Evaluation of a genetic method for estimating contemporary population size in cetaceans based on linkage disequilibrium

ROBIN S. WAPLES

NOAA Fisheries, Northwest Fisheries Science Center, 2725 Montlake Blvd. East, Seattle, WA 98117 USA
Robin.waples@noaa.gov

ABSTRACT

In all natural populations, random linkage disequilibrium (LD; non-random associations of alleles at different gene loci) occurs, with magnitude inversely proportional to effective population size, N_e. By measuring LD from a population sample, one can estimate N_e and (if the ratio N_e/N is known or can be estimated) the census population size, N. In contrast to the more commonly used temporal method, the LD method requires only a single sample. However, the LD method has seen relatively few applications, and its performance has not been evaluated with highly variable genetic markers. I use simulated data to evaluate bias and precision of the LD method as a function true N_e, the number of individuals sampled, and the number and frequencies of alleles included in the analysis. Results show that using a typical number of microsatellite markers, the LD method can estimate N_e very precisely in small populations and can provide useful information about the lower bound of population size for large populations. When true population size is large ($N_e > 1000$), estimates of effective size are largely unbiased if sample sizes are large enough (100 or more individuals), but precision is low, particularly for the upper bound of N_e. Restricting the analyses to alleles with frequency ≥ 0.02 appears to provide a good balance between maximizing precision and minimizing bias.

INTRODUCTION

Population size can be difficult to estimate in many natural populations, including cetaceans. This difficulty has motivated efforts to explore genetic methods for estimating abundance. Most genetic methods provide direct information not about the census population size (N) but rather the effective population size (N_e). Generally, $N_e < N$ in natural populations, and if the ratio N_e/N is known or can be estimated, a genetic estimate of N_e can provide insights into population abundance as well. Because N_e directly or indirectly determines the rates of change in response to the evolutionary processes genetic drift, migration, selection, and mutation, there is also considerable interest in estimates of N_e for their own sake. This paper evaluates the practical usefulness of one genetic method for estimating N_e (and hence N), based on linkage disequilibrium (LD; non-random associations of alleles at different gene loci).

The LD method has some advantages over other methods for estimating N_e. First, unlike methods that estimate long-term effective size (e.g., Roman and Palumbi 2003), the LD method does not require equilibrium assumptions and provides an estimate of contemporary N_e, which often is of more interest for management applications. Second, unlike the more commonly used temporal method (Nei and Tajima 1981; Waples 1989; Wang 2001), which requires at least two samples separated in time, the LD method requires only a single sample. Third, because the number of allelic combinations the estimate is based upon increases with the square of the numbers of loci and alleles per locus, precision of the method (in theory at least) increases quickly with addition of more and more genetic markers. However, the LD method has seen relatively few practical applications, and its performance has not been evaluated with highly polymorphic genetic markers (such as microsatellites) that are commonly used today.

In this paper I examine precision and bias of the LD method and how these indices vary as a function of sample size, effective population size, and the number and frequencies of alleles used. I also consider some questions of more direct relevance to management, such as, "If N_e is small, how often does the method mistakenly estimate a large N_e?" and "If N_e is large, how often does the method mistakenly estimate a small N_e?"

METHODS

Genotypic data were generated for populations of known effective size using the software EASYPOP (Balloux 2001). The populations had discrete generations, equal sex ratio, and mated randomly. One thousand replicate populations were generated for each population size considered (N = 100, 500, 1000, 5000). Each simulated

individual had data for $L = 20$ independent gene loci which had a mutational model approximating that of microsatellites (mutation rate $\mu = 5\text{x}10^{-4}$; *k*-allele model (KAM) with 10 possible allelic states). Each simulation was initiated with maximal diversity (initial genotypes randomly drawn from all possible allelic states) and run for successive generations until the mean within-population heterozygosity (H_e) reached 0.8 (comparable to levels found in many studies of natural populations using microsatellites). Simulations with $N = 5000$ used a lower mutation rate ($\mu = 5\text{x}10^{-5}$) because $\mu = 5\text{x}10^{-4}$ led to mutation-drift equilibrium values of H_e that were larger than 0.8; in addition, because of memory and time constraints, simulations with $N = 5000$ were evaluated over 500 rather than 1000 replicate populations. After the $H_e = 0.8$ criterion was met, samples of $S = 200$, 100, or 50 individuals were taken from the final generation and subjected to genetic analysis. [For simulations with $N = 100$, the sample size of $S = 200$ was skipped.] Since the populations were "ideal," effective size and census size were the same (actually, for otherwise ideal populations in species with separate sexes, $N_e \approx N + 0.5$; Balloux 2004).

Effective population size was estimated from the genetic data via the linkage disequilibrium method (Hill 1981), which depends on random associations of alleles at different gene loci that occur in all finite populations. The squared correlation of allele frequencies at pairs of loci ($\hat{r}^2$) was calculated using the composite Burrows method (Weir 1996). For each sample, an overall mean $\bar{\hat{r}}^2$ was computed as the weighted average $\hat{r}^2$ over the $L(L\text{-}1)/2$ pairwise comparisons among loci. With $L = 20$ loci initially segregating and a high mutation rate, virtually every replicate had 20 polymorphic loci at the time of sampling, yielding 20x19/2 = 190 pairwise comparisons of loci. The weights for each locus pair were the relative number of independent alleles used in the comparison. A locus with K alleles has the equivalent of K-1 independent alleles. For two loci, with K_1 and K_2 alleles, respectively, there are the equivalent of $(K_1 - 1)(K_2 - 1)$ (presumably) independent allelic comparisons. For each locus pair, therefore, I computed a mean $\bar{\hat{r}}^2$ as the average over all K_1K_2 allelic comparisons and gave that $\bar{\hat{r}}^2$ a relative weight equal to $(K_1 - 1)(K_2 - 1)$. The total degrees of freedom associated with the overall weighted mean $\bar{\hat{r}}^2$ was computed as $n = \Sigma(K_i - 1)(K_j - 1)$, with the sum over all pairwise comparisons of loci. To assess the possible biases from use of numerous low frequency alleles (as commonly occur with microsatellites), $\bar{\hat{r}}^2$ was computed separately after excluding alleles with frequencies below the following criteria: Critical P = 0.1; 0.05; 0.02; 0.01.

Because Hill's (1981) method was recently shown to be biased (England et al. 2006), the empirical bias correction developed by Waples (2006; Table 2, random mating model) was used to estimate effective size from the overall mean $\bar{\hat{r}}^2$. Parametric confidence intervals to $\hat{N}_e$ were computed using Equation 12 in Waples (2006) assuming n degrees of freedom. Accuracy was evaluated by using the overall mean $\bar{\hat{r}}^2$ to compute a mean $\hat{N}_e$ for each parameter set and comparing this with the true N_e; bias in $\hat{N}_e$ was computed as $\hat{N}_e$ – true N_e and expressed as the ratio $\hat{N}_e / N_e$. Precision was measured by the coefficient of variation (CV) of $\bar{\hat{r}}^2$ and by confidence intervals (CIs) for $\hat{N}_e$. A metric that measures both accuracy and precision (mean squared error (MSE) = mean squared deviation of the point estimates $\hat{N}_e$ from the true value) was computed as MSE($\hat{N}_e$) = Var($\hat{N}_e$) + Bias($\hat{N}_e$)^2). Because $\hat{N}_e$ can be infinite or negative (which occurs if the empirical $\bar{\hat{r}}^2$ is equal to or less than, respectively, the value expected from sampling error alone), a limit of 10x true N_e was imposed on $\hat{N}_e$. A similar approach is used with maximum likelihood methods for estimating N_e from genetic data, which require that an upper bound for $\hat{N}_e$ be stipulated (Wang 2005).

RESULTS

With four different population sizes, three different sample sizes (two only for $N = 100$) and four different critical P values, results were obtained for 44 different scenarios for each metric evaluated. These results are arranged below by population size and plotted as a function of sample size and critical P value. The results are further organized into three categories: metrics that measure bias of $\hat{N}_e$; those that measure precision (CV of $\bar{\hat{r}}^2$; CIs for $\hat{N}_e$); and a metric that measures both accuracy and precision (MSE($\hat{N}_e$)).

Bias

Two patterns are noteworthy with respect to the ratio $\hat{N}_e / N_e$, which is shown in the top panel of Figures 1-4. First, estimates using Critical $P = 0.1$ and 0.05 consistently produced estimates of N_e with little or no bias, whereas with the less stringent Critical $P = 0.02$ or 0.01 low frequency alleles resulted in an upward bias in $\hat{N}_e$. Second, there is an interaction effect between N, S, and Critical P with respect to bias. As alleles of frequency < 0.05 are included, bias increases faster for smaller samples. This interaction effect is most clearly seen with large populations ($N = 5000$). With large samples ($S = 200$), $\hat{N}_e$ has only a minor upward bias (< 7%) for Critical $P = 0.05$ and only slightly more for Critical $P = 0.02$ or 0.01; in contrast, with $S = 50$ $\hat{N}_e$ has 13% upward bias for Critical $P = 0.05$, 59% bias for Critical $P = 0.02$, and almost 300% upward bias when alleles at frequency as low as 0.01 are included in the analysis (Figure 4, top panel).

Precision

Precision is directly related to the amount of information available for computing $\bar{\hat{r}}^2$. Table 1 shows the mean degrees of freedom (n) associated with each parameter set. Relatively little variation in n occurred among replicates within a parameter set (data not shown). In addition, for a given N and Critical P, sample size had little effect on mean n, except for Critical $P = 0.01$, where larger samples had slightly more alleles that met the criterion (Table 1). In contrast, n increased sharply as lower frequency alleles were admitted into the computations. For example, n was more than twice as high for Critical $P = 0.05$ than for 0.1 and was about three times as high for Critical $P = 0.02$. For $S = 50$ use of Critical $P = 0.01$ provided little additional information, but for larger samples (which contain more rare alleles) use of Critical $P = 0.01$ led to further (albeit relatively modest) increases in n (Table 1).

The direct effects on precision when lower frequency alleles are used are seen in the second panel in Figures 1-4. For all values of N, the CV of $\bar{\hat{r}}^2$ is highest for Critical $P = 0.1$, drops by about 40-50% for Critical $P = 0.05$, and declines further (but more modestly) for Critical $P = 0.02$ and 0.01. This effect is essentially independent of sample size.

Parametric 95% CIs for the overall $\hat{N}_e$ were calculated using the presumed degrees of freedom shown in Table 1. To assess their accuracy, these CIs were compared to 2.5% and 97.5% tails of the empirical distribution of $\hat{N}_e$ estimates across the 1000 (500 in the case of $N = 5000$) replicates. In all cases the empirical CIs were broader than the parametric CIs, indicating that the latter are slightly optimistic. Results for $N = 500$ depicted in Figure 5 illustrate this pattern. For example, with $S = 200$, overall $\hat{N}_e$ was 504.7 (less than 1% upward bias) and the parametric 95% CI was (388-716), whereas in the simulations the empirical range was somewhat broader (358-833). This figure also dramatically shows the decline in precision with smaller sample size. With $S = 50$, the parametric and empirical 95% CIs were (239-∞) and (208-∞), respectively. Regardless of the Critical P, with samples of 50 individuals the LD method could not reliably distinguish a population with true effective size = 500 from one with an infinite effective size. As a rough general rule, the empirical CIs could be approximated by subtracting 10% from the lower bound of the parametric CI and adding 10% (or a bit more) to the upper bound.

Another way to evaluate precision in a form that is relevant to management considerations is to ask, How often do we obtain an estimate of N_e that is substantially larger (or smaller) than the true value? Two pairs of criteria were considered: values that were twice (or ten times) as large as true N_e, and values that were one-half (or one-tenth) the true N_e. Table 2 makes apparent the asymmetry in the power of the LD method: small effective size can be detected much more reliably than can large N_e. For example, with $N = 100$ and $S = 100$, in not a single replicate did $\hat{N}_e$ exceed $2N_e$ or fall below $0.5N_e$ (Table 2). For Critical $P = 0.05$ or lower, all of the estimates fell within the range $\hat{N}_e$ = 69-160 (data not shown). Even with small samples ($S = 50$), no estimates were 10x (or 1/10th) the true value and only a small fraction were twice (or half) as large as the true N_e (Table 2).

As N increases, the fraction of $\hat{N}_e$ estimates that are outside these bounds increases rapidly, and the magnitude of the increase depends heavily on S and Critical P. Furthermore, more estimates exceed the upper bound than the lower bound. With $N = 1000$, $S = 200$, none of the $\hat{N}_e$ estimates were less than $0.5N_e$ for Critical $P = 0.01$ and only 2.6% were lower than $0.5N_e$ for Critical $P = 0.1$. With small sample sizes ($S = 50$), these numbers jumped to 11.8% and 26.2%, respectively (Table 2). Still, none of the estimates under any scenario with $N = 1000$ were as small as $0.1N_e$. Upper bounds on $\hat{N}_e$ are much less reliable. With $N = 1000$, $S = 50$, a quarter to a third of the estimates were more than ten times the true N_e, depending on the Critical P. Even with $S = 200$,

4.8% of the estimates were > $2N_e$ for Critical $P = 0.02$ or 0.01 and 12.6% of the estimates were > $2N_e$ for Critical $P = 0.1$ (Table 2). With $N = 5000$, a large fraction (>24%) of the $\hat{N}_e$ estimates exceeded $10N_e$ even with large samples, regardless of Critical P. In general, the fraction of estimates that exceeded a given range around the true N_e was negatively correlated with S and positively correlated with N and Critical P.

Mean squared error

To facilitate comparisons across multiple values of N and S, for each parameter set mean squared error (MSE) was scaled to the minimum found for any of the four Critical P values. Results (presented in the bottom panel of Figures 1-4) indicate that MSE was highest for Critical $P = 0.1$ and lowest for Critical $P = 0.02$ and 0.01. The exception was $N = 100$, in which case the lowest MSE occurred with Critical $P = 0.05$. The sensitivity of scaled MSE to the frequencies of alleles used varied in a complex way with sample size. Given the asymmetry in the distribution of $\hat{N}_e$, the primary factor determining the MSE was the fraction of estimates with very high values (constrained here to be no larger than $10N_e$). Conditions that led to a large difference in scaled MSE can be identified from the data in Table 2. For example, with $N = 500$ and $S = 200$, none of the $\hat{N}_e$ was as large as $10N_e$ for Critical $P = 0.01$, while $\hat{N}_e = 10N_e$ occurred in 2.9% of the trials with Critical $P = 0.1$. As a consequence, scaled MSE was much higher for the more restrictive Critical P (Figure 2, bottom panel). In contrast, with $S = 50$, in a reasonable fraction (9-17%) of the trials $\hat{N}_e = 10N_e$ for all Critical P; although each of these scenarios led to a much higher absolute MSE than was found with $S = 200$, the relative difference as a function of Critical P was not as large with the smaller sample size.

DISCUSSION

The LD method provides an estimate of contemporary effective size, and the resulting $\hat{N}_e$ can be interpreted as the effective number of breeders that produced the sample under consideration (Waples 2005). England et al. (2006) showed that estimates of N_e from the standard LD method are biased downwards if S is small relative to N_e. This bias arises because the expectation of $\hat{r}^2$ used by Hill (1981) ignored second order terms in S and N_e, and these terms become important in some cases (Waples 2006). The empirical correction used here to adjust for this bias was based on a wide range of sample and effective sizes ($S = 10$ to 200; $N_e = 10$ to 500; Waples 2006) but did not consider very large populations, such as might occur with some cetaceans. Results presented here show that the adjusted method can lead to essentially unbiased estimates of N_e even for populations as large as $N_e = 5000$. However, certain combinations of sample size, true N_e, and Critical P produce biased estimates. In general, $\hat{N}_e$ increases as alleles with lower frequency are included in the analysis, and bias is minimized by using Critical $P = 0.05$ or 0.02 (Figure 1-4, top panel). In most of the situations that lead to biased estimates the bias is relatively small (< 10-20%), which is modest compared to various other sources of uncertainty associated with estimating population size. However, in some scenarios the bias is more pronounced (large N, small S, Critical $P = 0.01$). Large samples ($S = 200$) had little bias under all scenarios considered.

Precision of the LD method with highly polymorphic markers has not previously been evaluated. The pairwise comparisons of loci are not entirely independent, so treating them as if they are leads to a somewhat optimistic view of precision (Waples 2006). However, this effect is relatively modest. A larger uncertainty was how precision would increase as a function of the number of alleles segregating at each locus. In these analyses, a pair of loci with K_1 and K_2 alleles was considered to have the equivalent of $(K_1 - 1)(K_2 - 1)$ independent allelic comparisons. The comparison of theoretical and empirical CIs (Figure 5) shows that this approximation is not far off. The theoretical CIs based on the assumed numbers of independent alleles and loci are consistently too narrow, but the disparity is not large. Expanding the theoretical CIs by about 10% in each direction appears to be reasonable as a rough approximation to allow use of this method until more thorough evaluations can be made. Bootstrapping $\hat{r}^2$ values across loci and/or alleles within loci might prove to be a viable means of generating confidence bounds for $\hat{N}_e$ for a particular dataset.

The MSE analyses are useful because they combine information about both precision and bias. Precision always increases with lower Critical P (which allows a larger number of alleles to be used; Table 1 and middle panels in Figures 1-4), but in many cases bias increases at the same time. Critical $P = 0.01$ led to the lowest MSE for $N = 500$-5000, whereas MSE was lowest for Critical $P = 0.05$ with $N = 100$. MSE has some limitations as a metric. Because of the asymmetry in the ability of genetic methods to detect N_e (much higher precision occurs when N_e is small), MSE is dominated by the (sometimes few) estimates that result in large $\hat{N}_e$. If underestimates are more of a management concern than overestimates, then MSE would not be the best performance measure.

Collectively, the results lead to the following conclusions regarding practical usefulness of the LD method.

1. The method can lead to very precise estimates for small populations (e.g., with S = true N_e = 100, none of the estimates exceeded $2N_e$ or were less than $0.5N_e$; Table 2).

2. As true N_e increases, the upper bound of $\hat{N}_e$ becomes increasingly broad. As a consequence, it is very difficult to distinguish a large population ($N_e > 1000$) from one that is infinitely large.

3. The LD method has more power to establish a lower bound for large populations. For example, with N = 5000 and S = 200, none of the estimates was as small as $0.1N_e$, and for N = 1000 and S = 200, only 2.6% of the estimates were as small as $0.5N_e$ (Table 2). Performance was much worse for smaller sample sizes.

4. Small samples (S = 50) produce reliable estimates of effective size only if N_e is also small. Since N_e is generally unknown, it is risky to estimate N_e based on small samples unless one is confident the sampled population is relatively small. Samples from populations that are believed to be, or potentially could be, relatively large should include at least 100 individuals whenever possible.

5. Critical P = 0.1 is too harsh a criterion; it often leads to bias in $\hat{N}_e$ and sacrifices too much precision. Critical P = 0.01 leads to the highest precision but also to upward bias in $\hat{N}_e$ in most cases. Use of Critical P = 0.02 (using all alleles with frequencies ≥ 0.02) appears to provide a good balance between maximizing precision and minimizing bias.

In all of the above I considered ideal populations, in which $N_e = N$. The ratio N_e/N is generally less than 1 in natural populations, and perhaps a great deal less. The ratio has not been evaluated in detail in cetaceans, but Felsenstein (1971) estimated the N_e/N ratio in humans (who like cetaceans have a Type-I survivorship curve) to be 0.34. This agrees well with the mean values Frankham (1995) reported for single-generation N_e/N ratios across a wide range of species. This means that the census size of cetacean populations could probably be two or three times as large as the $N = N_e$ values considered here and still result in the same precision in terms of estimating N_e. Departures from an ideal population that cause the N_e/N ratio to be less than one do not by themselves bias the estimate of N_e based on the LD method (Waples 2006).

All of the simulations used 20 "microsatellite-like" gene loci, which is at or near the upper limit of what is typically used in studies of natural populations. However, the analyses with more restrictive Critical P excluded large numbers of alleles, and results for these analyses provide some indication of the precision that can be expected from use of fewer gene loci. For example, if every locus has 10 alleles, each locus pair has the equivalent of 9x9 = 81 independent allelic comparisons. With only 10 loci, there are 10x9/2 = 45 pairwise comparisons of loci, leading to a total n = 81x45 = 3645. This is comparable to the number of alleles used in analyses using Critical P = 0.1 (Table 1). Many microsatellite loci have more than 10 alleles, which can at least partially compensate for use of fewer loci.

Like other genetic methods for estimating N_e, the LD method is based on numerous assumptions (random sampling and random mating; no immigration; no selection) that are likely to be violated in natural populations. The consequences of violating these assumptions remain to be evaluated in detail, but a discussion of these factors can be found in Waples (2006). If population size fluctuates, the LD method provides an estimate that is most strongly influenced by N_e in the parental generation (Waples 2005). Bottlenecks can be detected quickly, but if a population expands rapidly from a bottleneck the residual LD from the bottleneck could downward bias the estimate of N_e for a few generations. Cetacean populations are age-structured, and the consequences for estimates of N_e of sampling from age-structured populations remains to be evaluated.

REFERENCES

Balloux, F. 2001. *EASYPOP* (version 1.7): A computer program for population genetics simulations. J. Heredity 92:301-302.

Balloux, F. 2004. Heterozygote excess in small populations and the heterozygote-excess effective population size. Evolution 58:1891-1900.

England, P.R., Cornuet, J-M., Berthier, P., Tallmon, D.A., and Luikart, G. 2006. Estimating effective population size from linkage disequilibrium: severe bias using small samples. Cons. Genet. (in press; available now in Online early).

Felsenstein, J. 1971. Inbreeding and variance effective numbers in populations with overlapping generations. Genetics **68:**581–597.

Frankham, R. 1995. Effective population size/adult population size ratios in wildlife: a review. Genet. Res. (Cambridge) 66:95-107.

Hill, W.G. 1981. Estimation of effective population size from data on linkage disequilibrium. Genet. Res. 38:209–216.

Nei, M. and Tajima, F. 1981. Genetic drift and estimation of effective population size. Genetics **98:**625–640.

Roman, J. and Palumbi, S. R. 2003. Whales before whaling in the North Atlantic. Science 301:508-510.

Wang, J. 2001. A pseudo-likelihood method for estimating effective population size from temporally spaced samples. Genet. Res. 78:243-257.

Wang, J. 2005. Estimation of effective population sizes from data on genetic markers. Phil. Trans. Royal Soc., Series B 360:1395-1409.

Waples, R. S. 1989. A generalized approach for estimating effective population size from temporal changes in allele frequency. Genetics **121:**379–391.

Waples, R. S. 2005. Genetic estimates of contemporary effective population size: To what time periods do the estimates apply? Mol. Ecol. 14:3335-3352.

Waples, R. S. 2006. A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. Conservation Genetics (in press; available now in Online early).

Weir, B.S. 1996. *Genetic data analysis, 2nd Edition*. Sinauer, Sunderland, MA.

Table 1. Means (across all replicates) of the presumed number of degrees of freedom (n) associated with each estimate of N_e.

		Critical P			
N	S	0.1	0.05	0.02	0.01
100	50	3309	6674	8636	8849
	100	3219	6686	8998	9400
500	50	3220	6623	8904	9229
	100	3127	6650	9297	9949
	200	3077	6650	9510	10306
1000	50	3198	6691	9314	9744
	100	3104	6695	9716	10594
	200	3056	6685	9916	11000
5000	50	3293	6845	9248	9556
	100	3179	6857	9657	10267
	200	3139	6855	9858	10643

Table 2. Percentage of replicate estimates of N_e that fell outside the indicated lower and upper bounds relative to true $N_e = N$.

			Critical P					Critical P			
		Lower					Upper				
N	S	Bound	0.1	0.05	0.02	0.01	Bound	0.1	0.05	0.02	0.01
100	50	$< 0.1N$	0.0	0.0	0.0	0.0	$10N$	0.0	0.0	0.0	0.0
		$< 0.5N$	0.1	0.0	0.0	0.0	$> 2N$	2.5	0.6	1.3	4.6
	100	$< 0.1N$	0.0	0.0	0.0	0.0	$10N$	0.0	0.0	0.0	0.0
		$< 0.5N$	0.0	0.0	0.0	0.0	$> 2N$	0.0	0.0	0.0	0.0
500	50	$< 0.1N$	0.0	0.0	0.0	0.0	$10N$	17.2	11.1	10.8	9.4
		$< 0.5N$	15.5	7.0	2.4	2.1	$> 2N$	31.5	26.0	26.7	32.0
	100	$< 0.1N$	0.0	0.0	0.0	0.0	$10N$	3.6	0.9	0.5	0.4
		$< 0.5N$	3.5	1.0	0.0	0.0	$> 2N$	15.8	9.0	6.6	8.4
	200	$< 0.1N$	0.0	0.0	0.0	0.0	$10N$	0.1	0.0	0.0	0.0
		$< 0.5N$	0.1	0.0	0.0	0.0	$> 2N$	2.8	0.7	0.1	0.0
1000	50	$< 0.1N$	0.0	0.0	0.0	0.0	$10N$	33.3	28.6	25.2	26.7
		$< 0.5N$	26.2	21.6	14.8	11.8	$> 2N$	42.2	39.4	39.1	42.5
	100	$< 0.1N$	0.0	0.0	0.0	0.0	$10N$	17.0	8.8	5.9	6.2
		$< 0.5N$	15.1	5.8	2.2	2.2	$> 2N$	29.8	22.0	21.1	23.1
	200	$< 0.1N$	0.0	0.0	0.0	0.0	$10N$	2.9	0.7	0.1	0.0
		$< 0.5N$	2.6	0.2	0.0	0.0	$> 2N$	12.6	6.7	4.8	4.8
5000	50	$< 0.1N$	14.8	6.6	4.0	3.4	$10N$	47.2	47.4	48.2	49.8
		$< 0.5N$	45.2	39.8	37.4	34.6	$> 2N$	48.8	50.4	51.4	53.6
	100	$< 0.1N$	2.8	0.0	0.0	0.0	$10N$	43.4	40.0	39.4	40.6
		$< 0.5N$	41.2	35.2	28.8	28.6	$> 2N$	47.0	46.0	45.8	48.0
	200	$< 0.1N$	0.0	0.0	0.0	0.0	$10N$	39.4	31.8	27.0	24.4
		$< 0.5N$	31.4	24.6	17.8	15.2	$> 2N$	44.4	40.8	37.0	39.0

Figure 1. Summary of results for simulations with $N = 100$, plotted as a function of Critical P. Top panel: Bias in the overal $\hat{N}_e$ compared to the true $N_e = N$. Dotted line shows unbiased expectation $\hat{N}_e / N_e = 1.0$. Middle Panel: Precision of $\bar{\hat{r}}^2$, meaured as CV across 1000 replicate $\bar{\hat{r}}^2$ values computed using data for 20 gene loci. Bottom Panel: mean squared error (MSE) of $\hat{N}_e$, scaled so that the lowest MSE = 1.0

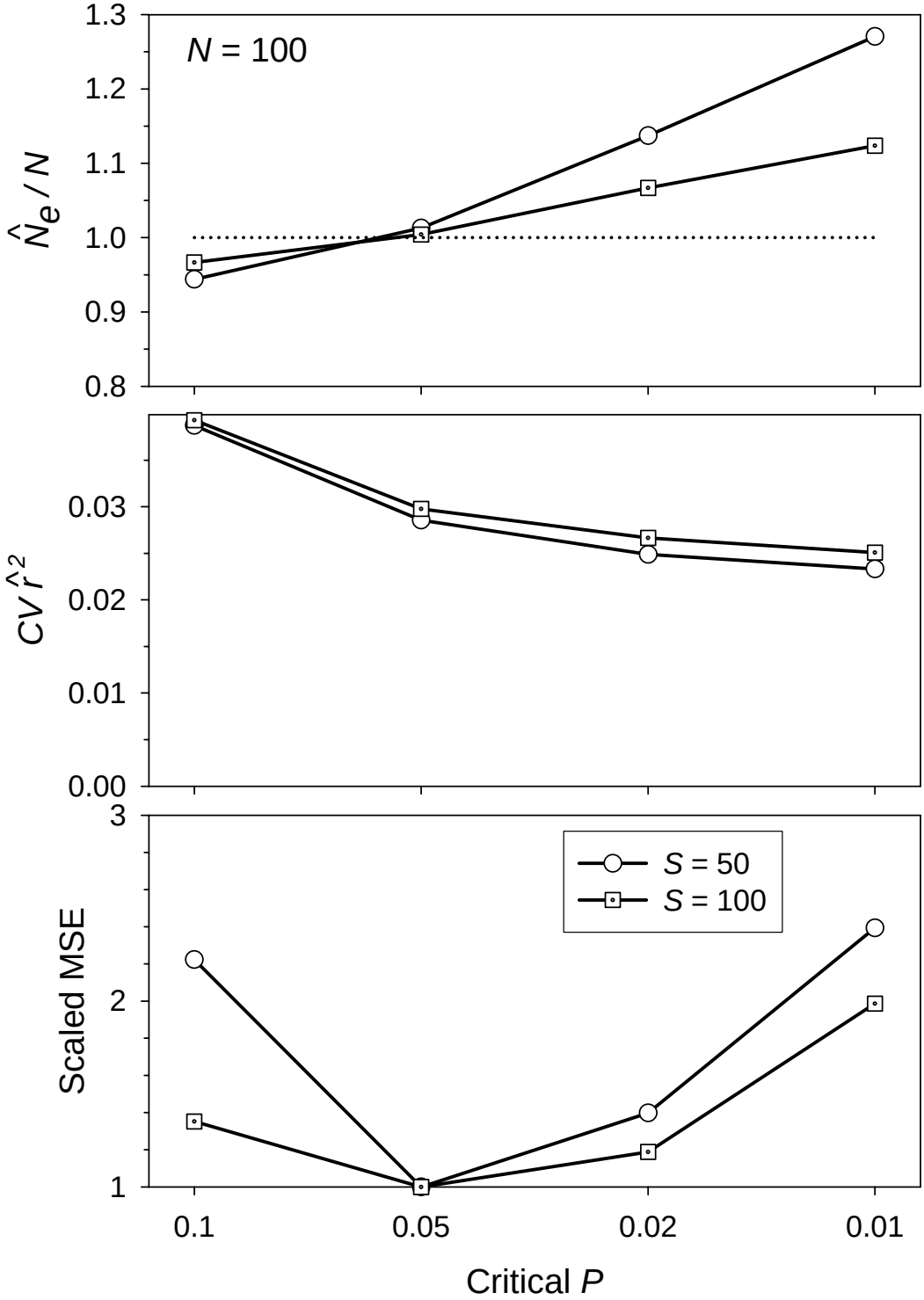

Figure 2. As in Figure 1, but for $N = 500$.

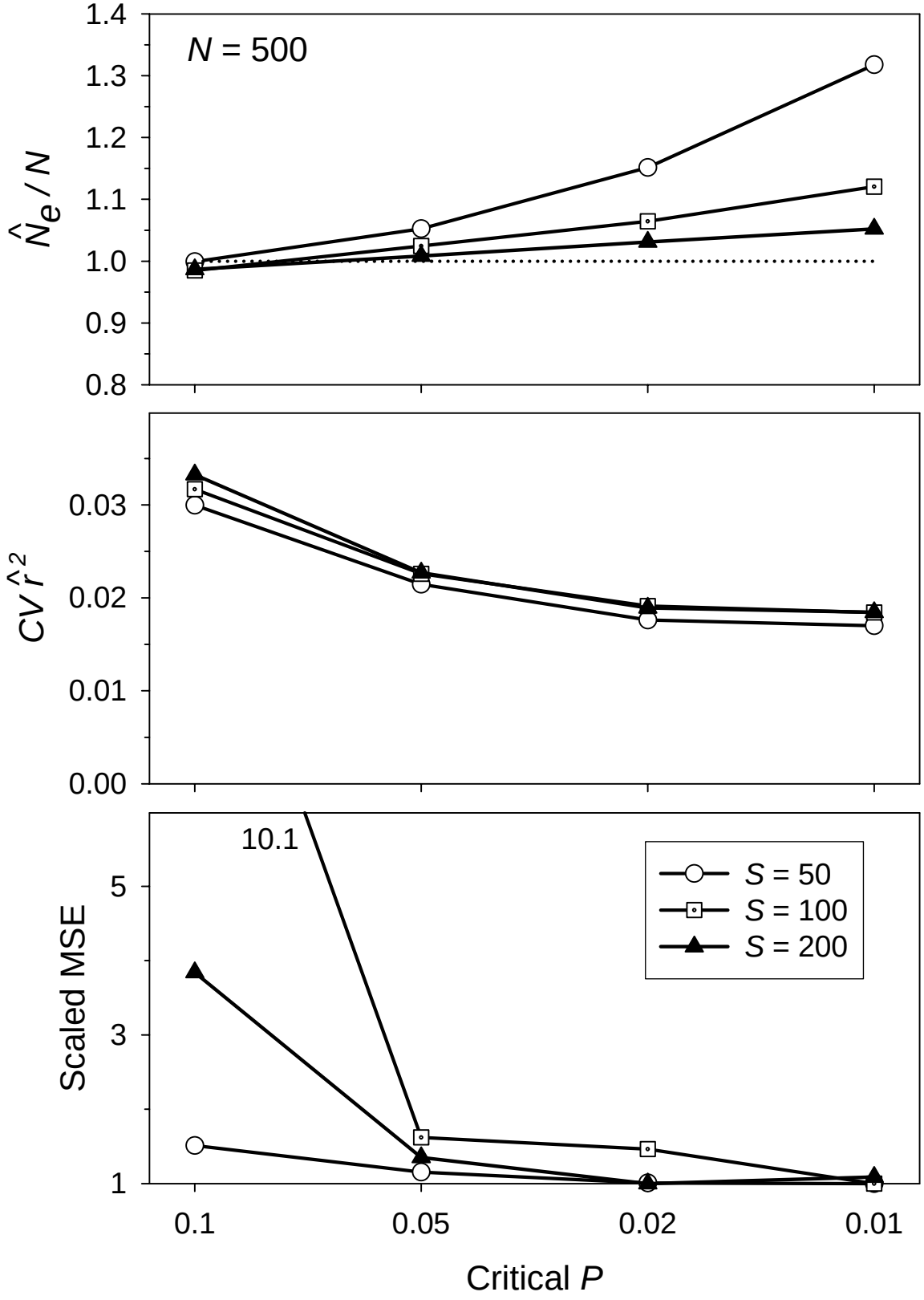

Figure 3. As in Figure 1, but for $N = 1000$.

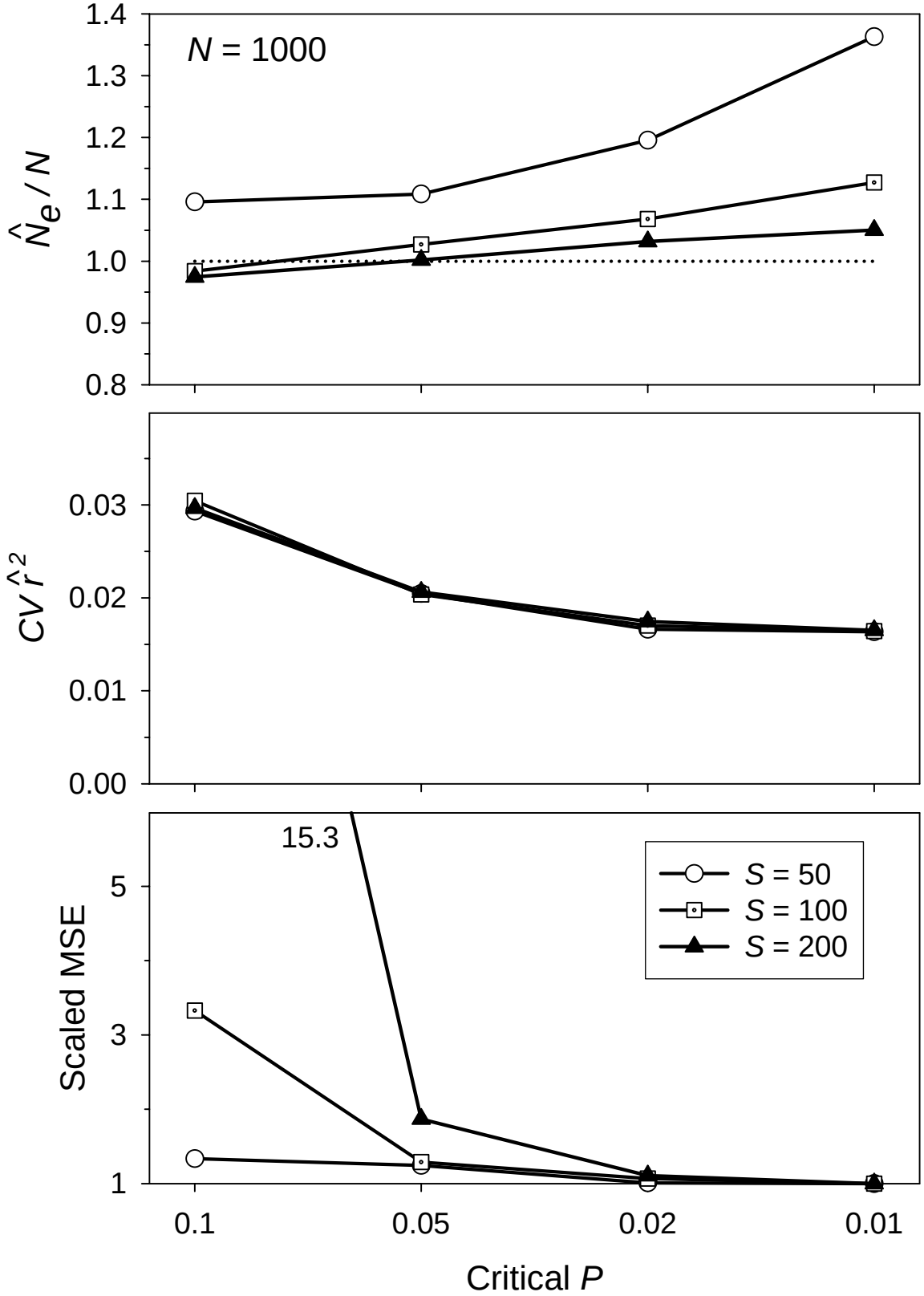

Figure 4. As in Figure 1, but for $N = 5000$.

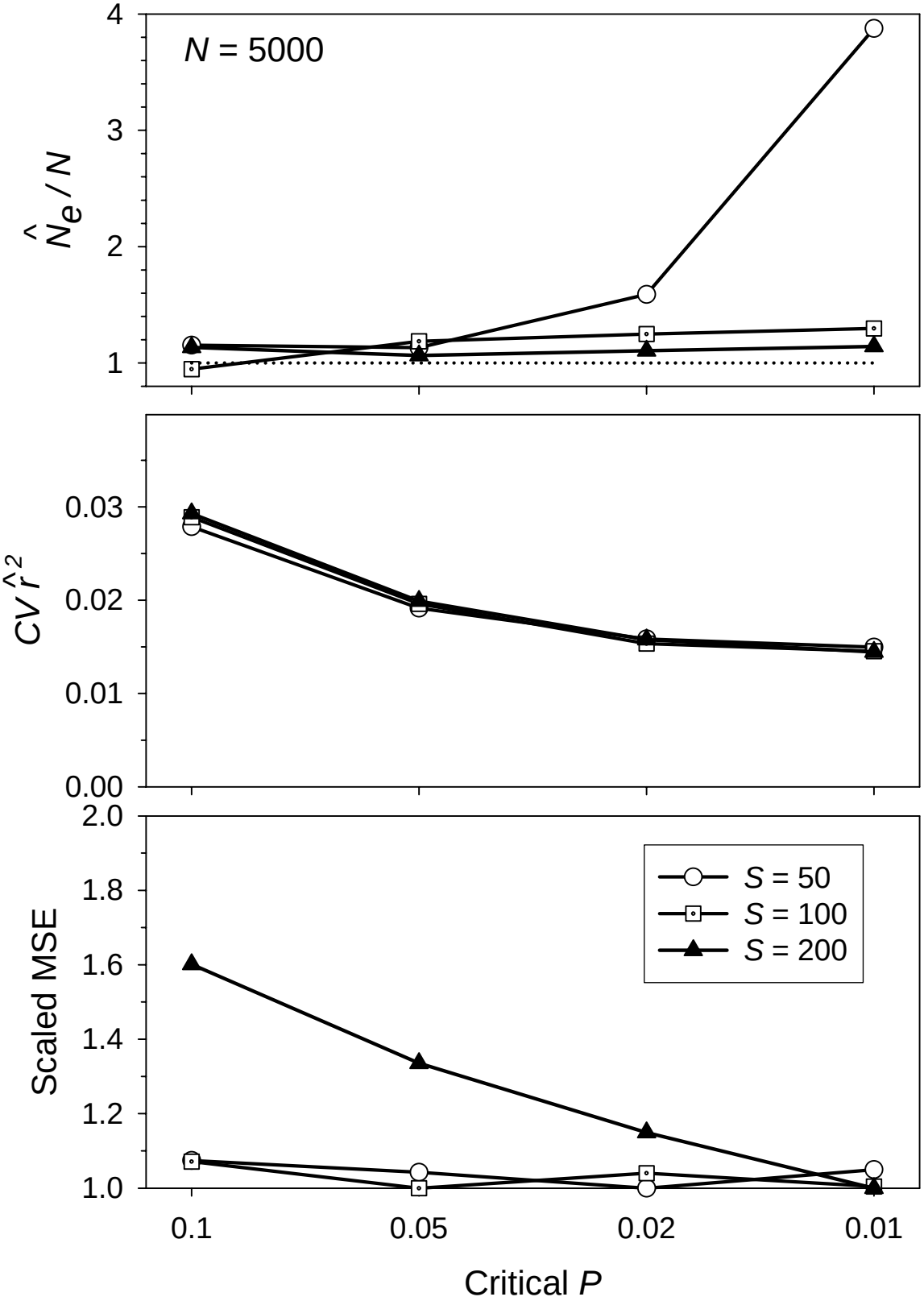
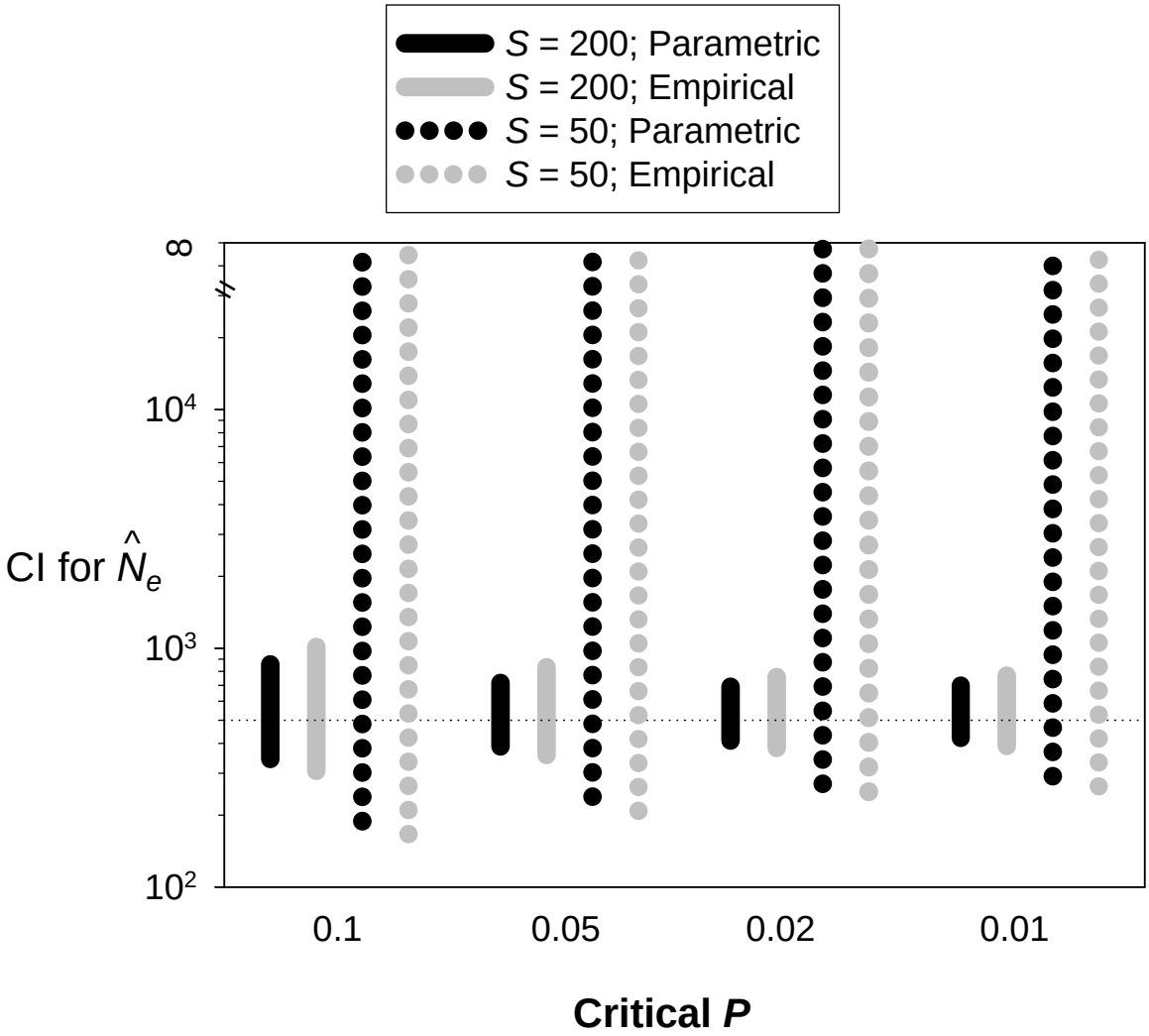
Figure 5. 95% confidence intervals (CIs) for $\hat{N}_e$ in the LD method as a function of sample size and critical P. Parametric CIs (black lines) were calculated from Equation 12 in Waples (2006); empirical CIs (gray lines) reflect simulated data. True N_e = 500 is shown by the horizontal dotted line.