



Catalogue no. 11-522-XIE

**Statistics Canada International Symposium
Series - Proceedings**

**Symposium 2004: Innovative
Methods for Surveying
Difficult-to-reach Populations**

2004



LINK-TRACING SAMPLING WITH AN INITIAL SAMPLE OF SITES SEQUENTIALLY SELECTED: ESTIMATION OF THE POPULATION SIZE

Martín H. Félix-Medina and Pedro E. Monjardin¹

ABSTRACT

In this paper we modify the variant of link-tracing sampling proposed by Félix-Medina and Thompson (2004) by considering a sequential sample of sites instead of a simple random sample of sites. Our proposed variant allows the sampler to have certain control of the final sample size or the precision of the estimator of the population size. We propose that the population size be estimated either by a maximum likelihood estimator proposed by Félix-Medina and Thompson (2004) or by an estimator derived under the Bayesian approach and proposed by Félix-Medina and Monjardin (2004). In addition, we propose that confidence intervals for the population size be constructed by bootstrap methods. The results of a simulation study indicate that our proposed design performs acceptably well.

KEYWORDS: Bayesian Inference Approach; Bootstrap; Design-Based Approach; Finite Population; Maximum Likelihood; Sequential Sampling; Stopping Rule Principle.

1. INTRODUCTION

Link-tracing sampling (LTS) has been found appropriate for sampling hidden and hard-to-access human populations, such as drug-user, homeless-person, or illegal-worker populations. The idea behind this method is to start with an initial sample of members of the target population, and then to increase the sample size by adding to the sample the new members that are nominated by the people in the initial sample. The sample could be increased again by adding to the sample the new members that are nominated by the previously nominated people. The sampling process continues in this way until a specified stopping rule is satisfied.

Although several variants of LTS have been suggested, recently Félix-Medina and Thompson (2004) proposed a new variant in which the initial sample is a simple random sample (SRS) without replacement of clusters or accessible sites, such as bars, parks or blocks, selected from a frame that covers only a portion of the population. Then, as in ordinary LTS, the people in each selected site are asked to nominate other members of the population. By using this initial sampling design the authors achieve two goals. First, they avoid making the ordinary and difficult to satisfy assumption of an initial Bernoulli sample; and second, they used the probability distribution of the initial sampling design to construct design-based variance estimators that are robust to certain model assumptions.

In this paper we present a modification of the design proposed by Félix-Medina and Thompson (2004). This new variant will allow the sampler the possibility of having certain control of the size of the final sample, or of the size of any of its subsets, such as the members in the initial sample or the nominated members in the portion of the population not covered by the frame. In addition, it could allow the sampler to control the precision of the estimators of the population size. Our strategy is to select an initial sequential sample of sites and to make the nominations as they are being included in the initial sample. By using appropriate stopping rules, the sampler could achieve the specified size of the final sample or of any of its components, or achieve the specified precision of the estimators.

¹ Martín H. Félix-Medina and Pedro E. Monjardin, Escuela de Ciencias Físico-Matemáticas, Universidad Autónoma de Sinaloa, Ciudad Universitaria, Culiacán Sinaloa, México, 80010. E-mails: mhfelix@uas.uasnet.mx and pemo@uas.uasnet.mx.

The structure of the paper is as follows. In Section 2 we introduce the notation to be used throughout the paper and describe the proposed sampling design. In Section 3 we present the different estimators of the population sizes to be used along with this design. In Section 4 we describe the use of the bootstrap to construct design-based confidence intervals for the population sizes. In Section 5 we present the results of a simulation study carried out to observe the performance of the proposed design. Finally, in Section 6 we present some conclusions.

2. NOTATION AND SAMPLING DESIGN

We will assume the same population structure as that considered by Félix-Medina and Thompson (2004). Thus, we will suppose that a finite hidden human population $U = \{u_1, \dots, u_\tau\}$ of unknown size τ can be divided into two portions U_1 and U_2 of unknown sizes τ_1 and $\tau_2 = \tau - \tau_1$. In addition, we will suppose that the portion U_1 is covered by a sampling frame of N accessible sites A_1, \dots, A_N , such as bars, parks, or blocks, of sizes (numbers of persons in the sites) m_1, \dots, m_N . Finally, we will assume that we can determine whether or not a person belongs to the region covered by the sampling frame and, in the affirmative case, the site to which that person belongs. Notice that $\tau_1 = \sum_1^N m_i$ as a person belong to only one site.

The sampling design is as follows. A site is selected from the frame by simple random sampling without replacement. Let A_1 denote the selected site. The m_1 members who belong to A_1 are identified and they are asked to nominate other members of the population outside A_1 , that is, in $U_1 - A_1$. As a convention, we will say that a person is nominated by a site if at least one member of the site nominates that person. It is worth noting that one of different possible nomination strategies can be used. For instance, the m_1 members, as a group, could make the nominations, or each of the m_1 members could separately make the nominations. Finally, for each nominated person the region U_1 or U_2 to which that person belongs is registered.

After the nomination process is concluded in this site, a stopping rule is verified to see whether or not it has been satisfied. If the stopping rule has been satisfied, the sampling process ends, otherwise the previous process is carried out again until the stopping rule is satisfied. It is worth noting that the nominations made from the different sites must be carried out independently, but different nomination strategies could be used in different sites. Also, noting that at the end of the sampling process we will have an initial ordered sample $S_0 = (A_1, \dots, A_n)$ of the n sites that were sequentially selected. We will call the design that gives rise to S_0 a sequential simple random sampling (SSRS) design.

With respect to the stopping rule, any stopping rule that is a function only of the observed data could be used. For instance, if the goal were to control the size of the final sample or the size of one of its subsets, such as the nominated members in U_1 or in U_2 , then we could end the sampling process when the size reaches a specified upper bound or it exceeds that upper bound for first time. Similarly, if the goal were to control the sampling cost, then we could end the sampling process when the sampling cost reaches a specified value or it exceeds that value for first time. As a final example, if the goal were to control the precision of one of the estimators of τ_1 , τ_2 or τ , then we could end the sampling process when the estimator of the variance of that estimator, or the length of the confidence interval constructed using that estimator reaches a specified value or it is below that value for first time. It is worth noting that the proposed sampling design makes difficult to end the sampling process when the stopping rule exactly reaches the specified value because, firstly, the members in the initial sample are selected by clusters, and not one by one, and secondly, because the nomination process carried out from a site is stopped until no more people are nominated.

To end this section we will introduce most of the notation to be used throughout the paper. Let $m = \sum_1^n m_i$ be the number of members in S_0 ; r_1 and r_2 be the numbers of distinct nominated members in $U_1 - S_0$ and in U_2 ; $z_1^{(1)}$ and $z_1^{(2)}$ be the numbers of members in $U_1 - A_i$ and in U_2 that are nominated by the site A_i , $i = 1, \dots, n$; $x_{ij}^{(1)}$ be equal to 1 if person u_j in $U_1 - A_i$ is nominated by the site A_i , and be equal to 0 otherwise; and $x_{ij}^{(2)}$ be equal to 1 if person u_j in U_2 is nominated by the site A_i , and be equal to 0 otherwise.

3. ESTIMATORS OF THE POPULATION SIZES

In the design proposed by Félix-Medina and Thompson (2004) those authors suggest maximum likelihood estimators (MLE's) of the population sizes τ_1 , τ_2 and τ . To obtain those estimators they assume that the sizes m_1, \dots, m_N of the sites A_1, \dots, A_N are realizations of independent Poisson random variables M_1, \dots, M_N with mean λ_1 ; and that given the initial sample, the values of $x_{ij}^{(1)}$ and $x_{ij}^{(2)}$ are realizations of independent Bernoulli random variables $X_{ij}^{(1)}$ and $X_{ij}^{(2)}$ with means $p_i^{(1)}$ and $p_i^{(2)}$, respectively, $i = 1, \dots, n$. From these assumptions and following Darroch's (1958) approach they obtain that the MLE's $\tilde{\tau}_1$ and $\tilde{\tau}_2$ of τ_1 and τ_2 are given as the solution to the following nonlinear equations:

$$\tilde{\tau}_1 = \frac{M + R_1}{1 - (1 - n/N) \prod_1^n (1 - \tilde{p}_i^{(1)})} \quad \text{and}$$

$$\tilde{\tau}_2 = \frac{R_2}{1 - \prod_1^n (1 - \tilde{p}_i^{(2)})}$$

where $\tilde{p}_i^{(1)} = Z_i^{(1)} / (\tilde{\tau}_1 - M_i)$ and $\tilde{p}_i^{(2)} = Z_i^{(2)} / \tilde{\tau}_2$ are the MLE's of $p_i^{(1)}$ and $p_i^{(2)}$, $i = 1, \dots, n$; and M , R_1 , R_2 , $Z_i^{(1)}$ and $Z_i^{(2)}$ are the random variables that give rise to the values m , r_1 , r_2 , $z_1^{(1)}$ and $z_1^{(2)}$, respectively. The MLE $\tilde{\tau}$ of τ is given by $\tilde{\tau} = \tilde{\tau}_1 + \tilde{\tau}_2$.

Those authors carried out a simulation study and found that when the nomination probabilities $p_i^{(2)}$, $i = 1, \dots, n$, are small, the estimator $\tilde{\tau}_2$ is very unstable and seriously overestimates τ_2 . This problem motivated Félix-Medina and Monjardin (2004) to use the Bayesian approach to construct estimators that do not have problems of bias. Although, those authors proposed three sets of estimators, here we will only consider the set obtained from the following initial distributions of τ_1 and τ_2 :

Poisson-Gamma distributions

$$\pi(\tau_1 | \lambda_1) \propto (N\lambda_1)^{\tau_1} / \tau_1! \quad \text{and} \quad \pi(\lambda_1) \propto \lambda_1^{a_1 - 1} e^{-b_1 \lambda_1},$$

$$\pi(\tau_2 | \lambda_2) \propto \lambda_2^{\tau_2} / \tau_2! \quad \text{and} \quad \pi(\lambda_2) \propto \lambda_2^{a_2 - 1} e^{-b_2 \lambda_2},$$

where a_1, b_1, a_2 and b_2 are known constants, τ_1 and τ_2 are conditionally independent given λ_1 and λ_2 ; and λ_1 and λ_2 are also independent random variables.

In the case of the nomination probabilities $p_i^{(k)}, i=1, \dots, n, k=1, 2$, those authors do not consider initial distributions for these variables, but they consider the logits $\alpha_i^{(k)} = \log[p_i^{(k)}/(1-p_i^{(k)})]$ of the $p_i^{(k)}$'s, and they assume the following initial distributions of the $\alpha_i^{(k)}$'s:

$$\alpha_i^{(k)} | \theta_k \sim N(\theta_k, \sigma_k^2) \text{ and } \theta_k \sim N(\mu_k, \gamma_k^2);$$

$i=1, \dots, n, k=1, 2$, where $N(\theta_k, \sigma_k^2)$ stands for the normal distribution with mean θ_k and variance σ_k^2 ; σ_k^2, μ_k and γ_k^2 are known constants; and the $\alpha_i^{(k)}$'s are conditionally independent given θ_k .

They complete the assumptions about the initial distributions by supposing that the random vectors (τ_k, λ_k) and $(\alpha_i^{(k)}, \theta_k), i=1, \dots, n, k=1, 2$, are mutually independent.

Those authors propose that $(\tau_1, \tau_2, \alpha_1, \alpha_2)$, where $\alpha_k = (\alpha_1^{(k)}, \dots, \alpha_n^{(k)})$, $k=1, 2$ be estimated by means of the mode of their joint posterior distribution. Using that strategy they obtained that an estimator $(\hat{\tau}_1, \hat{\tau}_2, \hat{\alpha}_1, \hat{\alpha}_2)$ of $(\tau_1, \tau_2, \alpha_1, \alpha_2)$ is given as the solution to the following system of nonlinear equations:

$$\hat{\tau}_1 = \frac{M + R_1 + (1 - n/N)[N(a_1 - 1)/(N + b_1)] \prod_1^n (1 - \hat{p}_i^{(1)})}{1 - (1 - n/N)[N/(N + b_1)] \prod_1^n (1 - \hat{p}_i^{(1)})};$$

$$\hat{p}_i^{(1)} = \frac{Z_i^{(1)}}{\hat{\tau}_1 - M_i} - \frac{\hat{\alpha}_i^{(1)} - \hat{\alpha}_1}{(\hat{\tau}_1 - M_i)\sigma_1^2} - \frac{\hat{\alpha}_1 - \mu_1}{n(\hat{\tau}_1 - M_i)\nu_1}; \quad i=1, \dots, n;$$

$$\hat{\tau}_2 = \frac{R_2 + [(a_2 - 1)/(1 + b_2)] \prod_1^n (1 - \hat{p}_i^{(2)})}{1 - [1/(1 + b_2)] \prod_1^n (1 - \hat{p}_i^{(2)})};$$

$$\hat{p}_i^{(2)} = \frac{Z_i^{(2)}}{\hat{\tau}_2} - \frac{\hat{\alpha}_i^{(2)} - \hat{\alpha}_2}{\hat{\tau}_2 \sigma_2^2} - \frac{\hat{\alpha}_2 - \mu_2}{n \hat{\tau}_2 \nu_2}; \quad i=1, \dots, n;$$

where $\hat{\alpha}_k = \sum_1^n \hat{\alpha}_i^{(k)} / n$ and $\nu_k = \gamma_k^2 + \sigma_k^2 / n, k=1, 2$. They suggest that τ be estimated by $\hat{\tau} = \hat{\tau}_1 + \hat{\tau}_2$.

In the case of our design, the only difference between it and the one proposed by Félix-Medina and Thompson (2004) is that in the first design the size of the initial sample is random, whereas in the second design the size is fixed. Therefore, by the stopping rule principle, the MLE's to be used in our case are the same as those proposed by Félix-Medina and Thompson (2004). Furthermore, by the same reason, the estimators derived under the Bayesian approach by Félix-Medina and Monjardin (2004) could also be used in our case.

Before we end this section let us make the following two observations. First, although Félix-Medina and Thompson (2004) found that the MLE $\tilde{\tau}_2$ is unstable when the nomination probabilities are small, in the case of our design this problem can be avoided by using a stopping rule that guarantees a minimum number of nominated members in U_2 , or a specified precision of the estimator $\tilde{\tau}_2$. Second, Félix-Medina and Monjardin (2004) used the Bayesian approach only to construct estimators of the population sizes, but they made inferences under the frequentist approach, that is, the population sizes are treated as fixed parameters, and the probability distributions of the estimators are used to make inferences about the parameters. Those authors called this approach a Bayesian assisted approach because it resembles the Model Assisted Survey Sampling approach proposed by Särndal et al. (1992). We will also use the Bayesian assisted approach for making inferences about the population sizes when we employ the previously described estimators derived under the Bayesian approach.

4. BOOTSTRAP CONFIDENCE INTERVALS

In this work, we will use the variant of bootstrap proposed by Félix-Medina and Monjardin (2004) to construct design-based confidence intervals. This variant takes into account both the sampling design used to select the initial sample of sites and the nomination procedure. To do that, the sites are sampled from an artificial finite population using the bootstrap variant proposed by Gross (1980), and the nomination procedure is considered by sampling the indicator variables $X_{ij}^{(k)}$'s from their estimated distributions using the parametric variant of bootstrap. (See Efron and Tibshirani, 1993 for a description of this variant.) The intervals are obtained using either the basic or the percentile method.

The steps of the procedure that we propose are the following. (i) Construct an artificial finite population of N values of m_i 's by repeating N/n times, assuming that N/n is an integer, the selected sample of the n site sizes m_1, \dots, m_n . If N/n is not an integer, use Booth et al. (1994) procedure to construct the finite population, that is, if $N = kn + r$, where k and r are positive integers, then repeat k times the selected sample of n site sizes, and add to this set of m_i 's a sample of r values of m_i 's selected by simple random sampling without replacement from the observed sample of n site sizes. (ii) Select a simple random sample without replacement of size one from the artificial population of the m_i 's. Let m_j be the element that was selected. (iii) Draw samples of sizes $\hat{\tau}_1 - m_j$ and $\hat{\tau}_2$ from Bernoulli distributions with means $\hat{p}_j^{(1)}$ and $\hat{p}_j^{(2)}$, respectively, where $\hat{\tau}_1$, $\hat{\tau}_2$, $\hat{p}_j^{(1)}$ and $\hat{p}_j^{(2)}$ are the estimates of τ_1 , τ_2 , $p_j^{(1)}$ and $p_j^{(2)}$ computed from the original sample. These samples simulate the nominations made from the site A_j . (iv) Determine whether or not the sample selected in steps (ii) and (iii) satisfies the stopping rule used for the original sample. In the positive case, end the sampling procedure and compute the estimates $\hat{\tau}_1$, $\hat{\tau}_2$ and $\hat{\tau}$ of τ_1 , τ_2 and τ using the same expressions as those used for the original estimates. In the negative case, repeat steps (i) to (iv). (v) Repeat steps (i) to (iv) a large number B of times. The bootstrap distributions of τ_1 , τ_2 and τ are the empirical distributions computed from the sets of B values of $\hat{\tau}_1$, $\hat{\tau}_2$ and $\hat{\tau}$, respectively. (vi) Construct the $100(1-\alpha)\%$ bootstrap confidence intervals for τ_1 , τ_2 and τ by either the basic or the percentile method (see Davison and Hinkley, 1997, Chapter 5, for descriptions of these methods), that is, in the first case, the interval for τ is $[2\hat{\tau} - \tau^{(1-\alpha/2)}, 2\hat{\tau} - \tau^{(\alpha/2)}]$, whereas in the second is $[\tau^{(\alpha/2)}, \tau^{(1-\alpha/2)}]$, where $\tau^{(\alpha/2)}$ and $\tau^{(1-\alpha/2)}$ are the $\alpha/2$ and $1-\alpha/2$ points of the bootstrap distribution of τ , and $\hat{\tau}$ is the estimate of τ obtained from the original sample.

It is worth noting that if the stopping rule depended on the precision of one of the estimators $\hat{\tau}_1$, $\hat{\tau}_2$ or $\hat{\tau}$, then in step (iv) we would need to compute that estimator to determine whether or not the stopping rule has been satisfied.

In addition to the construction of confidence intervals, simple design-based estimators of the variances of τ_1 , τ_2 and τ could be obtained by computing the sample variances from the corresponding sets of B values of the estimators.

5. MONTE CARLO STUDY

We considered two artificial finite populations that were obtained using the same parameter values as those used by Félix-Medina and Monjardin (2004). A description of each population is presented in Table 1. The nomination probabilities $p_i^{(k)}$, $k = 1, 2$, were generated using the model $p_i^{(k)} = 1 - \exp(-\beta_k m_i)$, where the values of β_k were set so that the following approximated values of $E(p_i^{(k)})$ were obtained: $(E(p_i^{(1)}), E(p_i^{(2)})) \cong (0.05, 0.03)$, and $(E(p_i^{(1)}), E(p_i^{(2)})) \cong (0.01, 0.006)$. The values of the parameters of the initial distributions were the following: $\sigma^2 = 9$, $\mu_k = -3.5$, $\gamma_k^2 = 9$, $k = 1, 2$, $a_1 = 1$, $b_1 = 0.1$, $a_2 = 8$, $b_2 = 0.01$, so that $E(\lambda_1) = 10$, $V(\lambda_1) = 100$, $E(\lambda_2) = 800$, and $V(\lambda_2) = 80000$.

Table 1. Parameters of simulated populations

Population	N	Distribution of M_i	$E(M_i)$	$V(M_i)$	τ_1	τ_2	τ	τ_1/τ
I	250	Poisson	7.2	7.2	1828	700	2528	0.72
II	250	Negative Binomial	7.2	24.4	1861	700	2561	0.73

The simulation study was performed as follows. From each population of $N = 250$ values of m_i 's, a sequential sample of values was selected using the SSRS design. The sampling procedure was stopped when the number of nominees in U_2 reached 250 or exceeded that value for first time. The numbers of nominated people in U_1 and U_2 made by the site A_i in the sample were obtained by $X_{i1}^{(1)} + \dots + X_{i\tau_1 - m_i}^{(1)}$ and $X_{i1}^{(2)} + \dots + X_{i\tau_2}^{(2)}$, respectively, where the $X_{ij}^{(1)}$'s and $X_{ij}^{(2)}$'s were samples from Bernoulli distributions with means $p_i^{(1)}$ and $p_i^{(2)}$. This procedure was repeated $r = 2000$ times. We also considered the variant of LTS proposed by Félix-Medina and Thompson (2004), that is, the one in which the initial sample of sites is selected by simple random sampling. In this case, the sizes of the initial samples were set equal to the average size of the 2000 initial sequential simple random samples.

We observed the performance of the MLE's $\tilde{\tau}_1$, $\tilde{\tau}_2$ and $\tilde{\tau}$, and that of the estimators $\hat{\tau}_1$, $\hat{\tau}_2$ and $\hat{\tau}$, derived under the Bayesian approach. The performance of an estimator $\hat{\tau}$, say, was evaluated by its relative bias and the square root of its mean square error, defined as $r\text{-bias} = \sum_1^r (\hat{\tau}_i - \tau) / (r\tau)$ and $\sqrt{r\text{-mse}} = \sqrt{\sum_1^r (\hat{\tau}_i - \tau)^2 / (r\tau^2)}$ respectively, where $\hat{\tau}_i$ was the value of $\hat{\tau}$ obtained in the i -th simulated sample.

In addition, we observed the performance of the 95% confidence intervals for the population sizes obtained by the basic bootstrap method and by the percentile bootstrap method. The performance of a confidence interval was evaluated by its average length and by its probability of coverage.

The main results of the simulation study were the following. With respect to the performance of the estimators of the population sizes (see Table 2), we have that in the case of the variant of LTS with initial SSRS, the r-biases of

the estimators of τ_2 and τ were greater than the corresponding r-biases obtained in the case of the variant with initial SRS. However, the r-biases were not large, they were less than 0.1, and did not affect seriously the performance of the estimator. In fact, in the case of the variant with initial SSRS, the $\sqrt{r - \text{mse}}$ of the estimators of τ_2 and τ were only slightly greater than those obtained in the case of the variant with initial SRS. On the other hand, the performance of each of the estimators of τ_1 was practically the same in both variants. Finally, the estimators derived under the Bayesian approach were slightly more efficient than the MLE's, and every one of the observed estimators was robust to the deviation from the assumed Poisson distribution of the site sizes.

With respect to the performance of the confidence intervals (see Table 3), the intervals obtained in the case of the variant of LTS with initial SSRS were shorter than the intervals obtained in the case of the variant with initial SRS. In addition, in the first variant, the coverage probabilities of the basic bootstrap intervals were relatively close to the nominal value, 0.95, whereas in the second variant, the coverage probabilities of the interval obtained from the estimator $\hat{\tau}_1$, derived under the Bayesian approach, were not as close to 0.95 as the coverage probabilities of the other intervals. The coverage probabilities of the percentile bootstrap intervals were not as close to 0.95 as those of the basic bootstrap intervals. In some cases the coverage probabilities of the percentile intervals were below 0.9, and in some others they were even below 0.8. Confidence intervals based on the estimators derived under the Bayesian approach performed better than those based on the MLE's. The deviation from the assumed Poisson distribution of the site sizes increased the lengths of the intervals for τ_1 and τ , and slightly moved away the coverage probabilities from the nominal value. The lengths and coverage probabilities of the intervals for τ_2 were not affected by the deviation from the Poisson distribution.

Table 2. Relative biases and square roots of relative mean square errors of the estimators of the population sizes. Results based on 2000 iterations.

Initial SSRS	Population I							Population II						
		$\tilde{\tau}_1$	$\tilde{\tau}_2$	$\tilde{\tau}$	$\hat{\tau}_1$	$\hat{\tau}_2$	$\hat{\tau}$		$\tilde{\tau}_1$	$\tilde{\tau}_2$	$\tilde{\tau}$	$\hat{\tau}_1$	$\hat{\tau}_2$	$\hat{\tau}$
$E(p_i^{(1)}) \approx .05$	$E(n)=14.56$.001	.051	.015	.001	.044	.013	$E(n)=14.47$.002	.039	.012	.001	.033	.010
$E(p_i^{(2)}) \approx .03$	$E(R_2)=263.3$.035	.129	.044	.034	.114	.040	$E(R_2)=264.2$.038	.126	.045	.038	.112	.041
$E(p_i^{(1)}) \approx .01$	$E(n)=68.96$	-.000	.088	.024	-.001	.074	.020	$E(n)=67.78$.001	.079	.023	.001	.067	.019
$E(p_i^{(2)}) \approx .006$	$E(R_2)=257.2$.022	.153	.045	.022	.131	.039	$E(R_2)=257.3$.030	.146	.046	.030	.126	.041
Initial SRS														
$E(p_i^{(1)}) \approx .05$	$E(n)=15$	-.001	.010	.002	-.001	.007	.001	$E(n)=15$	-.004	.012	.000	-.005	.008	-.001
$E(p_i^{(2)}) \approx .03$	$E(R_2)=259.7$.033	.115	.040	.033	.103	.038	$E(R_2)=263.0$.039	.118	.042	.039	.104	.040
$E(p_i^{(1)}) \approx .01$	$E(n)=69$	-.000	.016	.004	-.001	.009	.002	$E(n)=68$	-.001	.013	.003	-.002	.006	.000
$E(p_i^{(2)}) \approx .006$	$E(R_2)=241.2$.023	.131	.039	.023	.114	.035	$E(R_2)=241.9$.030	.128	.041	.030	.112	.037

First row in each cell contains the relative biases of the estimators; second row in each cell contains the square roots of the relative mean square errors of the estimators. $\tilde{\tau}_k$, maximum likelihood estimator; $\hat{\tau}_k$, estimator obtained under the Bayesian approach and Poisson prior distribution. Expected values of n and R_2 were obtained by simulation.

Table 3. Average lengths and coverage probabilities of 95% bootstrap confidence intervals for the population sizes. Results based on 2000 iterations.

Initial SSRS	Population I						Population II					
	\tilde{t}_1	\tilde{t}_2	\tilde{t}	\hat{t}_1	\hat{t}_2	\hat{t}	\tilde{t}_1	\tilde{t}_2	\tilde{t}	\hat{t}_1	\hat{t}_2	\hat{t}
$E(p_i^{(1)}) \approx .05$	$E(n)=14.56$						$E(n)=14.47$					
$E(p_i^{(2)}) \approx .03$	$E(R_2)=263.3$						$E(R_2)=264.2$					
Average length	254.9	360.8	446.1	254.9	304.8	398.1	294.7	356.7	466.1	294.7	302.0	421.2
Coverage probability	.947	.946	.961	.945	.932	.946	.968	.929	.960	.969	.915	.948
	.939	.905	.902	.939	.937	.928	.880	.920	.895	.880	.941	.912
$E(p_i^{(1)}) \approx .01$	$E(n)=68.96$						$E(n)=67.78$					
$E(p_i^{(2)}) \approx .006$	$E(R_2)=257.2$						$E(R_2)=257.3$					
Average length	178.4	402.2	442.5	177.2	322.1	368.3	250.6	397.4	471.8	248.0	319.4	404.2
Coverage probability	.949	.963	.975	.948	.948	.965	.985	.947	.971	.985	.935	.962
	.946	.762	.802	.945	.814	.864	.869	.790	.807	.867	.841	.837
Initial SRS												
$E(p_i^{(1)}) \approx .05$	$E(n)=15$						$E(n)=15$					
$E(p_i^{(2)}) \approx .03$	$E(R_2)=259.7$						$E(R_2)=263.0$					
Average length	218.2	336.7	414.5	218.2	290.7	364.9	254.9	360.8	446.1	254.9	304.8	398.1
Coverage probability	.942	.930	.959	.896	.913	.924	.968	.943	.981	.846	.916	.908
	.924	.953	.949	.889	.964	.938	.859	.966	.934	.734	.963	.871
$E(p_i^{(1)}) \approx .01$	$E(n)=69$						$E(n)=68$					
$E(p_i^{(2)}) \approx .006$	$E(R_2)=241.2$						$E(R_2)=241.9$					
Average length	162.4	359.5	396.7	135.6	306.0	335.6	222.7	360.9	426.2	151.3	304.6	341.4
Coverage probability	.925	.929	.949	.831	.905	.914	.981	.931	.963	.771	.910	.897
	.912	.943	.946	.814	.966	.943	.835	.954	.921	.598	.967	.868

Sizes of bootstrap samples: 1000. First row in each cell contains the results for the basic bootstrap intervals. Second row in each cell contains the results for the percentile intervals.

6. CONCLUSIONS

In this paper we have proposed a variant of LTS in which the initial sample is a sequential SRS of sites. This variant allows the sampler to have certain control of the final sample size, the number of nominees, the sampling cost or the precision of the estimators. The population sizes can be estimated either by the MLE's proposed by Félix-Medina and Thompson (2004) or by the Bayesian estimators proposed by Félix-Medina and Monjardin (2004). Confidence intervals for the population sizes can be constructed by using bootstrap methods.

The results of the simulation study indicate that the estimators of the population sizes used with the proposed variant of LTS perform acceptably well. However, because of the randomness of the initial sample size the estimators are slightly less efficient in this variant than in the variant with initial SRS. The estimators derived under the Bayesian approach are slightly more efficient than the MLE's. Finally, confidence intervals obtained by using the basic bootstrap method have better coverage properties than those obtained by using the percentile method.

ACKNOWLEDGEMENTS

This research was supported by grants UASIN-EXB-01-01 and PIFI-2003-25-28 of the Secretaría de Educación Pública, and a grant of the Coordinación General de Investigación y Postgrado de la Universidad Autónoma de Sinaloa.

REFERENCES

- Booth, J. G., R. W. Butler and P. Hall (1994), "Bootstrap methods for finite populations", *Journal of the American Statistical Association*, 89, pp. 1282-1289.
- Davison, A. C. and D.V. Hinkley (1997), *Bootstrap Methods and their Applications*, New York: Cambridge University Press.
- Efron, B. and R.J. Tibshirani (1993), *Introduction to the Bootstrap*, New York: Chapman & Hall.
- Félix-Medina, M. H. and S. K. Thompson (2004), "Combining Link-Tracing Sampling and Cluster Sampling to Estimate the Size of Hidden Populations", *Journal of Official Statistics*, 20, pp. 19-38.
- Félix-Medina, M. H. and P. E. Monjardin (2004), "Combining Link-Tracing Sampling and Cluster Sampling to Estimate the Size of Hidden Populations: a Bayesian Assisted Approach", paper submitted to *Survey Methodology*.
- Gross, S. (1980), "Median Estimation in Sample Surveys", *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 181-184.
- Särndal, C. E., B. Swensson and J. H. Wretman (1992), *Model Assisted Survey Sampling*, New York: Springer Verlag.