Survey Methodology, December 1991
Vol. 17, No. 2, pp. 123–130
Statistics Canada

Variance Estimation in Dual Registration Under Population Heterogeneity

JUHA M. ALHO[1]

ABSTRACT

The usual dual system estimator for population size can be severely biased, if there is population heterogeneity in the capture probabilities. In this note we investigate the bias of the corresponding variance estimator under heterogeneity. We show that the usual estimator is conservative, *i.e.*, it gives too large values, if the two registration systems are negatively correlated, uncorrelated, or when the correlation is positive, but small. In the case of high positive correlation the usual estimator may yield too low values. Two alternative estimators are proposed. One is conservative under arbitrary heterogeneity. The other is conservative under Gaussian heterogeneity. The methods are applied to occupational disease data from Finland.

KEY WORDS: Capture-recapture; Dual system; Heterogeneity; Occupational diseases.

1. INTRODUCTION

Suppose there are N individuals in a closed population. The problem is to estimate the unknown N using dual registration. We sample twice with n_j individuals captured at the jth time, $j = 1, 2$. Let m be the number captured twice. Define indicator variables u_{ji} and m_i for $i = 1, \ldots, N$ such that $u_{ji} = 1$, if and only if individual i is captured at the jth time only, $j = 1, 2$; and $m_i = 1$, if and only if individual i is captured twice. Otherwise u_{ji} and m_i are zero. Define $n_{ji} = u_{ji} + m_i$ as the indicator of capture at the jth time, $j = 1, 2$. Let $M_i = u_{1i} + u_{2i} + m_i$ indicate capture at least once. Define the individual capture probabilities as $p_{ji} = E[n_{ji}], j = 1, 2$; and $p_{12i} = E[m_i]$. Assume that the probabilities are strictly between zero and one. The fact that the probabilities are allowed to vary by individual indicates that we may have population heterogeneity in the capture probabilities. We complete the definition of the dual registration (or capture-recapture) model by assuming that the captures are independent for each individual, or $p_{12i} = p_{1i}p_{2i}$, and that the multinomial vectors

$$(u_{1i}, u_{2i}, m_i, 1 - M_i) \sim \text{Mult}(1; p_{1i}q_{2i}, p_{2i}q_{1i}, p_{1i}p_{2i}, 1 - \phi_i),$$

where $q_{ji} = 1 - p_{ji}, j = 1, 2$, and $\phi_i = p_{1i} + p_{2i} - p_{1i}p_{2i}$, are independent for $i = 1, \ldots, N$.

It is well-known that when capture probabilities do not vary by individual, or $p_{ji} = p_j$, $j = 1, 2$, the maximum likelihood estimator of N is $\hat{N} = n_1n_2/m$ (or more precisely, the largest integer short of this value; *cf.*, Feller 1968, p. 46). This classical estimator can be severely biased under population heterogeneity (Seber 1982, p. 565; Burnham and Overton 1979, Table 4, pp. 931-932). As shown, *e.g.*, in Example 1 below, under homogeneous capture probabilities the asymptotic variance of $\hat{N}$ is $\text{Var}(\hat{N}) = Nq_1q_2/(p_1p_2)$, where $q_j = 1 - p_j$, $j = 1, 2$. Then $\text{Var}(\hat{N})$ can be estimated by $V_1 = n_1n_2u_1u_2/m^3$ (Sekar and Deming 1949, pp. 114-115).

[1] Juha M. Alho, Institute for Environmental Studies and Department of Statistics, University of Illinois at Urbana-Champaign, 1101 W. Peabody Dr., Urbana IL, 61801, U.S.A.

The purpose of this note is to investigate the adequacy of the variance estimator V_1, and compare the bias of V_1 to the bias of $\hat{N}$. One motive for investigating V_1 is that it has not been previously known whether V_1 is adequate in the case in which there is population heterogeneity, but $\hat{N}$ is, nevertheless, consistent. This turns out to be the case. Similarly, it has not been clear when V_1 gives overestimates and thus can lead to valid confidence intervals, despite the bias of $\hat{N}$. This turns out to be possible for one-sided intervals in special circumstances.

In Section 2 we calculate the asymptotic variance of $\hat{N}$, as $N \to \infty$, and derive a conservative estimator V_2 for this variance under arbitrary heterogeneity. In other words, V_2 overestimates the true asymptotic variance. One might hope that an overestimate of variance could compensate for the typically negative bias of $\hat{N}$ and still yield valid confidence intervals. Unfortunately, this appears possible only when the bias of $\hat{N}$ is small, or when N is small. In Section 3 the adequacy of V_1 is studied under Gaussian heterogeneity and an estimator V_3 is derived, which is conservative under this restricted type of heterogeneity. Gaussianity *per se* is not required for the arguments, only that the moments of the pairs (p_{1i}, p_{2i}) agree with those of a bivariate Gaussian distribution. This setup permits the ready examination of the effect of correlation between p_{1i}'s and p_{2i}'s on variance estimation, because correlation is expressible in terms of just one parameter, the ordinary moment correlation coefficient. In Section 4 we compare the bias in variance estimates to the bias of $\hat{N}$ using empirical data relating to the registration of occupational diseases in Finland.

2. BIAS AND VARIANCE UNDER HETEROGENEITY

Define $\bar{p}_{jN}$ as the average probability of capture at the jth time, $j = 1, 2$; and let $\bar{p}_{12N}$ be the average of the products $p_{1i}p_{2i}$, $i = 1, \ldots, N$. Then, $C_N = \bar{p}_{12N} - \bar{p}_{1N}\bar{p}_{2N}$ is the covariance of the pairs (p_{1i}, p_{2i}). Assume that the limits $\bar{p}_{jN} \to \bar{p}_j$, $j = 1, 2$; $\bar{p}_{12N} \to \bar{p}_{12}$, and $C_N \to C$ exist. Then we have that $\hat{N}/N \to \bar{p}_1\bar{p}_2/\bar{p}_{12}$, so $\hat{N}/N - 1 \to -C/\bar{p}_{12}$, as $N \to \infty$. This is the asymptotic bias of the classical estimator under population heterogeneity. Interestingly, it only depends on the first two moments of the distribution of the pairs (p_{1i}, p_{2i}). As is well-known (Sekar and Deming 1949, pp. 105-106; Seber 1982, p. 86), when the covariance is zero $(C = 0)$, then the classical estimator is consistent; if $C > 0$, $\hat{N}$ gives an underestimate; and if $C < 0$, it gives an overestimate. As noted above the adequacy of V_1, when the p_{ji}'s vary from one individual to the next but still $C = 0$, is of particular interest.

We shall now calculate the asymptotic variance of the classical estimator under our general heterogeneity model. Note that the finite variance does not exist, because there is a positive probability that $m = 0$. Therefore, "asymptotic variance" properly refers here to the variance of the limiting distribution rather than to limit of the variances, as $N \to \infty$.

Lemma 1. The asymptotic variance of $\hat{N}$ is

$$\operatorname{Var}(\hat{N}) = N\left\{\frac{\bar{p}_1^2\bar{p}_2^2}{\bar{p}_{12}^3} - \frac{\bar{p}_1^2\bar{p}_2}{\bar{p}_{12}^2} - \frac{\bar{p}_1\bar{p}_2^2}{\bar{p}_{12}^2} - \frac{\bar{p}_2^2}{\bar{p}_{12}^2}\bar{S}_1 - \frac{\bar{p}_1^2}{\bar{p}_{12}^2}\bar{S}_2 - \frac{\bar{p}_1^2\bar{p}_2^2}{\bar{p}_{12}^4}\bar{S}_3 + 2\left(\frac{\bar{p}_1\bar{p}_2^2}{\bar{p}_{12}^3}\bar{S}_4 + \frac{\bar{p}_1^2\bar{p}_2}{\bar{p}_{12}^3}\bar{S}_5\right)\right\},$$

where $\bar{S}_j = S_j/N$ for $j = 1, \ldots, 5$, with

$$S_1 = \sum_{i=1}^{N} p_{1i}^2, \quad S_2 = \sum_{i=1}^{N} p_{2i}^2, \quad S_3 = \sum_{i=1}^{N} p_{1i}^2 p_{2i}^2,$$

$$S_4 = \sum_{i=1}^{N} p_{1i}^2 p_{2i}, \quad S_5 = \sum_{i=1}^{N} p_{1i} p_{2i}^2.$$

The proof is sketched in the Appendix. We note that unlike the bias of $\hat{N}$ that depends on the first two moments of the pairs (p_{1i}, p_{2i}) only, $\mathrm{Var}(\hat{N})$ depends on moments up to fourth order. In special cases, such as the ones considered in Example 2 and Proposition 2, a simpler representation is possible.

Example 1. Suppose there is no heterogeneity in the probabilities, or $p_{ji} = p_j, j = 1, 2$. Then $\bar{p}_j = p_j, j = 1, 2$; $\bar{p}_{12} = p_1 p_2$; $\bar{S}_j = p_j^2, j = 1,2$; $\bar{S}_3 = p_1^2 p_2^2$, $\bar{S}_4 = p_1^2 p_2$, and $\bar{S}_5 = p_1 p_2^2$. Hence, the asymptotic variance is $\mathrm{Var}(\hat{N}) = N(1 - p_1 - p_2 + p_1 p_2)/(p_1 p_2) = Nq_1 q_2/(p_1 p_2)$. Consistent estimators for $Np_1 p_2$ and Np_j are m and $n_j, j = 1, 2$. In other words, $Np_j/n_j \rightarrow 1, j = 1, 2$, and $Np_1 p_2/m \rightarrow 1$, as $N \rightarrow \infty$. This gives us V_1 as an estimator for $\mathrm{Var}(\hat{N})$.

Example 2. Suppose that the pairs (p_{1i}, p_{2i}), $i = 1, \ldots, N$, are independent in the sense that the distribution of p_{1i}'s is the same for each distinct value of the p_{2i}'s. Then, $\bar{p}_{12} = \bar{p}_1 \bar{p}_2$, $\bar{S}_3 = \bar{S}_1 \bar{S}_2$, $\bar{S}_4 = \bar{p}_2 \bar{S}_1$, $\bar{S}_5 = \bar{p}_1 \bar{S}_2$. Substituting into the Lemma we get

$$\begin{aligned}\mathrm{Var}(\hat{N}) &= N\left(\frac{1}{\bar{p}_1 \bar{p}_2} - \frac{1}{\bar{p}_2} - \frac{1}{\bar{p}_1} - \frac{\bar{S}_1 \bar{S}_2}{\bar{p}_1^2 \bar{p}_2^2} + \frac{\bar{S}_1}{\bar{p}_1^2} + \frac{\bar{S}_2}{\bar{p}_2^2}\right) \\ &= N\left(\frac{\bar{q}_1 \bar{q}_2}{\bar{p}_1 \bar{p}_2} - cv(p_{1i})^2 \, cv(p_{2i})^2\right),\end{aligned}$$

where $cv(p_{ji}) = (\bar{S}_j - \bar{p}_j^2)/\bar{p}_j$, is the coefficient of variation of the p_{ji}'s, $j = 1, 2$. Obviously, $\mathrm{Var}(\hat{N}) \leq N\bar{q}_1 \bar{q}_2/(\bar{p}_1 \bar{p}_2)$. A comparison with Example 1 shows that V_1 is a conservative estimator of $\mathrm{Var}(\hat{N})$ (*i.e.*, V_1 is asymptotically too large), when p_{1i}'s are independent of p_{2i}'s. Another way of saying this is that, given the means $\bar{p}_j, j = 1, 2$, the largest value of the variance is obtained at homogeneity. This is analogous to the variance of the number of successes in Bernoulli trials with variable probabilities of success, *cf.* Feller 1968, pp. 230-231. A comparison with Example 1 shows that V_1 is a conservative estimator of $\mathrm{Var}(\hat{N})$ (*i.e.*, V_1 is asymptotically too large), when the pairs (p_{1i}, p_{2i}) are independent. Note that the independence condition implies that $C = 0$.

When the probabilities are not independent, the classical variance estimator is not guaranteed to be conservative. A conservative estimator exists, however. It is obtained by majorizing $\mathrm{Var}(\hat{N})$ by a quantity that can be estimated in terms of the observable variables. We prove in the Appendix the following general proposition.

Proposition 1. A conservative estimator of $\mathrm{Var}(\hat{N})$ is

$$V_2 = (n_1^2 n_2^2 + n_2^2 m u_1 + n_1^2 m u_2)/m^3,$$

where $u_j = n_j - m, j = 1, 2$.

3. GAUSSIAN HETEROGENEITY

We shall now turn to a special case in which the sample moments of the pairs (p_{1i}, p_{2i}), $i = 1, \ldots, N$, agree with those of a bivariate normal, or Gaussian, distribution. This will permit a much sharper specification of a conservative variance estimator than the one obtained in the general case above. Assume that

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} v_1^2\mu_1^2 & \rho v_1 v_2 \mu_1 \mu_2 \\ \rho v_1 v_2 \mu_1 \mu_2 & v_2^2 \mu_1^2 \end{bmatrix}\right),$$

where $|\rho| < 1$, and $0 < \mu_j < 1, j = 1, 2$. Note that v_j's can be interpreted as the coefficients of variation of the distributions of p_{ji}'s. Write $\bar{S}_j = S_j/N$ for $j = 1, \ldots, 5$, as before. Then substitute the moments of the bivariate normal distribution into Lemma 1 as follows,

$$\bar{p}_j = E[X_j] = \mu_j, j = 1, 2;$$

$$\bar{S}_j = E[X_j^2] = \mu_j^2(1 + v_j^2), j = 1, 2;$$

$$\bar{p}_{12} = E[X_1 X_2] = \mu_1\mu_2(1 + \rho v_1 v_2);$$

$$\bar{S}_3 = E[X_1^2 X_2^2] = \mu_1^2\mu_2^2(1 + v_1^2 + v_2^2 + 4\rho v_1 v_2 + (2\rho^2 + 1)v_1^2 v_2^2);$$

$$\bar{S}_4 = E[X_1^2 X_2] = \mu_1^2\mu_2(1 + 2\rho v_1 v_2 + v_1^2);$$

$$\bar{S}_5 = E[X_1 X_2^2] = \mu_1\mu_2^2(1 + 2\rho v_1 v_2 + v_2^2).$$

Straightforward, but slightly tedious calculations prove then the following proposition (details omitted).

Proposition 2. With the above assumptions

$$\mathrm{Var}(\hat{N}) = A_1 A_2 + RN,$$

where

$$A_1 = N/(1 + \rho v_1 v_2)^2;$$

$$A_2 = [1 - (\mu_1 + \mu_2)(1 + \rho v_1 v_2 + \mu_1\mu_2(1 + \rho v_1 v_2)^2]/[\mu_1\mu_2(1 + \rho v_1 v_2)^2];$$

$$R = (2\rho v_1 v_2 + 3\rho^2 v_1^2 v_2^2 - \rho^2 v_1^4 v_2^2 - \rho^2 v_1^2 v_2^4 - v_1^2 v_2^2)/(1 + \rho v_1 v_2)^4.$$

We can evaluate the classical variance estimator $V_1 = n_1 n_2 u_1 u_2 / m^3$ using this result. Note first that $\{n_1 n_2/m\}/A_1 \rightarrow 1$, as $N \rightarrow \infty$. Similarly, $\{u_1 u_2/m^2\}/A_2 \rightarrow 1$. This proves the following corollary to Proposition 2: $(V_1 - \mathrm{Var}(\hat{N}))/N \rightarrow -R$, as $N \rightarrow \infty$. For example, if $\rho = 0$, then $-R = v_1^2 v_2^2$, so that V_1 is seen to overestimate the asymptotic variance. This is in accordance with Example 2.

How reasonable is the assumption of Gaussian moments? Certainly the capture probabilities cannot have strictly Gaussian distributions, because the Gaussian distribution always puts some probability mass outside the unit interval. On the other hand, suppose we generate the p_{ji}'s by taking $\text{logit}(p_{ji}) = a_j + b_j Y_{ji}$, where the pairs (Y_{1i}, Y_{2i}) are a sample from a bivariate normal distribution with mean zero, unit variances, and correlation ρ. If we have the relations $a_j = \text{logit}(\mu_j), j = 1, 2$, and $b_j = \nu_j(1 + \mu_j)^2$, then the assumption of Gaussian moments is approximately true. In fact, even the distribution of the pairs (p_{1i}, p_{2i}) is in that case approximately bivariate Gaussian.

Let us consider the adequacy of V_1 further, under the Gaussian moments. The fact that probabilities are constrained to be between 0 and 1 means that μ_j's are between zero and one. Moreover, to be sure that most of the probability mass is in the unit square, let us assume that $0 < \nu_j \le 1/2, j = 1, 2$. If μ_j's are close to one, a much smaller upper bound would be needed. Assume now that $\rho \le 0$. Then, one can show that

$$-R \ge (\rho^2 \nu_1^4 \nu_2^2 + \rho^2 \nu_1^2 \nu_2^4)/(1 + \rho \nu_1 \nu_2)^4 > 0,$$

so that V_1 overestimates $\text{Var}(\hat{N})$ for $\rho \le 0$ also. Note that by continuity V_1 must overestimate $\text{Var}(\hat{N})$ for some positive values of ρ, as well.

One can show that $R = R(\rho)$ is an increasing function of ρ for at least $\rho > 0$. In the limit we have

$$-R(\rho) \to (-2\nu_1\nu_2 - 2\nu_1^2\nu_2^2 + \nu_1^4\nu_2^2 + \nu_1^2\nu_2^4)/(1 + \nu_1\nu_2)^4,$$

as $\rho \to 1$. When $0 < \nu_j \le 1/2, j = 1, 2$, the smallest value of the above limit occurs at $\nu_1 = \nu_2 = 1/2$. The minimum value is $-152/625 > -1/4$. Consequently, for $\rho > 0$, V_1 can either underestimate or overestimate $\text{Var}(\hat{N})$.

The practical implications of the above results are as follows. First, if $\rho \le 0$, then $\hat{N}$ is either consistent or it overrestimates N and V_1 gives an overestimate of the variance, so we can calculate a conservative upper confidence limit for N. When $\rho > 0$, $\hat{N}$ gives an underestimate of N. If, in addition, ρ is small, then V_1 gives an overestimate, and we can get a conservative lower confidence limit for N. Obviously, these are rather special circumstances that one would not expect to be of wide practical utility.

Under the present model the asymptotic bias of V_1 is $> -N/4$ for all values of ρ. We can derive a conservative variance estimator by noting that in the Gaussian case the asymptotic relative bias of $\hat{N}$ is $-\rho\nu_1\nu_2/(1 + \rho\nu_1\nu_2) \ge -1/5$. Hence, asymptotically $5\hat{N}/4 \ge N$. A conservative estimator of $\text{Var}(\hat{N})$ is, for example, $V_3 = V_1 + 5\hat{N}/16$. This can be much smaller than V_2 indicating that the Gaussian assumption is a very powerful one.

4. AN APPLICATION TO OCCUPATIONAL DISEASE REGISTRATION DATA

To get an idea of how large the biases may be in practice, let us look at occupational disease data from Finland as an example. The Finnish Register of Occupational Diseases has been in operation since 1964. It is kept by the Institute of Occupational Health in Helsinki. Since 1975 the number of new cases reported to the Register has varied from about 4,000 to over 7,000 annually (0.2 - 0.4 % of the employed population). Noise-induced hearing loss, diseases caused by repetitive of monotonous work (epicondylitis, bursitis, tendinivaginitis), and skin diseases are the major diagnostic groups (*cf.* Vaaranen *et al.* 1985). The Register can be viewed as a dual registration system, because each case of disease should, under existing regulations, be reported to the Register both from the appropriate insurance company and the examining physician.

It is likely that the probability of reporting a case depends on diagnosis, for example. Indeed, based on data from the year 1981 we get the following statistics. Reports from the insurance companies, $n_1 = 3{,}769$; reports from the physicians, $n_2 = 3{,}053$; and cases reported from both sources, $m = 1{,}591$. Thus the usual dual registration estimate is $\hat{N} = 7{,}232$ with $V_1^{1/2} = 97$, $V_2^{1/2} = 222$, and $V_3^{1/2} = 108.0$. The closeness of V_3 to V_1 is striking. Stratifying the data into four categories by diagnosis (the three diagnostic groups mentioned above, and the remaining "other" category) yields the following estimates. Noise-induced hearing loss: $\hat{N} = 2{,}230$, $V_1^{1/2} = 33.4$, $V_2^{1/2} = 47.2$, and $V_3^{1/2} = 42.6$; diseases caused by repetitive or monotonous work: $\hat{N} = 3{,}572$, $V_1^{1/2} = 201.4$, $V_2^{1/2} = 303.8$, and $V_3^{1/2} = 204.2$; skin diseases: $\hat{N} = 1{,}441$, $V_1^{1/2} = 30.9$, $V_2^{1/2} = 86.2$, and $V_3^{1/2} = 37.5$; other diseases $\hat{N} = 1{,}015$, $V_1^{1/2} = 32.7$, $V_2^{1/2} = 79.1$, and $V_3^{1/2} = 37.2$. Adding the results yields the following estimates for the total number of diseases: $\hat{N} = 8{,}258$, $V_1^{1/2} = 209.0$, $V_2^{1/2} = 340.3$, and $V_3^{1/2} = 215.2$. We see that diseases caused by repetitive or monotonous work are underreported to a particularly great extent.

The analysis was extended further by stratifying the data by diagnosis (4 categories), insurance company (11 categories), and main groups of industry (7 categories). *A priori*, these factors could be thought to have an influence on reporting probabilities. However, the stratification did not alter the point estimate materially. It did increase the estimated standard deviations by over a third, apparently because some of the strata became very small. We conclude that the bias in the point estimator caused by diagnosis is the dominant source of error in the classical estimator in this application.

The same data were further analyzed using a logistic regression technique that allows us to take into account observable population heterogeneity due to both discrete and continuous explanatory variables. In this application age was shown to have an effect on reporting probabilities within the diagnostic groups for one source of information, but not for the other. Therefore, the point estimates remained unchanged and the conclusion regarding the role of diagnosis could not be refuted (Alho 1990).

5. DISCUSSION

Our theoretical results indicate that the usual variance estimator V_1 is conservative when the two registration systems are negatively correlated or independent. By continuity the estimator may be conservative also when the correlation is positive but small. Under high positive correlation V_1 gives too low values. We introduced an alternative estimator V_2, which is conservative under arbitrary population heterogeneity. However, it appears to be unduly conservative in view of the numerical comparisons with V_3, which is guaranteed to be conservative under Gaussian heterogeneity. The closeness of V_3 to V_1 suggests that, in practice, V_1 may be fairly robust against population heterogeneity.

Unfortunately, even the use of the conservative estimator V_2 would not have been sufficient to cover the bias in the classical point estimator in our empirical example. Perhaps this was to be expected, since the bias of $\hat{N}$ and the degree of overestimation provided by V_2 are both of order N. Hence, the use of V_2 inflates the width of a confidence interval by a factor of order $N^{1/2}$ only. Therefore, V_2 can compensate for the bias of $\hat{N}$, if the bias is small, or if N itself is small. Hence, it seems that the successfull application of the dual registration method requires that either we have roughly uncorrelated registration systems, or that the heterogeneity is observable. In the latter case we may use stratification as suggested already by Sekar and Deming (1949), or logistic regression modeling as suggested by Huggins (1989) and Alho (1990), to adjust for the bias of the classical estimator of population size.

ACKNOWLEDGEMENT

The author would like to thank Bruce Spencer and an anonymous referee for comments that helped to improve the presentation. Part of the empirical results were first presented at the 11th Nordic Conference on Mathematical Statistics in Uppsala, Sweden, in June 1986.

APPENDIX

Proof of Lemma 1. Apply a linear Taylor-series development to $\hat{N} = n_1 n_2 / m$ at $E[n_1] E[n_2]/E[m] = N\bar{p}_1\bar{p}_2/\bar{p}_{12}$, or

$$\hat{N} \approx \frac{N\bar{p}_1\bar{p}_2}{\bar{p}_{12}} + \frac{\bar{p}_2}{\bar{p}_{12}}(n_1 - N\bar{p}_1) + \frac{\bar{p}_1}{\bar{p}_{12}}(n_2 - N\bar{p}_2) - \frac{\bar{p}_1\bar{p}_2}{\bar{p}_{12}^2}(m - N\bar{p}_{12}).$$

Hence, we have

$$E\left[\left(\hat{N} - \frac{N\bar{p}_1\bar{p}_2}{\bar{p}_{12}}\right)^2\right] \approx \left(\frac{\bar{p}_2}{\bar{p}_{12}}\right)^2 \mathrm{Var}(n_1) + \left(\frac{\bar{p}_1}{\bar{p}_{12}}\right)^2 \mathrm{Var}(n_2) + \left(\frac{\bar{p}_1\bar{p}_2}{\bar{p}_{12}^2}\right)^2 \mathrm{Var}(m)$$

$$- \frac{\bar{p}_1\bar{p}_2^2}{\bar{p}_{12}} \mathrm{Cov}(n_1, m) - 2\frac{\bar{p}_1^2\bar{p}_2}{\bar{p}_{12}^3} \mathrm{Cov}(n_2, m).$$

Under our independence assumptions $\mathrm{Var}(n_j) = N\bar{p}_j - S_j, j = 1, 2$; $\mathrm{Var}(m) = N\bar{p}_{12} - S_3$, $\mathrm{Cov}(n_1, m) = -S_4 + N\bar{p}_{12}$, and $\mathrm{Cov}(n_2, m) = -S_5 + N\bar{p}_{12}$. Substituting these into the mean squared error gives the result.

Proof of Proposition 1. We ignore the negative term containing S_3 in Lemma 1. Since $0 < p_{ji} < 1$, we have $S_4 < N\bar{p}_{12}$, and $S_4 < S_1$. Therefore,

$$\frac{2\bar{p}_1\bar{p}_2^2}{\bar{p}_{12}^3} S_4 < \frac{\bar{p}_1\bar{p}_2^2}{\bar{p}_{12}^3} N\bar{p}_{12} + \frac{\bar{p}_2^2}{\bar{p}_{12}^2} S_1 + \left(\frac{\bar{p}_1 - \bar{p}_{12}}{\bar{p}_{12}}\right) \frac{\bar{p}_2^2}{\bar{p}_{12}^2} N\bar{p}_{12}.$$

Similarly,

$$\frac{2\bar{p}_1^2\bar{p}_2}{\bar{p}_{12}} S_5 < \frac{\bar{p}_1^2\bar{p}_2}{\bar{p}_{12}^3} N\bar{p}_{12} + \frac{\bar{p}_1^2}{\bar{p}_{12}^2} S_2 + \left(\frac{\bar{p}_2 - \bar{p}_{12}}{\bar{p}_{12}}\right) \frac{\bar{p}_1^2}{\bar{p}_{12}^2} N\bar{p}_{12}.$$

Substituting these bounds to the expression of Lemma 1 we get

$$\mathrm{Var}(\hat{N}) < \frac{\bar{p}_1^2\bar{p}_2^2}{\bar{p}_{12}^3} N + \frac{(\bar{p}_1 - \bar{p}_{12})\bar{p}_2^2}{\bar{p}_{12}^2} N + \frac{(\bar{p}_2 - \bar{p}_{12})\bar{p}_1^2}{\bar{p}_{12}^2} N.$$

Estimating $N\bar{p}_j$ by n_j, $j = 1, 2$; and $N\bar{p}_{12}$ by m we get the result.

REFERENCES

ALHO, J. (1990). Logistic regression in capture-recapture models. *Biometrics*, 46, 623-635.

BURNHAM, P.K., and OVERTON, W.S. (1979). Robust estimation of population size when capture probabilities vary among animals. *Ecology*, 60, 927-936.

FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications*, (Vol. I, 3rd ed.). New York: Wiley.

HUGGINS, R.M. (1989). On the statistical analysis of capture experiments. *Biometrika*, 76, 133-140.

SEBER, G.A.F. (1982). *The Estimation of Animal Abundance*, (2nd ed.). New York: Griffin.

SEKAR, C.C., and DEMING, W.E. (1949). On a method of estimating birth and death rates and the extent of registration. *Journal of the American Statistical Association*, 44, 101-115.

VAARANEN, V., VASAMA, M., and ALHO, J. (1985). Occupational diseases in Finland in 1984. Reviews 11, Institute of Occupational Health, Helsinki.