, both the sampling design and the detection probabilities must be taken into account. The complexity of this situation is more difficult with ACS designs than with conventional designs because the probability of including an object depends on both its detection probability and the detection probability of objects in nearby units. Fortunately, results that adjust for detectability and extensions of this theory to ACS for a variety of cases have been worked out in a straightforward fashion (Thompson and Seber, 1994, 1996). Two main ideas were utilized. First, an estimator of τ, unbiased under the assumption of perfect detectability, can be adjusted, according to the case, by the detection probability or its estimate. Second, for derivation purposes, expectations and variances with respect to the initial sample conditional on the detectability were determined followed by expectations with respect to detectability.

Two common cases involving the probability of detection g are (i) g is assumed to be known, constant, and independent for each object in the population and (ii) g is constant, unknown, and has to be estimated. Additional cases include modification of cases (i) and (ii) such that $g = g_{ij}$, i.e. the detection probability is not constant and can vary among j objects in i units in the population. Obviously, the level of complexity increases accordingly in these latter situations. Refer to (Thompson and Seber, 1996, Chap. 9) for the associated estimators, their variances and estimated variances, and the underlying theory.

The effect of imperfect detectability on the relative efficiency of SRS and ACS would be to add an identical amount to the variance under each strategy, so that the relative efficiency, whether greater than or less than 1, would move slightly closer to 1 (Thompson and Seber, 1996).

12. Bootstrapping and interval estimates

Confidence intervals for τ are typically based on the assumption of asymptotic normality of the sampling distribution of the estimator. However, for any ACS

design, the sampling distribution of $\hat{\tau}$ can be very asymmetric (Christman, 1997; Christman and Pontius, 2000) and discontinuous (Christman, 2000). Thus, a confidence interval based on the assumption of normality is inappropriate. As an option, bootstrapping can be used to generate an empirical sampling distribution for $\hat{\tau}$ which allows computation of interval estimates.

Pontius and Christman (1997) noted that any bootstrapping procedure used would have to account for the finite population size. This was accomplished by modifying the bootstrapping algorithm of Sitter (1992a, b). The authors conducted a small ACS simulation study and compared normal approximation, percentile, and bias-corrected and accelerated bootstrap confidence intervals for τ using their modified finite population bootstrapping algorithm in conjunction with $\hat{\tau}_{HT}$. For the specific population and various network sizes, Sitter's bootstrap algorithm yielded bootstrapped estimates of τ that were, in general, close to the sample estimates. However, the authors cautioned that the network sizes were small relative to N and that the variability in interval estimates might increase with larger networks. All three types of confidence intervals yielded minimal biases (all slightly underestimated the nominal confidence level) and similar interval widths for all given sample sizes. The confidence interval based on the normal approximation had, not surprisingly, the worst coverage for all sample sizes, whereas the bias-corrected and accelerated confidence interval always had the best coverage.

Christman and Pontius (2000) addressed these topics when using $\hat{\tau}_{HH}$. They did not, however, consider bootstrapping interval estimators based on $\hat{\tau}_{HT}$ due to theoretical complications arising from an unequal probability sampling design that involves a combination of WR and WOR selections for networks. Four nonparametric, finite population bootstrap methods and three types of confidence intervals were investigated where the input sample consisted of n_1 network totals. The results depended on the type of population being simulated and n_1. Results were also obtained using the data from Smith *et al.* (1995). Bootstrap percentile intervals, in general, provided better results when compared to the asymptotic normal and the bias-corrected and accelerated intervals. Moreover, the authors recommended using the Bootstrap With Replacement method of McCarthy and Snowden (1985) with percentile intervals.

DiBattista and DiSpalatro (1999) presented a resampling procedure, which they called the Bootstrap for Adaptive Cluster Sampling (BACS). BACS was used to estimate the bias of the sample mean and sample variance of the final ACS sample given initial SRSWR. They showed through proof and a simulation study that the BACS method is consistent, i.e. as $n \rightarrow \infty$, bias estimated via the BACS method converges to the actual bias generated via ACS.

Conclusion

By definition, a sampling strategy for estimating or predicting a population quantity $z(\mathbf{y})$ consists of a sampling design $P(s|\mathbf{y})$ and an estimator $\hat{z}(d_s)$. In Section 5 (with $z(\mathbf{y}) = \phi(\boldsymbol{\theta})$), it was stated that no UMVUE exists for the class of all adaptive and conventional designs in the fixed-population setting considered in this review. Consequently, there is no optimal strategy in the fixed-population strategy such that:

$$\mathrm{Var}[\hat{z}_{opt}(d_s)|\mathbf{y}] = \min_{(P,\hat{z}) \in B} Var[\hat{z}(d_s)|\mathbf{y}]$$

for all **y**, where B is the class of all design-unbiased strategies (Thompson and Seber, 1996, pp. 235–236). This result was supported by the numerous studies and simulations cited in this review.

Despite the fact that there is no optimal strategy, ACS can provide efficient estimation with judicious choice of design type, estimator, and design factors (such as critical value, neighborhood choice, unit size, and sample size). Smith *et al*. (2003) provided a proper assessment of the current state of affairs in ACS when they said "The ultimate (and currently unanswered) question is how to tell when ACS should be used, i.e., when is a population sufficiently rare and clustered for ACS to be efficient and practical?".

References

Borkowski, J.J. (1999) Network inclusion probabilities and Horvitz–Thompson estimation for adaptive simple Latin square sampling. *Environmental and Ecological Statistics*, **6**, 291–311.

Brown, J.A. (1994) The application of adaptive cluster sampling to ecological studies. In *Statistics in Ecology and Environmental Monitoring*. pp. 86–97.

Brown, J.A. (1996) The relative efficiency of adaptive cluster sampling for ecological surveys, Mathematical and Information Sciences Report Series B:96/08, Massey University, Palmerton North, Department of Statistics.

Brown, J.A. (2003) Designing an efficient adaptive cluster sample. *Environmental and Ecological Statistics*, **10**, 95–105.

Brown, J.A. and Manly, B.J.F. (1998) Restricted adaptive cluster sampling. *Environmental and Ecological Statistics*, **5**, 49–63.

Christman, M.C. (1996a) Comparison of efficiency of adaptive sampling in some spatially clustered populations. In *ASA Proceedings of the Section on Statistics and the Environment*, pp. 122–126.

Christman, M.C. (1996b) Efficiency of adaptive sampling designs for spatially clustered populations, Technical Report, Department of Statistics, Stanford University.

Christman, M.C. (1997) Efficiency of some sampling designs for spatially clustered populations. *Environmetrics*, **8**, 145–166.

Christman, M.C. (2000) A review of quadrat-based sampling of rare, geographically clustered populations. *Journal of Agricultural, Biological, and Environmental Statistics*, **5**(2), 168–201.

Christman, M.C. and Lan, F. (1998) Sequential adaptive sampling designs to estimate abundance in rare populations, In *ASA Proceedings of the Section on Statistics and the Environment*, pp. 87–96.

Christman, M.C. and Lan, F. (2001) Inverse adaptive cluster sampling. *Biometrics*, **57**, 1096–1105.

Christman, M.C. and Pontius, J.S. (2000) Bootstrap confidence intervals for adaptive cluster sampling. *Biometrics*, **56**(2), 503–510.

DiBattista, T. (2003) Resampling methods for estimating dispersion indices in random and adaptive designs. *Environmental and Ecological Statistics*, **10**, 83–93.

DiBattista, T. and DiSpalatro, D. (1999) A bootstrap method for adaptive cluster sampling. In *Classification and Data Analysis: Theory and Application – Proceedings of the Biannual Meeting of the Classification Group of the Societ Italiana di Statistica (SIS)*, pp. 19–26.

DiConsiglio, L. and Scanu, M. (2001) Some results on aymptotics in adaptive cluster sampling. *Statistics & Probability Letters*, **52**, 189–197.

Dryver, A.L. (2003) Performance of adaptive cluster sampling estimators in a multivariate setting. *Environmental and Ecological Statistics*, **10**, 107–113.

Dryver, A.L. and Thompson, S.K. (1998) Improving unbiased estimators in adaptive cluster sampling. In *ASA Proceedings of the Section on Survey Research Methods*, pp. 727–731.

Felix-Medina, M.H. (2000) Analytical expressions for Rao–Blackwell estimators in adaptive cluster sampling. *Journal of Statistical Planning and Inference*, **84**, 221–236.

Felix-Medina, M.H. (2003) Asymptotics in adaptive cluster sampling. *Environmental and Ecological Statistics*, **10**, 61–82.

Felix-Medina, M.H. and Thompson, S.K. (1999) Adaptive cluster double sampling. In *ASA Proceedings of the Section on Survey Research Methods*, pp. 339–344.

Hansen, M.M. and Hurwitz, W.N. (1943) On the theory of sampling from finite populations. *Annals of Mathematical Statistics*, **14**, 333–362.

Horvitz, D.G. and Thompson, D.J. (1952) A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, **47**, 663–685.

Lehmann, E.L. (1983) *Theory of Point Estimation*. John Wiley & Sons.

McCarthy, P.J. and Snowden, C.B. (1985) The bootstrap and finite population sampling. In *Vital and Health Statistics*, Series 2, Volume 95, Public Health Service Publication. Department of Health and Human Services, Washington, DC, pp. 85–1369.

Murthy, M.N. (1957) Ordered and unordered estimators in sampling without replacement. *Sankhya*, **18**, 379–390.

Muttlak, H.A. and Khan, A. (2002) Adjusted two-stage adaptive cluster sampling. *Environmental and Ecological Statistics*, **9**, 111–120.

Pontius, J.S. (1996) Probability proportional to size selection in strip adaptive sampling. Technical Report Series II, Theory and Methods, Kansas State University, Department of Statistics.

Pontius, J.S. (1997) Strip adaptive cluster sampling: Probability proportional to size selection of primary units. *Biometrics*, **53**, 1092–1096.

Pontius, J.S. and Christman, M.C. (1997) Bootstrap confidence intervals from adaptive sampling of an insect population. In *Proceedings of the Ninth Annual Kansas State University Conference on Applied Statistics in Agriculture*, 1997, pp. 191–203.

Raj, D. (1956) Some estimators in sampling with varying probabilities without replacement. *Journal of the American Statistical Association*, **51**, 269–284.

Rocco, E. (2003) Constrained inverse adaptive cluster sampling. *Journal of Official Statistics*, **19**, 45–57.

Roesch, F.A. Jr. (1993) Adaptive cluster sampling for forest inventories. *Forest Science*, **39**, 655–669.

Salehi, M.M. (1999) Rao–Blackwell versions of the Horvitz–Thompson and Hansen–Hurwitz in adaptive cluster sampling. *Environmental and Ecological Statistics*, **6**, 183–195.

Salehi, M.M. (2003) Comparison between Hansen–Hurwitz and Horvitz–Thompson estimators for adaptive cluster sampling. *Environmental and Ecological Statistics*, **10**, 115–127.

Salehi, M.M. and Seber, G.A.F. (1997a) Adaptive cluster sampling with networks selected without replacement. *Biometrika*, **84**, 209–219.

Salehi, M.M. and Seber, G.A.F. (1997b) Two-stage adaptive cluster sampling. *Biometrics*, **53**, 959–970.

Salehi, M.M. and Seber, G.A.F. (2001) A new proof of Murthy's estimator which applies to sequential sampling. *Australian and New Zealand Journal of Statistics*, **43**, 281–286.

Salehi, M.M. and Seber, G.A.F. (2002) Unbiased estimators for restricted adaptive cluster sampling. *Australian and New Zealand Journal of Statistics*, **44**, 63–74.
Seber, G.A.F. and Thompson, S.K. (1994) Environmental adaptive sampling. In *Handbook of Statistics*. Volume 12: *Environmental Statistics*, pp. 201–220.
Sitter, R.R. (1992a) A resampling procedure for complex survey data. *Journal of the American Statistical Association*, **87**, 755–765.
Sitter, R.R. (1992b) Comparing three bootstrap methods for survey data. *The Canadian Journal of Statistics*, **20**, 135–154.
Smith, D.R., Conroy, M.J. and Brakhage, D.H. (1995) Efficiency of adaptive cluster sampling for estimating density of wintering waterfowl. *Biometrics*, **51**, 777–788.
Smith, D.R., Villella, R.F. and Lemarié, (2003) Application of adaptive cluster sampling to low-density populations of freshwater mussels. *Environmental and Ecological Statistics*, **10**, 7–15.
Su, Z. and Quinn, T.J. II (2003) Estimator bias and efficiency for adaptive cluster with order statistics and a stopping rule. *Environmental and Ecological Statistics*, **10**, 17–41.
Thompson, S.K. (1990) Adaptive cluster sampling. *Journal of the American Statistical Association*, **85**, 1050–1059.
Thompson, S.K. (1991a) Adaptive cluster sampling: Designs with primary and secondary units. *Biometrics*, **47**, 1103–1115.
Thompson, S.K. (1991b) Stratified adaptive cluster sampling. *Biometrika*, **78**, 389–397.
Thompson, S.K. (1992) *Sampling*. John Wiley & Sons.
Thompson, S.K. (1994) Factors influencing the efficiency of adaptive cluster sampling. Technical Report Number 94-0301, Technical Reports and Reprints Series, Pennsylvania State University, Department of Statistics.
Thompson, S.K. (1996) Adaptive cluster sampling based on order statistics. *EnvironMetrics*, **7**, 123–133.
Thompson, S.K. and Seber, G.A.F. (1994) Detectability in conventional and adaptive sampling. *Biometrics*, **50**, 712–724.
Thompson, S.K. and Seber G.A.F. (1996) *Adaptive Sampling*. John Wiley & Sons.

Biographical Sketch

Philip Turk is a Visiting Professor of Statistics at Montana State University. His current research involves computer simulation studies of rare and clustered populations and the assessment of factors affecting the relative efficiency of ACS estimators.

John Borkowski is an Professor of Statistics at Montana State University. In addition to his current research in sampling and experimental design, he is involved in collaborative projects related to the impact of winter use vehicles on wildlife, estimation of predation rates and offtake in ungulate populations, and revegetation of weed-infested areas with native plant species.