

RESEARCH ARTICLE

# Statistical Efficiency in Distance Sampling

Robert Graham Clark\*

National Institute for Applied Statistics Research Australia (NIASRA), University of Wollongong, Wollongong, Australia

\* [rclark@uow.edu.au](mailto:rclark@uow.edu.au)



**OPEN ACCESS**

**Citation:** Clark RG (2016) Statistical Efficiency in Distance Sampling. PLoS ONE 11(3): e0149298. doi:10.1371/journal.pone.0149298

**Editor:** Shyamal D Peddada, National Institute of Environmental and Health Sciences, UNITED STATES

**Received:** November 29, 2015

**Accepted:** January 29, 2016

**Published:** March 7, 2016

**Copyright:** © 2016 Robert Graham Clark. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Data are based on simulations. Code to generate the simulated data and figures, etc. is included as supplementary material and is deposited into GitHub (<https://github.com/rgcstats/distance.simulation>).

**Funding:** The author has no support or funding to report.

**Competing Interests:** The author has declared that no competing interests exist.

## Abstract

Distance sampling is a technique for estimating the abundance of animals or other objects in a region, allowing for imperfect detection. This paper evaluates the statistical efficiency of the method when its assumptions are met, both theoretically and by simulation. The theoretical component of the paper is a derivation of the asymptotic variance penalty for the distance sampling estimator arising from uncertainty about the unknown detection parameters. This asymptotic penalty factor is tabulated for several detection functions. It is typically at least 2 but can be much higher, particularly for steeply declining detection rates. The asymptotic result relies on a model which makes the strong assumption that objects are uniformly distributed across the region. The simulation study relaxes this assumption by incorporating over-dispersion when generating object locations. Distance sampling and strip transect estimators are calculated for simulated data, for a variety of overdispersion factors, detection functions, sample sizes and strip widths. The simulation results confirm the theoretical asymptotic penalty in the non-overdispersed case. For a more realistic overdispersion factor of 2, distance sampling estimation outperforms strip transect estimation when a half-normal distance function is correctly assumed, confirming previous literature. When the hazard rate model is correctly assumed, strip transect estimators have lower mean squared error than the usual distance sampling estimator when the strip width is close enough to its optimal value ( $\pm 75\%$  when there are 100 detections;  $\pm 50\%$  when there are 200 detections). Whether the ecologist can set the strip width sufficiently accurately will depend on the circumstances of each particular study.

## 1 Introduction

The number of animals in a region is often of ecological importance. This paper considers the distance sampling approach to estimating abundance, in its usual conjunction with line transect sampling. Transect lines are laid down across a region, often parallel and equally spaced but not necessarily so. Observers move along the transects, and record observations of animals (or plants, or groups of animals, or other objects) and their perpendicular distances from the transect. Empirically, more objects are detected near to transect lines than far from them in many studies, suggesting that detectability is a decreasing function of distance. The distance sampling methodology exploits this phenomenon, by modelling the detection rate as a function

of distance. The number of detected objects can then be scaled to estimate the abundance  $N$  allowing for imperfect detection. Provided the assumptions of the method are met, the maximum range can be made fairly large, thereby increasing the sample size, while avoiding or reducing bias due to declining detection rate. Distance sampling is widely used in ecology: a Web of Science search found 276 articles on distance sampling in ecology journals in 2014 alone. The wide range of applications includes wild horses in the Australian Alps [1], large herbivores in South Africa's Kruger National Park [2] and odonata (dragonflies) in a rainforest locality in Papua New Guinea [3]. For a detailed description of the approach, see [4].

The major assumptions of the method are

1. Detection is perfect at zero distance.
2. The detection function is of known form, with some unknown parameters requiring estimation. Alternatively, model-averaging may be used provided the detection function is assumed to be one of a known set of alternatives.
3. Animals' distances to the nearest transect line are (at least approximately) uniformly distributed.

It is also assumed that there is no measurement error (for example, false positive detections, or mis-measurement of distances), that there is no movement of objects in response to the observer which could lead to multiple chances of detection, that detection events are independent, that there are sufficient transects for reliable estimation, and that transect lines are a good representation of the study area. This paper will consider the conventional distance sampling (CDS) scenario where there is only one observer. The same methodology can be used with multiple observers by pooling their detections. Mark recapture and mark recapture distance sampling (MRDS) are methods which more fully use data from multiple observers by matching their detections; MRDS, in particular, can be used to relax assumption (i). More recent approaches combine a spatial model with the detection model (see for example [5–7] and [8]). Spatial models allow abundances to be estimated for subregions, and can exploit spatial trends in estimation, however inference may be sensitive to the assumed spatial model which must therefore be carefully constructed. This paper focuses on CDS, as most applications of line transect sampling remain single observer. The use of spatial models in distance sampling is an important advance, but some researchers may decide that the extra effort and complication required to develop an adequate spatial model are not warranted for some studies, particularly when the number of detections is relatively small.

Robustness to the 3 assumptions above is explored in the literature. For example, MRDS can be used to achieve some level of robustness to (i). Assumption (ii) is dealt with by the use of flexible families of detection functions with two or more parameters, and the use of model averaging. [9] argue that (iii) is approximately satisfied provided transect lines are placed randomly or systematically from a random starting point. However, [10] question the uniformity assumption and find that CDS estimators are biased in a design-based framework, that is, under repeated random placement of transect lines. The matter remains in contention [11–13]. [11] suggest an alternative approach where the detection function is estimated from a separate calibration study, and [14] propose estimating detection probabilities using multiple observer data and possibly but not necessarily distance data.

A natural alternative to distance sampling is the simple scaling up of observations in a strip about the transect. When strips are too wide, this strip transect estimator is severely negatively biased due to non-detections, particularly of the more distant animals in the strip. When strips are sufficiently narrow, this bias becomes negligible, but the variance of the estimated abundance becomes large. Distance sampling aims to achieve lower variance by including

observations at greater distances, while reducing bias by adjusting for non-detection. But there is a hidden cost—the effect of unknown detection parameters on the precision of the estimated abundance—which reverses at least some of the benefit due to wider strips. This paper illuminates this cost both asymptotically and in small samples. This existence of a penalty due to unknown detection parameters is known, but no asymptotic expression has been derived in the literature, and the penalty has not been quantified except in a simulation of the half-normal and negative exponential detection functions [15]. A number of authors have compared CDS and strip transect estimators in particular applications (e.g. [1, 16]) but typically with only one or two options for strip width. [17] also suggested strip transects as an alternative to CDS but did not make a numerical comparison.

An example of the variability of CDS estimators may be found in [1], which compares strip transect estimates with two different strip widths to CDS estimates with various detection functions, and to mark-recapture and MRDS estimates of abundance. Detections of groups of wild horses are attempted up to 200m from a helicopter traversing 91 parallel transect lines in south-eastern Australia, resulting in 52 group detections (pooled from two observers). Table 1 here reproduces results from page 1145 of [1]. The strip transect estimates of abundance are the number of observed groups within 50m or 200m multiplied by the total area divided by the area lying within either 50m or 200m of a transect line, with no allowance for undercount. The values of the Akaike Information Criterion (AIC) are also shown for each detection model. Model averaged estimates are calculated using weights calculated from the AICs as described in [18]. Mark recapture and MRDS results are not reproduced here as they are outside the scope of the current paper. The CDS estimator shown in Table 1 corresponds to formula (6) in the next section of the current paper.

Table 1 shows that the CDS abundance estimators have high coefficients of variation (CV), with the two-parameter hazard rate model leading to a much higher CV (about 57%) than the other detection models which are one-parameter (CVs between 19% and 25%). The model-averaged estimator has a CV of 30%. The strip transect estimator using detections up to 200m has much lower CV, but is much lower than the CDS estimators, suggesting that it is negatively biased due to undetected groups. The strip transect estimator based on detections up to 50m is more plausible. It is close to the model-averaged CDS estimator, suggesting that its bias is small, presumably because the detection rate would decline relatively little over this shorter range. Surprisingly, the 0–50m strip transect estimator gives similar CVs to the 0–200m model-averaged CDS method, even though it only uses one quarter of the distance range and 44% of the detected groups. This motivated the research in the current paper on the efficiency of CDS estimators.

Section 2 proves for the first time that the usual CDS estimator of  $N$  is also a maximum likelihood estimator (MLE) of  $N$  under a particular model provided the likelihood is approximated using Stirling's Rule. Its asymptotic variance under the model is derived using this result. The asymptotic variance is identical to that of [19], but the use of the Stirling approximation allows a simpler derivation. The limiting variance of  $\hat{N}$  is expressed as the variance when the detection function is known multiplied by a penalty for unknown  $\theta$ . This asymptotic penalty is tabulated for various detection models. It can be substantial, and for many situations arising in practice is between 2 and 6. Section 3 summarizes a simulation study to evaluate the small sample performance of CDS estimators for various detection functions and strip transect estimators with differing strip widths. Unlike the theoretical result in Section 2, the simulation allows for overdispersion in the counts of animals falling in any given range. Section 4 contains conclusions about the magnitude of the penalty due to unknown detection parameters in CDS, and the relative performance of CDS and strip transect estimators.

**Table 1. Estimates of Density (abundance/area) of Horse Groups obtained from [1].** CV is the estimated coefficient of variation of the estimate, given by the square root of the estimated variance divided by the estimate.

Method	Detection Function	AIC	Density ( $\hat{N}/A$ ) (groups/km <sup>2</sup> )	CV%( $\hat{N}$ )	Strip Width	estimated average detection probability
CDS	neg. exponential	0.00	0.36	24.8	200m	0.50
CDS	uniform (cosine)	0.96	0.28	18.8	200m	0.64
CDS	half-normal	1.35	0.28	19.9	200m	0.64
CDS	hazard rate	1.87	0.36	57.2	200m	0.50
CDS	model-averaged	n/a	0.33	30.3	200m	0.55
Strip	n/a	n/a	0.31	22.1	50m	1.00
Strip	n/a	n/a	0.18	16.1	200m	1.00

doi:10.1371/journal.pone.0149298.t001

## 2 Theoretical results on maximum likelihood estimation of $N$

### 2.1 Notation and background

The aim is to estimate the number of objects, which may be animals, groups of animals or other objects, in a region. Let  $N$  be the number of objects in a defined study area and  $n$  be the number of detections made by an observer moving along predefined line transects. Observation may either be one-sided (only objects to the right, or only objects to the left, are observed) or two-sided. Only objects with perpendicular distance up to a pre-chosen limit  $w$  from a transect line have a chance of being observed (the *covered region*); let  $N_c$  be the number of such objects. Two-sided observation is the more common case, but one-sided observation is sometimes necessary, for example if the observer can only see out one side of a vehicle. It is assumed that the probability of observing an object at perpendicular distance  $d_i$  from a transect line is  $g(d_i, \theta)$  when  $0 \leq d_i \leq w$ , where  $\theta$  is a vector of  $p$  parameters specifying the function within a family. It is assumed that  $g(0, \theta) = 1$  and that  $g$  is a non-increasing function of distance. The *Distance* software [20], which implements both CDS and MRDS, allows four possible functional forms for  $g()$ , including the half-normal model,  $g(d_i) = \exp\left(-\frac{d_i^2}{\theta^2}\right)$ , and the hazard rate function,  $g(d_i) = 1 - \exp\left(-\left(d_i/\theta_1\right)^{-\theta_2}\right)$ . Both of these functions satisfy the *shoulder condition* that  $g'(0) = 0$ , with the hazard rate model giving greater flexibility in modelling the *shoulder width* (i.e. the range near 0 over which  $g$  is relatively flat). The use of “robust models”, which have enough flexibility to model a range of typical shapes, is recommended on pages 46–49 of [9], with the hazard rate model given as a particularly useful example.

It is also possible to include other covariates affecting the detectability of objects in the distance function, such as characteristics of the animal or plant (see [21] and [22]).

Let  $d_i, i = 1, \dots, N$ , be the perpendicular distances from the objects to the nearest transect line, and let  $\delta_i = 1$  for observed objects and  $\delta_i = 0$  for the rest. Under model assumption (iii) stated in the Introduction,  $d_i$  are independent and identically distributed uniform  $U(0, M)$  for  $i = 1, \dots, N$ , where  $M$  is the maximum possible distance from a transect line. The distribution of  $d_i$  given  $\delta_i = 1$  is easily derived as

$$g_{cond}(d_i; \theta) = g(d_i; \theta) / \int_0^w g(u; \theta) du \tag{1}$$

Let  $\bar{g}(\theta) = w^{-1} \int_0^w g(u; \theta) du$  be the unconditional probability of detection. (This is the probability that an animal in the covered region is detected, unconditional on its distance, denoted by  $P_a$  in equation (2) of [23]. The notation  $\bar{g}(\theta)$  is used here to emphasise that it is the mean

value of  $g(d_i; \theta)$  over  $d_i \leq w$ .) The conditional likelihood of  $\mathbf{d}_s = (d_1, \dots, d_n)^T$  given  $n$  is

$$L_d = \prod_{i=1}^n g_{cond}(d_i; \theta) = \left( \prod_{i=1}^n g(d_i; \theta) \right) \left( \int_0^w g(u; \theta) du \right)^{-n} = \left( \prod_{i=1}^n g(d_i; \theta) \right) w^{-n} \bar{g}(\theta)^{-n} \quad (2)$$

and the corresponding conditional log-likelihood given  $n$  is

$$l_d = \sum_{i=1}^n \log g(d_i; \theta) - n \log(\bar{g}(\theta)) - n \log w. \quad (3)$$

The parameters  $\theta$  can be obtained by setting the derivative of  $l_d$  with respect to  $\theta$  to 0. Prior to this, it is convenient to define

$$\mathbf{h}(u; \theta) = \frac{\partial}{\partial \theta} g(u; \theta) \quad (4)$$

and  $\bar{\mathbf{h}}(\theta) = w^{-1} \int_0^w \mathbf{h}(u; \theta) du$ . Notice that  $\mathbf{h}(u; \theta)$  is a  $p$ -vector where  $p$  is the number of parameters in  $\theta$ . The partial derivative of  $\bar{g}(\theta)$  with respect to  $\theta$  is  $\bar{\mathbf{h}}(\theta)$ , subject to regularity conditions allowing the derivative operator to be taken within the integral. Setting the derivative of  $l_d$  to 0 gives the following estimating equation for  $\theta$ :

$$0 = \sum_{i=1}^n g(d_i; \theta)^{-1} \mathbf{h}(d_i; \theta) - n \bar{g}(\theta)^{-1} \bar{\mathbf{h}}(\theta) \quad (5)$$

The most commonly used estimator of  $N$  in CDS is:

$$\hat{N}_{CDS} = \frac{n}{P \bar{g}(\hat{\theta}_{CML})}. \quad (6)$$

where  $P$  is the proportion of the area falling within perpendicular distance of  $w$  of a transect line, and  $\hat{\theta}_{CML}$  is the solution to Eq (5) (CML stands for conditional maximum likelihood, as the likelihood  $L_d$  in Eq (2) is conditional on  $n$ ). In CDS, the estimation of  $\theta$  is model-based and maximises the likelihood conditional on  $n$ , while the estimation of  $N$  in Eq (6) is design-based and motivated by the fact that  $E[n] = NP\bar{g}(\theta)$ . See equation 1.4 of [9]. See also [23] for a recent discussion of CDS and extensions. The value of  $P$  is assumed to be known, and is approximated in practice by the total length of all transects multiplied by  $w$  (and multiplied by 2 for two-sided observation) divided by the region's area.

Another alternative that has been proposed is maximum likelihood estimation of  $N$ , which is the minimum variance unbiased estimator for large samples subject to regularity conditions. Section 7.2 of [24] derives the likelihood for  $N$  and  $\theta$ . The conditional density of  $\mathbf{d}_s$  given  $n$  is  $L_d$  in Eq (2). Under the assumed model,  $\delta_i$  are independent Bernoulli random variables with expected value  $\bar{g}(\theta)$ . Hence

$$n \sim \text{bin}(N, P\bar{g}(\theta)). \quad (7)$$

The likelihood is the product of the probability function of  $n$  and  $L_d$ :

$$L = \binom{N}{n} \{P\bar{g}(\boldsymbol{\theta})\}^n \{1 - P\bar{g}(\boldsymbol{\theta})\}^n L_d \tag{8}$$

$$= \binom{N}{n} \{P\bar{g}(\boldsymbol{\theta})\}^n \{1 - P\bar{g}(\boldsymbol{\theta})\}^{N-n} \left( \prod_{i=1}^n g(d_i; \boldsymbol{\theta}) \right) w^{-n} \bar{g}(\boldsymbol{\theta})^{-n} \tag{9}$$

$$= \binom{N}{n} P^n w^{-n} \{1 - P\bar{g}(\boldsymbol{\theta})\}^{N-n} \prod_{i=1}^n g(d_i; \boldsymbol{\theta}) \tag{10}$$

(See equation 2.33 of [4]).  $L$  can be maximised with respect to  $N$  and  $\boldsymbol{\theta}$  to obtain a maximum likelihood estimator of  $N$ . In this approach, both  $N$  and  $\boldsymbol{\theta}$  are treated as unknown parameters. On pages 16–17, [4] recommend against this approach because in practice  $n$  is likely to be overdispersed and so to have higher variance than implied by Eq (7). For example, this would occur if there were positive correlations between the values of  $d_i$ . It is worth noting that any such overdispersion is likely to also invalidate Eq (2) and so  $\hat{\boldsymbol{\theta}}_{CML}$  and  $\hat{N}_{CDS}$  to some extent, as Eq (2) assumes that  $d_i$  are independent conditional on  $n$ .

When  $\boldsymbol{\theta}$  is known, the MLE of  $N$  is the integer part of  $n/(P\bar{g}(\boldsymbol{\theta}))$  (see pages 17–19 and 138 of [24]). The same reference also shows that if the factorial terms in Eq (10) are approximated using Stirling’s rule and  $N$  is treated as a continuous parameter, the MLE is then

$$\hat{N}_{known\boldsymbol{\theta}} = n/(P\bar{g}(\boldsymbol{\theta})). \tag{11}$$

It is straightforward to derive the variance of  $\hat{N}_{known\boldsymbol{\theta}}$  using the fact that  $n \sim \text{bin}(N, P\bar{g}(\boldsymbol{\theta}))$ :

$$\text{var}(\hat{N}_{known\boldsymbol{\theta}}) = P^{-2} \bar{g}(\boldsymbol{\theta})^{-2} \text{var}(n) = NP^{-1} \bar{g}(\boldsymbol{\theta})^{-1} (1 - P\bar{g}(\boldsymbol{\theta})) \tag{12}$$

When  $\boldsymbol{\theta}$  is unknown, [19] note that the MLE of  $N$  is the integer part of  $n/(P\bar{g}(\hat{\boldsymbol{\theta}}))$  where  $\hat{\boldsymbol{\theta}}$  is the MLE of these parameters. The next subsection of this paper extends this result by showing that  $\hat{\boldsymbol{\theta}}_{CML}$  and  $\hat{N}_{CDS}$  maximise a Stirling approximation to the full likelihood  $L$ . This enables a theoretical result on the large sample variance of  $\hat{N}_{CDS}$ , albeit under the strong assumption Eq (7). The simulation study in Section 3 relaxes this assumption by including overdispersion.

## 2.2 Derivations of the MLE and its variance based on Stirling’s approximation

The maximum likelihood estimator of  $N$  under Stirling’s approximation for factorials is derived in this subsection, where the likelihood  $L$  is given by Eq (10). (Note that in section 3.3.1, [9] consider maximization of  $L_d$ , not  $L$ , to estimate  $\boldsymbol{\theta}$ , where  $L_d$ , defined in Eq (2), is the conditional likelihood given  $n$ ). Stirling’s rule  $\log(x!) \approx x \log(x) - x$  implies that

$$\log \binom{N}{n} = \log\{N!/n!(N - n)!\} \tag{13}$$

$$\approx N \log(N) - N - (n \log(n) - n) - \{(N - n) \log(N - n) - (N - n)\} \tag{14}$$

$$= N \log(N) - (N - n) \log(N - n) - n \log(n). \tag{15}$$

There is a positive probability that  $n$  is equal to 0 or  $N$ , in which case  $\log(n)$  or  $\log(N - n)$  are not defined. The limit of the right hand side of Eq (15) can easily be shown to equal 0 in these cases, so the following refinement of Eq (15) is used:

$$\log \binom{N}{n} \approx \begin{cases} N \log(N) - (N - n) \log(N - n) - n \log(n) & \text{if } 0 < n < N \\ 0 & \text{if } n = 0 \text{ or } n = N \end{cases}. \quad (16)$$

$L$  will be maximised with respect to  $N$  and  $\theta$  treating  $N$  as a continuous parameter. Stirling's approximation for  $\log(x!)$  is very accurate even for small  $x$  as long as  $x$  is at least 2 or 3, and both  $N - n$  and  $n$  would be well above this in practice. Theorem 1 states the MLEs and the approximate Fisher information.

**Theorem 1** *The model defined by Eqs (1) and (7) is assumed, and it is assumed that the log of the likelihood Eq (10) can be approximated using Eq (16), leading to*

$$l = N \log(N) - (N - n) \log(N - n) - n \log(n) + n \log(P) - n \log(w) + (N - n) \log(1 - P\bar{g}(\theta)) + \sum_{i=1}^n \log(g(d_i; \theta)) \quad (17)$$

Let  $\Omega$  be an open set defining the set of feasible values of  $\theta$ . If there is a unique  $\hat{\theta}_{CML}$  satisfying Eq (5) then it is the maximum likelihood estimator, and the MLE of  $N$  on  $(0, \infty)$  is  $\hat{N}_{CDS}$  in Eq (6).

Let  $D$  be a random variable with density  $g(d)/\int_0^w g(u)du$  for  $0 \leq d \leq w$ , which is also the distribution of  $d_i$  conditional on detection. The Fisher Information of  $(N, \theta^T)^T$  is approximately equal to

$$\mathcal{I}(N, \theta) \approx \begin{bmatrix} \mathcal{I}_{NN} & \mathcal{I}_{N\theta}^T \\ \mathcal{I}_{N\theta} & \mathcal{I}_{\theta\theta} \end{bmatrix} \quad (18)$$

for large  $N$ , where

$$\mathcal{I}_{NN} = N^{-1}(1 - P\bar{g}(\theta))^{-1}P\bar{g}(\theta) \quad (19)$$

$$\mathcal{I}_{\theta\theta} = NPw^{-1} \int_0^w \mathbf{h}(u; \theta)\mathbf{h}(u; \theta)^T g(u; \theta)^{-1} du + NP\bar{\mathbf{h}}(\theta)\bar{\mathbf{h}}(\theta)^T (1 - P\bar{g}(\theta))^{-1} \quad (20)$$

$$= NP\bar{g}(\theta) \text{var}_D(\mathbf{h}(D; \theta)/g(D; \theta)) + NP\bar{\mathbf{h}}(\theta)\bar{\mathbf{h}}(\theta)^T \bar{g}(\theta)^{-1} (1 - P\bar{g}(\theta))^{-1} \quad (21)$$

$$\mathcal{I}_{N\theta} = (1 - P\bar{g}(\theta))^{-1}P\bar{\mathbf{h}}(\theta) \quad (22)$$

Surprisingly,  $\hat{\theta}_{CML}$  and  $\hat{N}_{CDS}$  maximise the Stirling approximation to the full likelihood  $L$ , even though  $\hat{\theta}_{CML}$  was defined as the maximizer of the conditional likelihood  $L_d$  given  $n$  (not  $L$ ), and  $\hat{N}_{CDS}$  was motivated by a design-based argument.

Let  $V$  be the inverse of the Fisher Information matrix in Eq (18). Subject to regularity conditions, maximum likelihood estimators are asymptotically normal with expectation equal to the true parameter values and variance-covariance matrix equal to the inverse of the Fisher Information matrix. Unfortunately these regularity conditions do not hold here, for example condition (M3) of the Central Limit Theorem on pages 499–500 of [25] is not met, because  $n$  and  $\mathbf{d}_i$  are dependent. Moreover, Eq (18) is only the approximate Fisher information based on a Taylor Series expansion, whereas the usual Central Limit Theorem requires the exact Fisher

information. [26] uses an alternative method of proof to derive a Central Limit Theorem for the MLE. It is shown in this working paper that  $V$  is indeed the limiting variance of  $(\hat{N}, \hat{\theta}^T)^T$ . The proof in [26] is essentially a simplified version of the proof of the result in [19] which does not use the Stirling approximation. Theorem 1 is an advance on the result in [19], because it shows that  $\hat{N}_{CDS}$  is the full maximum likelihood estimator (subject to the Stirling approximation), and also provides a simpler derivation of  $V$ . Theorem 1 helps explain the finding of [19] that the difference between  $\hat{N}_{CDS}$  and the exact maximum likelihood estimator (i.e. the MLE when the Stirling approximation is not made) is small asymptotically. The theorem also suggests that there is no need for the calculation of the exact MLE in practice, even when the model assumptions are justified, since  $\hat{N}_{CDS}$  maximises an excellent approximation to the full likelihood.

It is convenient to express  $V$  in block form. We will henceforth mostly write  $g(u)$ ,  $\mathbf{h}(u)$ ,  $\bar{g}$  and  $\bar{\mathbf{h}}$  for readability, rather than  $g(u; \theta)$  etc. Using a standard result on the inverse of a matrix in block form (e.g. 5.16a of [27]),  $V$  is equal to

$$V = \begin{bmatrix} V_{11} & V_{21}^T \\ V_{21} & V_{22} \end{bmatrix} \tag{23}$$

where

$$\left. \begin{aligned} V_{22} &= \{ \mathcal{I}_{\theta\theta} - \mathcal{I}_{N\theta} \mathcal{I}_{NN}^{-1} \mathcal{I}_{N\theta}^T \}^{-1} = N^{-1} P^{-1} \Delta^{-1} \bar{g}^{-1} \\ V_{11} &= \mathcal{I}_{\theta\theta}^{-1} + \mathcal{I}_{\theta\theta}^{-1} \mathcal{I}_{N\theta}^T V_{22} \mathcal{I}_{N\theta} \\ &= N(1 - P\bar{g})P^{-1}\bar{g}^{-1} + NP^{-1}\bar{g}^{-3}\bar{\mathbf{h}}^T \Delta^{-1} \bar{\mathbf{h}} \\ V_{21} &= V_{21}^T = -V_{22} \mathcal{I}_{N\theta} \mathcal{I}_{NN}^{-1} \\ &= -\bar{g}^{-2} P^{-1} \Delta^{-1} \bar{\mathbf{h}} \end{aligned} \right\} \tag{24}$$

and  $\Delta$  is the  $p$  by  $p$  matrix defined by

$$\Delta = \text{var}_D[\mathbf{h}(D)/g(D)]. \tag{25}$$

The limiting variance of  $\hat{N}_{CDS}$ ,  $V_{11}$  from Eq (24), is of primary interest. It can be expanded by elementary operations as

$$\text{var}(\hat{N}_{CDS}) = V_{11} = \text{var}(\hat{N}_{known\theta})F \tag{26}$$

where  $\text{var}(\hat{N}_{known\theta})$  is the variance of  $\hat{N}$  when  $\theta$  is known, as defined by Eq (12), and

$$F = 1 + \bar{\mathbf{h}}^T \Delta^{-1} \bar{\mathbf{h}} \bar{g}^{-2} (1 - P\bar{g})^{-1} \tag{27}$$

is a penalty term attributable to  $\theta$  requiring estimation. The penalty  $F$  is always 1 or greater, because  $\Delta$  is a variance-covariance matrix, and so is positive semi-definite.

The coefficient of variance (CV) of  $\hat{N}_{CDS}$  follows directly, noting that  $E[n] = NP\bar{g}$ :

$$CV^2 = \text{var}(\hat{N}_{CDS})/N^2 = FNP^{-1}\bar{g}^{-1}(1 - P\bar{g})/N^2 \tag{28}$$

$$= F(1 - E[n]/N)/E[n] \tag{29}$$

## 2.3 Numerical values of the asymptotic variance for selected models

Values of  $\Delta$  are obtained by rewriting Eq (25) as

$$\Delta = \frac{w^{-1} \int_0^w \mathbf{h}(u)\mathbf{h}(u)^T g(u)^{-1} du}{\bar{g}} - \frac{\bar{\mathbf{h}}\bar{\mathbf{h}}^T}{\bar{g}^2} \tag{30}$$

and calculating by quadrature using the `integrate` function in the R Statistical Environment [28]. Table 2 shows values of  $F$  numerically calculated for various hazard rate models. The parameter  $\theta_2$  determines the shape of the detection curve, with 1.1 giving a very narrow shoulder (i.e. steeply declining for small distances) and 3 giving a very wide shoulder. Hazard rate detection models for a number of values of  $\theta_2$  are illustrated in the next section. The parameter  $\theta_1$  is calculated numerically to give the specified  $\bar{g}$  in each row.

Table 2 shows that  $F$  increases as  $\theta_2$  decreases, i.e. as the shoulder becomes narrower. For a given  $\bar{g}$ ,  $F$  decreases as the coverage rate  $P$  increases. This is because increasing  $P$  improves the precision of  $\hat{N}_{CDS}$ , but it improves the precision of  $\hat{N}_{known\theta}$  even faster. For a given  $P$ ,  $F$  decreases as  $\bar{g}$  increases when  $\theta_2 \leq 2$  (narrow shoulder), but increases as  $\bar{g}$  increases when  $\theta_2 > 2$ . A possible reason is that when  $\theta_2$  is small and  $\bar{g}$  is large, the detection function is near 1 for much of its range but then decreases precipitously. The fact that it remains near 1 for much of the range may mean that  $\hat{N}$  is relatively insensitive to  $\hat{\theta}$ . In contrast, when  $\theta_2 > 2$  and  $\bar{g}$  is large, the detection function declines more smoothly, so that  $\hat{N}_{CDS}$  is more sensitive to  $\hat{\theta}$ .

Values of  $\bar{g}$  of 0.1, 0.3 or 0.6,  $P = 0.3$  and  $\theta_2 \geq 1.25$  are probably the most representative of studies in practice.  $F$  varies from 1.9 to 5.6 in this subset of Table 2.

Table 3 shows similar results for half-normal detection models. The values of  $F$  are generally much closer to 1 than in Table 2, varying from 1.5 to 1.9 in the subset of the table where  $\bar{g} \in \{0.1, 0.3, 0.6\}$  and  $P = 0.3$ .  $F$  increases with  $P$ , as in Table 2.  $F$  increases with  $\bar{g}$  for fixed  $P$ , similar to the wide-shouldered results in Table 2 where  $\theta_2 > 2$ .

## 3 Simulation study

### 3.1 Design of the simulation study

**Generation of distances  $d_i$  for  $i = 1, \dots, N$ .** Distance data are simulated for abundances  $N$  such that the expected numbers of detections are  $E[n] = 50, 100, 200, \dots, 1000$ , and the fraction of the area covered is  $P = 0.1$ . 10,000 simulations are used in every case.

Distances  $d_i$  are generated both with and without overdispersion. In the latter case,  $d_i$  are independent  $U(0, 10)$  random variables. The maximum range of observation is set to  $w = 1$ , so that objects are only eligible for detection when  $d_i \leq 1$ , and so the probability of any given object falling within the covered area is  $P = 0.1$ . One implication of this model is that the number of objects  $n(v)$  with distance falling into any given interval of length  $v$  within  $[0, 10]$  is distributed as  $n(v) \sim bin(N, vP)$ , and hence  $E[n(v)] = NvP$  and  $var[n(v)] = NvP(1 - vP)$ . For example,  $N_c$  is a special case of  $n(v)$  with  $v = 1$ . The assumption of independent uniform distances has been criticised because in practice  $n(v)$  is often observed to be overdispersed, with variance greater than  $NvP(1 - vP)$ .

Overdispersed  $d_i$  are generated by firstly replacing the  $U(0, 10)$  distribution by the discrete approximation with probability 0.001 at each of 1000 evenly spaced values between 0 and 10. The probability that  $d_i$  falls in any given interval would then be 0.001 in the non-overdispersed case. Overdispersed  $d_i$  are assumed to be discrete random variables with the same set of possible values, with probability  $\phi_k$  for value  $k = 1, \dots, 1000$ . The vector  $\phi$  is simulated as coming from a Dirichlet distribution with vector parameter equal to  $0.001\alpha \mathbf{1}_{1000}$  where  $\mathbf{1}_{1000}$  is a vector containing 1000 values all equal to 1, and  $\alpha$  is a parameter which controls the variance of  $\phi$ .

**Table 2. Asymptotic penalty (F) for the hazard rate model for selected values of the coverage rate P, the shape parameter (θ<sub>2</sub>) and the mean detection rate  $\bar{g}$ .** The largest distance for which detection is attempted is w = 1 in all cases.

P	$\bar{g}$	$\theta_2$					
		1.1	1.25	1.5	2	2.5	3
0.1	0.3	6.60	5.28	3.99	2.79	2.25	1.94
0.3	0.3	6.97	5.56	4.19	2.91	2.33	2.00
0.6	0.3	7.63	6.06	4.54	3.12	2.47	2.11
0.9	0.3	8.44	6.68	4.97	3.38	2.65	2.25
0.1	0.6	5.03	4.25	3.43	2.59	2.17	1.92
0.3	0.6	5.62	4.72	3.78	2.82	2.34	2.05
0.6	0.6	6.92	5.77	4.56	3.33	2.71	2.34
0.9	0.6	9.23	7.63	5.96	4.24	3.38	2.87
0.1	0.9	3.28	2.93	2.54	2.09	1.85	1.70
0.3	0.9	3.85	3.41	2.91	2.36	2.06	1.87
0.6	0.9	5.52	4.82	4.04	3.16	2.69	2.39
0.9	0.9	11.94	10.25	8.35	6.24	5.08	4.35

doi:10.1371/journal.pone.0149298.t002

When  $\alpha \rightarrow \infty$ ,  $\phi$  is equal to  $\mathbf{1}_{1000}/1000$  with probability 1, resulting in the non-overdispersed case. For  $0 < \alpha < \infty$ ,  $\phi$  are random variables each lying between 0 and 1, with  $\sum_{k=1}^{1000} \phi_{k=1}$ . The parameters  $\phi$  are generated anew for each simulation, and  $\{d_i; i = 1, \dots, N\}$  are then generated as independent discrete random variables with probabilities  $\phi$  at each of the 1000 evenly spaced values between 0 and 10.

The overdispersed  $d_i$  have the property that  $n(v)$  is beta-binomial distributed with parameters  $N, \alpha P v$ , and  $\alpha(1 - P v)$ . (This follows from the properties of the Dirichlet-multinomial and beta-binomial distributions—see for example [29].) The expected value of  $n(v)$  is then  $N v P$  as before, but the variance is inflated to  $var[n(v)] = c N v P (1 - v P)$  where  $c = (\alpha + N)/(\alpha + 1)$  is an overdispersion factor. Values of  $\alpha$  corresponding to  $c = 1, 2$  and  $3$  are used.

The above process is a discrete approximation of a Dirichlet process with base distribution given by  $U(0, 10)$ . Dirichlet processes are widely used as prior distributions for distribution functions (e.g. chapter 23 of [30]). This also makes them suitable to simulate overdispersed

**Table 3. Asymptotic penalty (F) for the half-normal model for selected values of the coverage rate P and the mean detection rate  $\bar{g}$ .** The largest distance for which detection is attempted is w = 1 in all cases.

P	$\bar{g}$	Penalty (F)
0.1	0.3	1.52
0.3	0.3	1.55
0.6	0.3	1.61
0.9	0.3	1.69
0.1	0.6	1.78
0.3	0.6	1.90
0.6	0.6	2.15
0.9	0.6	2.60
0.1	0.9	2.23
0.3	0.9	2.54
0.6	0.9	3.44
0.9	0.9	6.90

doi:10.1371/journal.pone.0149298.t003

data, particularly as the resulting  $d_i$  have the property that the number of distances falling in any interval is overdispersed to the same degree.

**Simulation of detection process.** Objects are detected with probability  $g(d_i; \theta)$  when  $d_i \leq w = 1$  for each  $i$ , with detection independent across objects. Two families of detection functions are used: the hazard rate and the half-normal. Both are among those proposed in [9]. The two-parameter hazard rate function meets the requirement specified on page 41 of [9] of being a flexible model, giving some robustness to mis-specified detection function; in particular, it allows the shoulder to be narrow or wide. The half-normal detection function is less flexible, but is also often used, and would generally be easier to fit from data as it has only one parameter. Fig 1 shows the 4 particular distance functions used: hazard rate (hr) functions  $g(d_i; \theta) = 1 - e^{-(d_i/\theta_1)^{-\theta_2}}$  with  $\theta = (0.405, 1.25)$ ,  $\theta = (0.448, 2)$  and  $\theta = (0.484, 3)$  corresponding to a very narrow, narrow, and wide shoulder respectively; and a half-normal (hn) function  $g(d_i; \theta) = 1 - e^{-(d_i/\theta)^2/2}$  with  $\theta = 0.502$ . This gives a variety of shapes of the detection function, with all 4 having the same average detection rate of  $\bar{g}(\theta) = 0.6$ . This means that the number of detections is approximately  $n \approx P\bar{g}(\theta)N = 0.06N$ . The values of  $g(w)$  for all three functions are at least 0.11, and all but the wide hazard rate function are at least 0.14, roughly in line with the rule of thumb for choice of  $w$  on page 16 of [9]. The figure also shows the asymptotic penalty  $F$  due to unknown  $\theta$  for each detection function. The penalty ranges from 2 to 5.4.

**Estimators of abundance.** The CDS estimator  $\hat{N}_{CDS}$  defined in Eq (6) is calculated using the hazard rate model and the half-normal model. The CDS estimator assuming known  $\theta$ , defined in Eq (11), is also calculated. This last option is of course unrealistic in practice, but is included to show the impact of uncertainty about  $\theta$ .

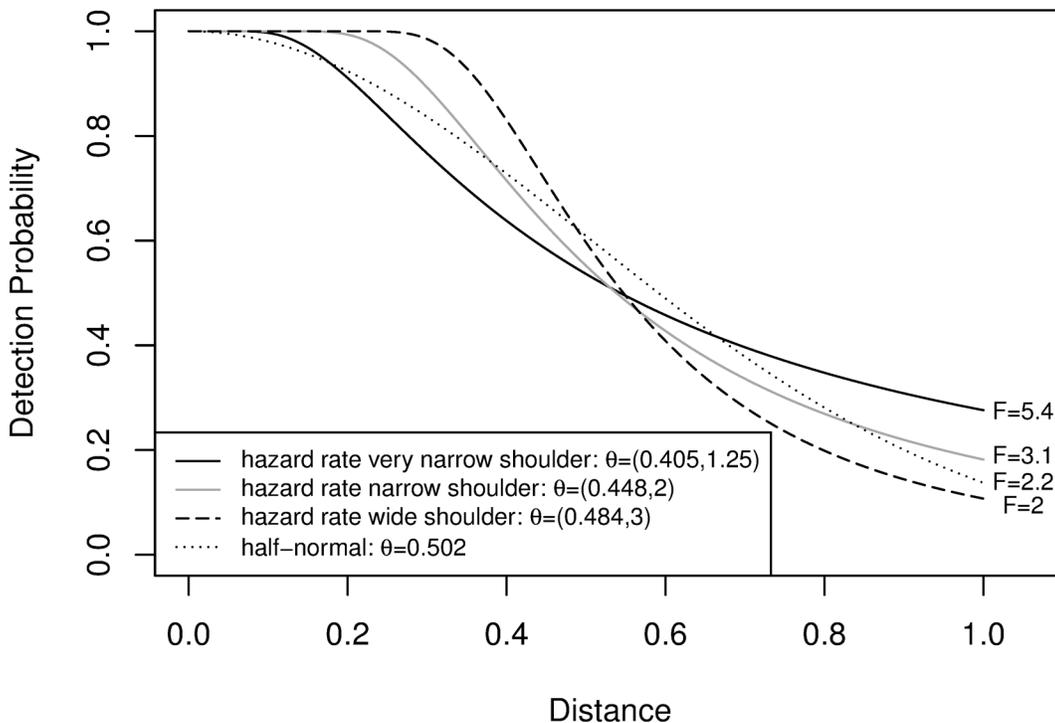
An estimate of  $N$  is also calculated using the strip transect estimator  $\hat{N}_{ST} = n/P$ . This is unbiased under the binomial model  $n \sim bin(N, P)$ , which incorrectly assumes perfect detection up to distance  $w$ . The strip transect estimator is also applied to restricted datasets using only those distances up to a range of  $w' = 0.01, 0.02, \dots, 1$ , with  $\hat{N}_{ST}(w') = n[d \leq w']/(Pw'/w)$ .

Each simulation corresponds to a single detection function, and a single value of  $E[n]$  and of  $c$ . 10,000 populations of distances and detections are generated in each simulation. All computations are carried out in the R statistical environment version 3.0.1 [28]. The `Distance` package [31] is not used because of occasional non-convergence (this would generally not be an issue in practice, but is a problem in a large simulation). The estimator  $\hat{\theta}_{CML}$  is calculated by maximising  $l_d$  from Eq (3) using the `optim` function, using the Nelder-Mead method for the two-parameter hazard rate function and the Brent method for the one-parameter half-normal function. The complete simulation requires approximately 14 hours on a Macbook Pro with a 2.7GHz Intel Core i7 processor and 16GB of RAM. The code to conduct the simulation and produce the figures and tables is in [S1 Code](#).

The aim of the simulation is to estimate the penalty factor due to unknown detection parameters for finite sample sizes, and to compare the mean squared errors (MSEs) of line transect and strip transect estimators in different scenarios. The focus of the paper is on variances and mean squared errors of abundance estimators, so variance estimators and confidence interval coverage are not reported on.

### 3.2 Simulation results for the asymptotic penalty

Fig 2 shows how the simulation estimates of  $F$  converge to the asymptotic values as  $n$  and  $N$  increase, for  $c = 1$ . For the hazard rate with very narrow shoulder and the halfnormal model, the asymptotic approximation is good even for  $E[n] = 100$ . For the other two models, the small sample penalties are higher than the asymptotic value for  $E[n] = 100$ , but converge to the



**Fig 1. Detection Functions used to Generate Simulated Data.** The variance penalty factors  $F$  from Eq (26) due to unknown  $\theta$  are also shown.

doi:10.1371/journal.pone.0149298.g001

asymptote as  $E[n]$  increases. The results provide a computational confirmation of the derivation of  $F$  in Section 3. Fig 3 shows the simulation estimates of  $F$  when the overdispersion parameter  $c$  is 2. No asymptotic result for  $F$  is available in this case. The values of  $F$  are 0.1–0.7 higher than when  $c = 1$ .

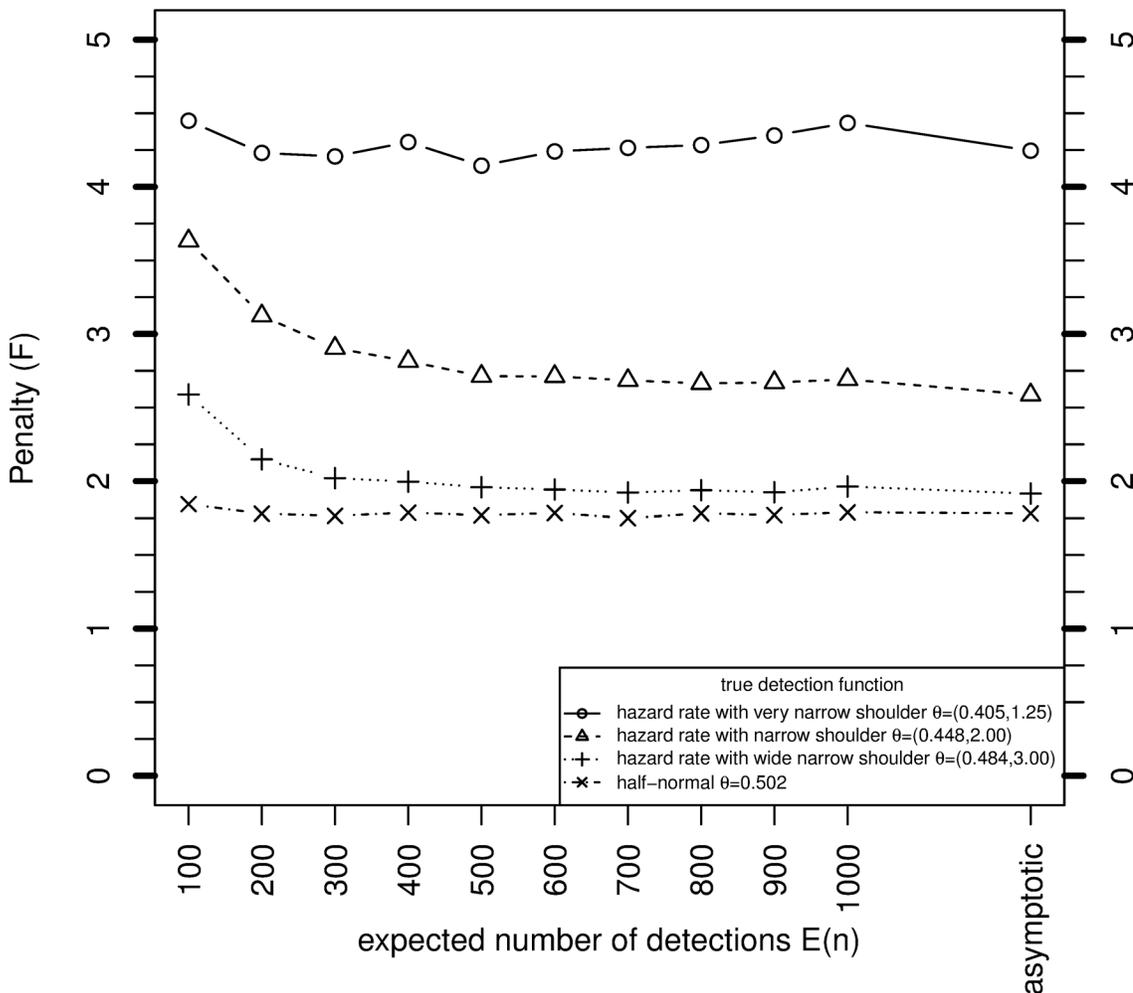
### 3.3 Simulation MSEs of CDS and strip transect estimators

Fig 4 compares the simulation MSEs of  $\hat{N}_{CDS}$ , and  $\hat{N}_{ST}$  with varying strip width, when  $E[n] = 50$ . The 3 horizontal red lines in each plot are the MSEs of  $\hat{N}_{CDS}$ , which make use of all detections up to distance  $w = 1$ . The 3 lines are for overdispersion values of  $c = 1, 2$  and  $3$ . The 3 blue curves show the MSEs of  $\hat{N}_{ST}$  for  $c = 1, 2$  and  $3$ , with strip widths of  $0.01, 0.02, \dots, 1$ . Fig 4(a) shows MSEs for the half-normal detection function. Fig 4(b), 4(c) and 4(d) are results for the hazard rate models with very narrow, narrow and wide shoulders, respectively.

The format of Fig 4 is loosely based on Figure 1 in [15]. [15] compared CDS and strip transect estimators for the half-normal and negative exponential detection functions (both one parameter families) for  $n = 40, 60$  and  $100$ , and  $c = 1, 2$  and  $3$ . Here we extend these results to the two-parameter hazard rate function, and we simulate over-dispersed distances corresponding to  $c = 1, 2$  and  $3$ , rather than using the approximate formula in [15] to convert results when  $c = 1$  to other values of  $c$ .

Figs 5, 6 and 7 are of the same format as Fig 4 and show results when  $E[n]$  is  $100, 200$  and  $400$  respectively.

For the half-normal function, Figs 4–7 replicate the finding of [15] that CDS dominates strip transect estimators, with the former having lower MSE for almost all values of  $c, E[n]$  and strip width.



**Fig 2. Variance of MLE of  $N$  when  $\theta$  is unknown relative to when it is known, for various sample sizes based on simulation, and asymptotically from Eq (26), where  $P = 0.1$ ,  $c = 1$  and  $\bar{g} = 0.6$ .**

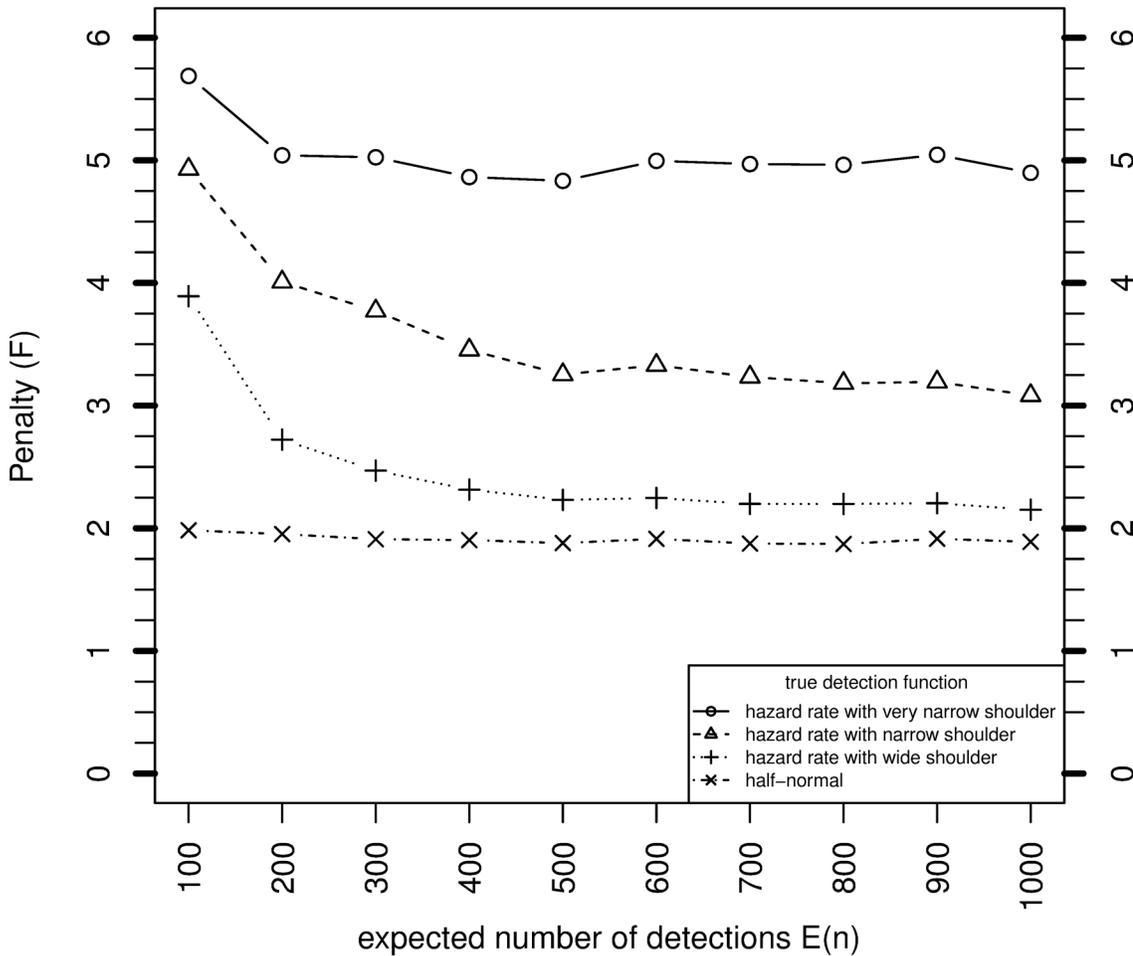
doi:10.1371/journal.pone.0149298.g002

For the hazard rate function, the picture is quite different, presumably because of the larger value of the penalty ( $F$ ) due to unknown detection parameters for this model. [15] argue that  $c = 2$  is the most practically relevant scenario out of  $c = 1, 2$  and  $3$ , so we concentrate on this case. For this value of  $c$ , and for the narrow shoulder detection function (panel (c) of each figure),  $\hat{N}_{ST}$  has lower MSE than  $\hat{N}_{CDS}$  when the strip width is  $0.1$  and above (when  $E[n] = 50$ ), between  $0.15$  and  $0.71$  (when  $E[n] = 100$ ),  $0.19$  and  $0.54$  (when  $E[n] = 200$ ) and  $0.21$  and  $0.45$  (when  $E[n] = 400$ ). The optimal strip widths are  $0.47, 0.42, 0.38$  and  $0.35$  for  $E[n]$  equal to  $50, 100, 200$  and  $400$ , respectively.

$\hat{N}_{CDS}$  performs better relative to  $\hat{N}_{ST}$  than the above when the hazard rate function has a wide shoulder, and worse than the above when the shoulder is very narrow.

The MSEs of both  $\hat{N}_{CDS}$  and  $\hat{N}_{ST}$  increase as  $c$  increases, particularly  $\hat{N}_{ST}$ . So when  $c$  increases,  $MSE[\hat{N}_{CDS}]/MSE[\hat{N}_{ST}]$  decreases slightly.

Simulations were also carried out with values of  $P$  other than  $0.1$ . Results for  $P$  equal to  $0.2, 0.25, 0.3$  and  $0.5$ , with  $E[n] = 100$ , are in S1, S2, S3 and S4 Figs, respectively. A comparison of



**Fig 3. Variance of MLE of  $N$  when  $\theta$  is unknown relative to when it is known, for various sample sizes based on simulation, and asymptotically from Eq (26), where  $P = 0.1$ ,  $c = 2$  and  $\bar{g} = 0.6$ .**

doi:10.1371/journal.pone.0149298.g003

these figures to Fig 4 shows that mean squared errors decrease slightly as  $P$  increases. The relative performance of the different methods for the various  $c$  and  $w$  is insensitive to  $P$ .

### 4 Conclusions

To achieve a given coefficient of variation  $CV$  for the line transect estimator, the required expected sample size is

$$E[n] = (N^{-1} + CV^2 F^{-1})^{-1} \tag{31}$$

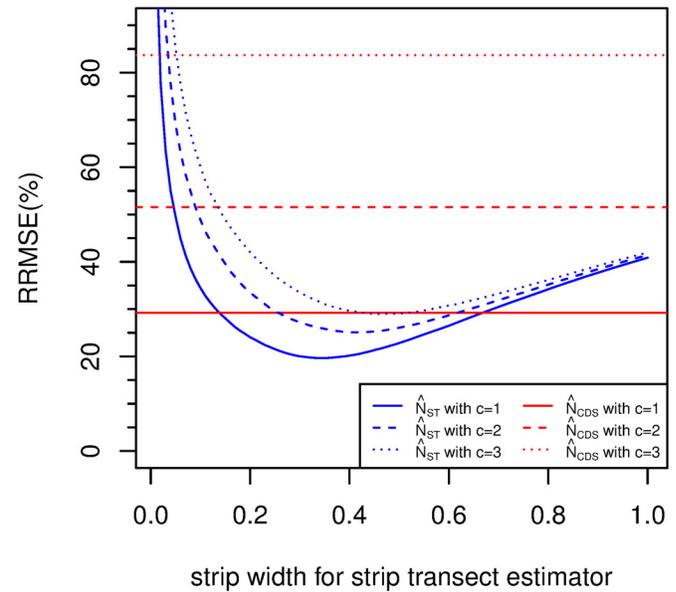
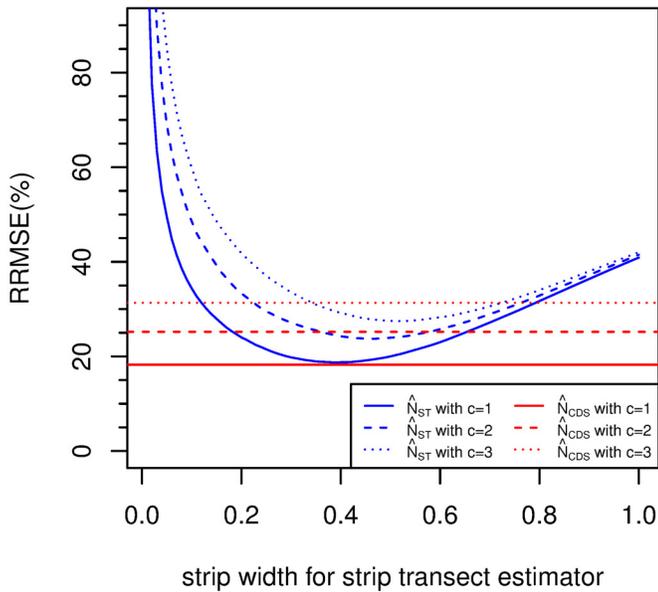
(follows from rearranging Eq (29)). When  $n$  is much smaller than  $N$ , this simplifies to

$$E[n] = F / CV^2. \tag{32}$$

The factor  $F$  is the inflation due to unknown detection function parameters. Asymptotic values of  $F$  are made available here for the first time (see Tables 2 and 3), albeit under a strong simplifying assumption that counts are binomially distributed. The asymptotic values of  $F$  for typical hazard rate models are between 2 and 6. Simulation confirms the accuracy of the asymptotic result when there is no overdispersion, and shows that  $F$  is larger in the presence of

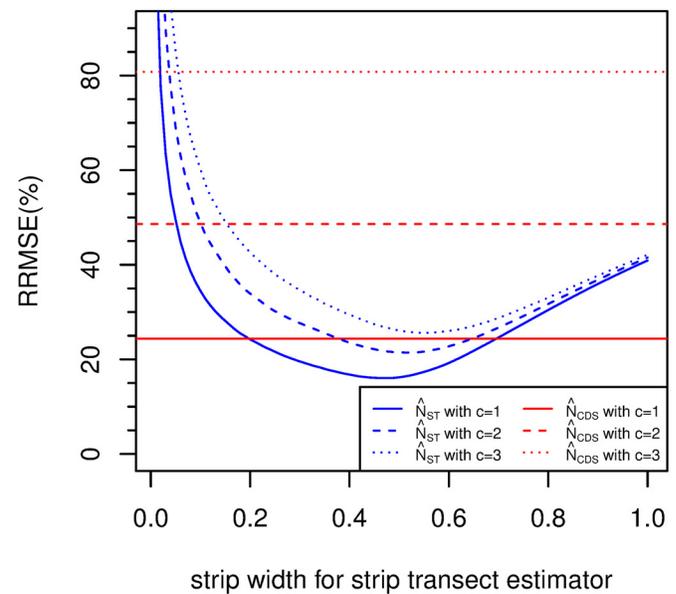
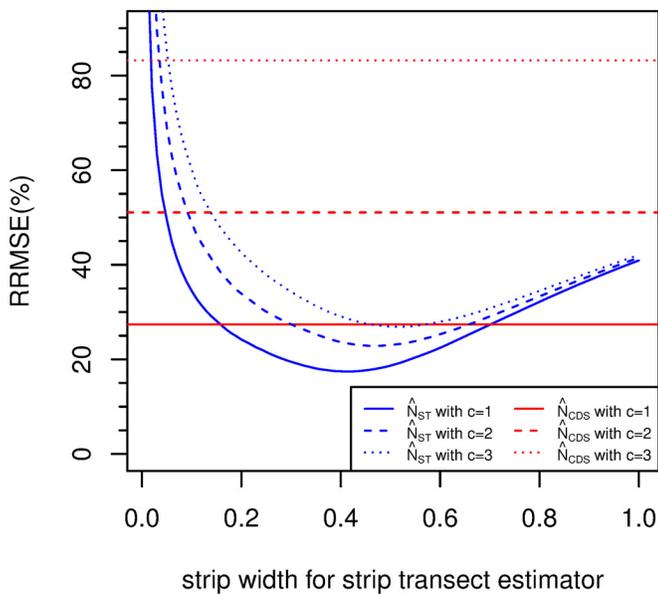
(a) half-normal  $\theta=0.502$

(b) hazard rate with very narrow shoulder  $\theta=(0.405,1.25)$



(c) hazard rate with narrow shoulder  $\theta=(0.448,2.00)$

(d) hazard rate with wide shoulder  $\theta=(0.484,3.00)$

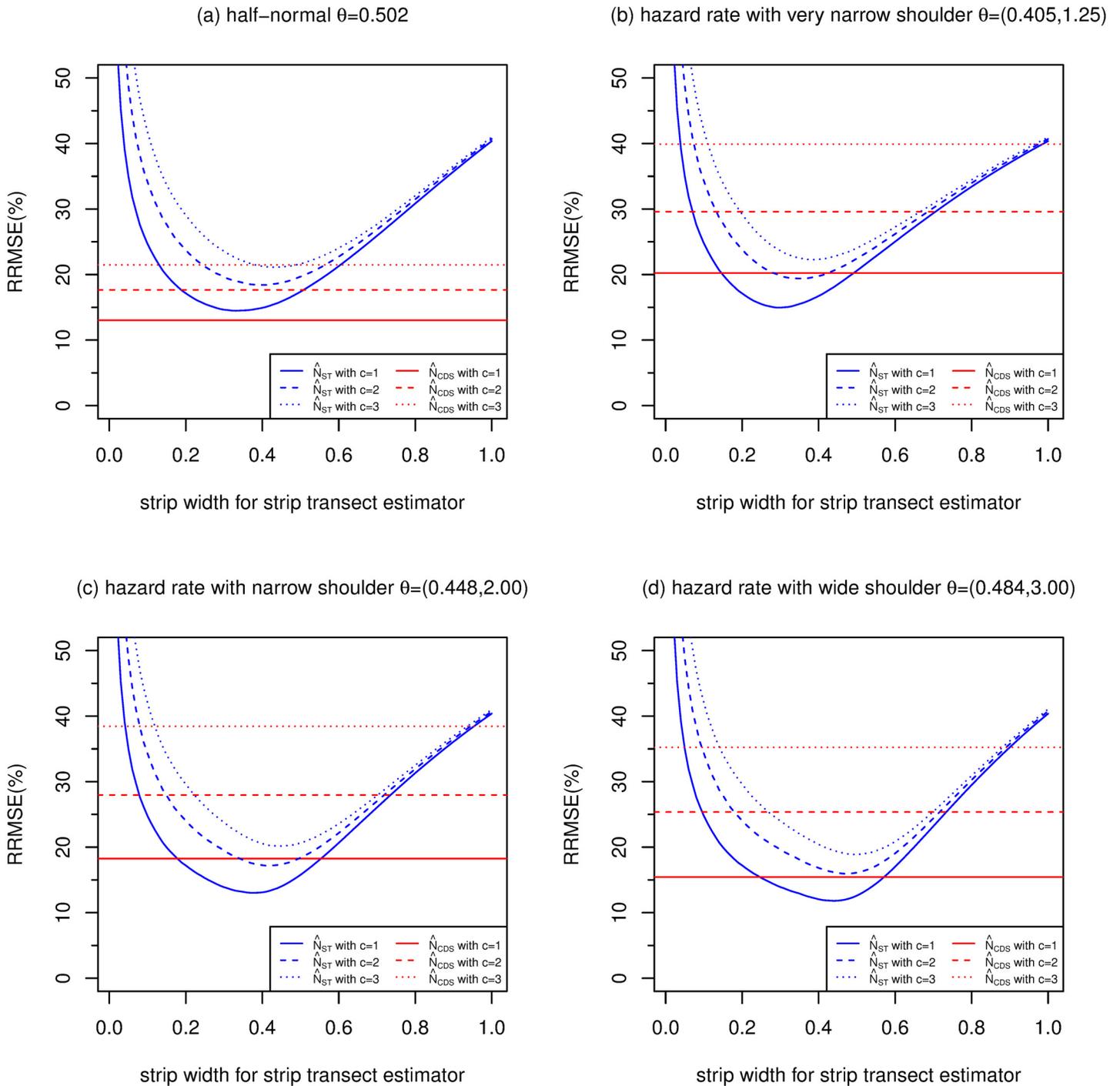


**Fig 4. Relative root mean squared errors (RRMSE) (%) of CDS and strip transect estimators (strip widths ranging from 0 to 1) of  $N$ , for 4 detection functions, with  $E[n] = 50$ ,  $P = 0.1$  and  $\bar{g} = 0.6$ .**

doi:10.1371/journal.pone.0149298.g004

overdispersion. The R package `distance.sample.size`[32] calculates  $F$  and required sample size using the methods described in the current paper.

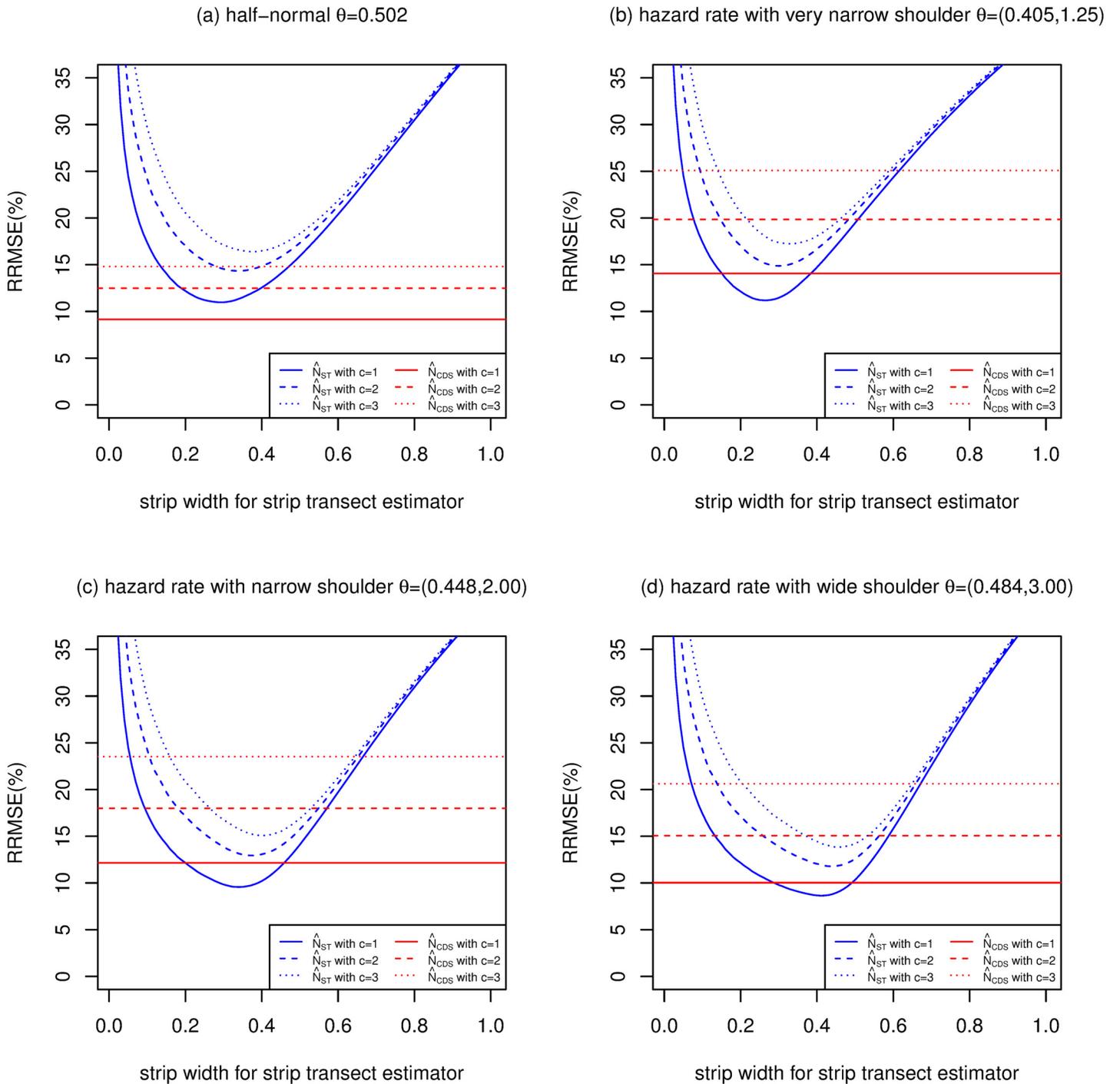
Note that the inflation factor  $F$  does not apply if  $g$  is modelled using mark-recapture data as suggested by [14].



**Fig 5. Relative root mean squared errors (RRMSE) (%) of CDS and strip transect estimators (strip widths ranging from 0 to 1) of  $N$ , for 4 detection functions, with  $E[n] = 100$ ,  $P = 0.1$  and  $\bar{g} = 0.6$ .**

doi:10.1371/journal.pone.0149298.g005

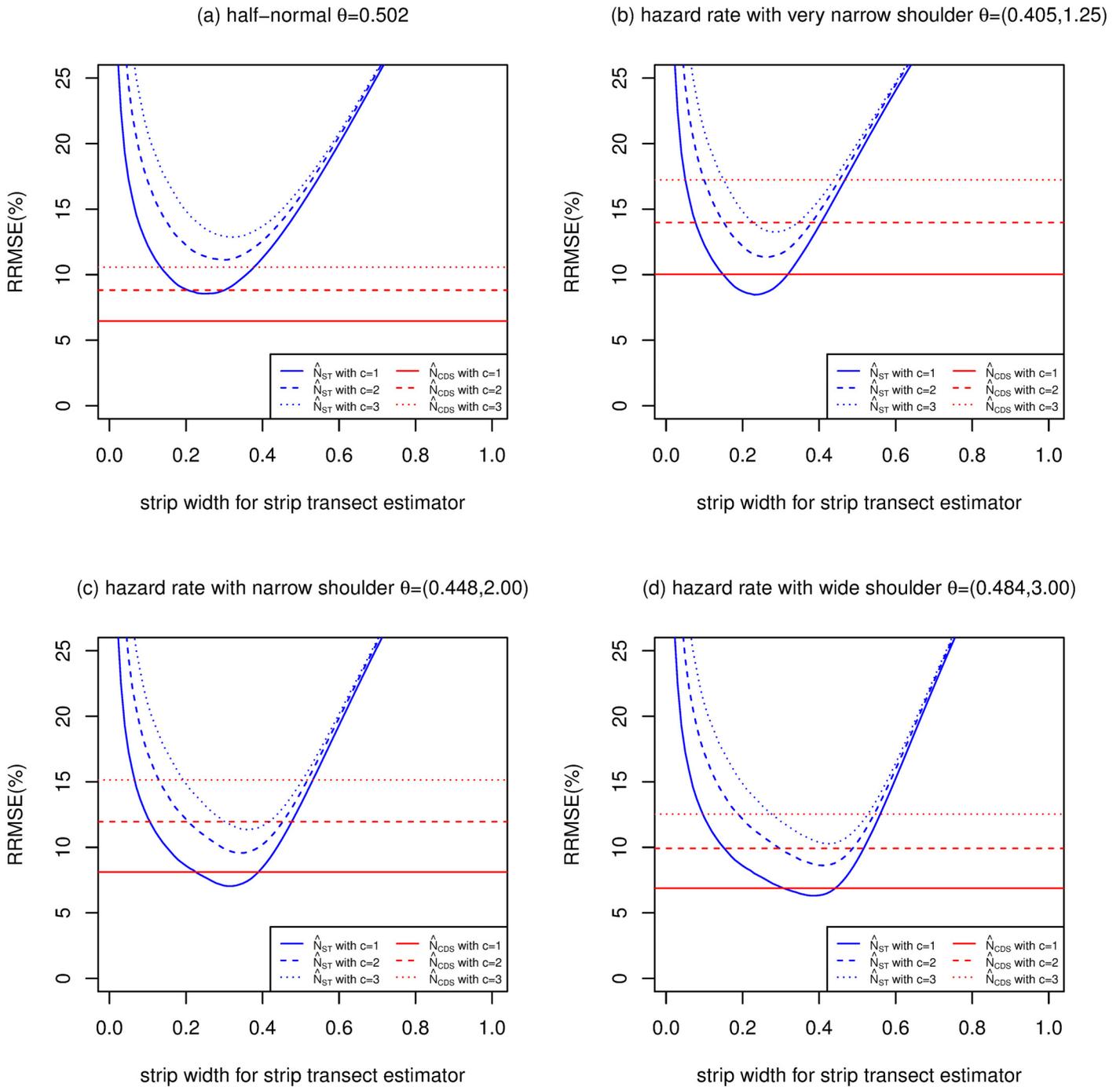
The results on  $F$  help to explain the relative mean squared errors of strip and CDS estimators in the simulation. When the number of detections  $n$  is sufficiently large (e.g. 400), CDS outperforms strip transect estimation unless the strip width is moderately close to its optimal value (within about  $\pm 25\%$ ). This is because variances of estimated abundances are then



**Fig 6. Relative root mean squared errors (RRMSE) (%) of CDS and strip transect estimators (strip widths ranging from 0 to 1) of  $N$ , for 4 detection functions, with  $E[n] = 200$ ,  $P = 0.1$  and  $\bar{g} = 0.6$ .**

doi:10.1371/journal.pone.0149298.g006

relatively small, so that bias considerations are paramount. For smaller  $n$ , the variance of the CDS estimator becomes more dominant, partly due to the penalty  $F$ . As a result, when the overdispersion factor is 2 and the hazard rate model applies, we find that strip transects give lower mean squared error (MSE) than CDS when:



**Fig 7. Relative root mean squared errors (RRMSE) (%) of CDS and strip transect estimators (strip widths ranging from 0 to 1) of  $N$ , for 4 detection functions, with  $E[n] = 400$ ,  $P = 0.1$  and  $\bar{g} = 0.6$ .**

doi:10.1371/journal.pone.0149298.g007

- $n = 50$  and the strip width is 0.1 or higher (with an optimal width of about 0.4), where a width of 1 corresponds to a typical detection range for distance sampling;
- $n = 100$  and the strip width is between 0.2 and 0.7, i.e. within about  $\pm 75\%$  of its optimal value of 0.42; or

- $n = 200$  and the strip width is within about 50% of its optimal value.

A Dirichlet process approach was able to generate non-uniform detections resulting in overdispersion factors of 1, 2 and 3. This approach has not been used in simulations in statistical ecology to the author’s knowledge. It would be applicable to any simulation where overdispersion or robustness relative to an assumed spatial model is of interest.

The simulation settings are more favourable to CDS than to strip transect estimators. In particular, it is assumed that the functional form used in CDS estimation matches the true detection function. This assumption will never be perfectly justified in reality, and its failure will impact on the performance of CDS estimators to some extent. Model selection or model averaging are often used to try to identify the correct model, but uncertainty about the model form will inflate variances and potentially lead to some bias as well. In contrast, strip transect estimators do not require a specification of the detection function. Of course, the properties of strip transect estimators depend on the mean detection rate in the strip, but beyond that there is no particular impact of one functional form rather than another. The simulations with  $c = 1$  are also favourable to both CDS and strip transects as they mean that distances are independent and uniformly distributed. This is relaxed to some extent by the overdispersed simulations with  $c$  equal to 2 and 3, however the Dirichlet process used is still centred on the uniform distribution. When there is a more severe departure from uniformity, CDS estimators will be biased, and strip transect estimators will also be affected to some degree.

The choice of methodology for assessing abundance, as well as the determination of required sample size, should be informed by consideration of all relevant biases and by the likely achievable precision. The results in this paper will help in this process, by providing a sample size formula reflecting the penalty due to unknown detection parameters, theoretical and simulation results on the size of this penalty, and a comparison of the mean squared errors of strip and line transect estimators in a wide-ranging simulation.

### Appendix: Proof of Theorem 1

Applying Eq (15), we approximate  $l = \log(L)$  by

$$l \approx N \log(N) - (N - n) \log(N - n) - n \log(n) + (N - n) \log(1 - P\bar{g}(\theta)) + n \log(P) - n \log(w) + \sum_{i=1}^n \log g(d_i; \theta) \tag{33}$$

The next step is to differentiate  $l$  to obtain the score function:

$$\frac{\partial l}{\partial N} = N \cdot N^{-1} + 1 \cdot \log(N) - (N - n) \cdot (N - n)^{-1} - \log(N - n) + \log(1 - P\bar{g}(\theta)) \tag{34}$$

$$= \log(N) - \log(N - n) + \log(1 - P\bar{g}(\theta)) \tag{35}$$

$$\frac{\partial l}{\partial \theta} = \sum_{i=1}^n g(d_i; \theta)^{-1} h(d_i; \theta) - (N - n)(1 - P\bar{g}(\theta))^{-1} P\bar{h}(\theta) \tag{36}$$

The MLE is obtained by setting Eqs (35) and (36) to 0. Firstly, set Eq (35) to 0 and exponentiate both sides:

$$1 = \hat{N}(\hat{N} - n)^{-1} \left(1 - P\bar{g}(\hat{\theta})\right) \tag{37}$$

which leads directly to

$$\hat{N} = n / \left\{ P\bar{g}(\hat{\theta}) \right\}. \tag{38}$$

Setting Eq (36) to 0, and then substituting for  $\hat{N}$  from Eq (38) gives an estimating equation for  $\theta$ :

$$0 = \sum_{i=1}^n g(d_i; \theta)^{-1} h(d_i; \theta) - (\hat{N} - n)(1 - P\bar{g}(\theta))^{-1} P\bar{h}(\theta) \tag{39}$$

$$= \sum_{i=1}^n g(d_i; \theta)^{-1} h(d_i; \theta) - (nP^{-1}\bar{g}(\theta)^{-1} - n)(1 - P\bar{g}(\theta))^{-1} P\bar{h}(\theta) \tag{40}$$

$$= \sum_{i=1}^n g(d_i; \theta)^{-1} h(d_i; \theta) - nP^{-1}\bar{g}(\theta)^{-1}(1 - P\bar{g}(\theta))(1 - P\bar{g}(\theta))^{-1} P\bar{h}(\theta) \tag{41}$$

$$= \sum_{i=1}^n g(d_i; \theta)^{-1} h(d_i; \theta) - n\bar{g}(\theta)^{-1} \bar{h}(\theta) \tag{42}$$

Results Eqs (38) and (42) give the Theorem result on the maximum likelihood estimators of  $N$  and  $\theta$ .

The Fisher Information is given by the variance of the score vector, the elements of which are given by Eqs (35) and (36). It can be written as

$$\mathcal{I}(N, \theta) = \begin{bmatrix} \mathcal{I}_{NN} & \mathcal{I}_{N\theta}^T \\ \mathcal{I}_{N\theta} & \mathcal{I}_{\theta\theta} \end{bmatrix} = \begin{bmatrix} \text{var}(\partial l / \partial N) & \text{cov}(\partial l / \partial N, \partial l / \partial \theta)^T \\ \text{cov}(\partial l / \partial N, \partial l / \partial \theta) & \text{var}(\partial l / \partial \theta) \end{bmatrix}. \tag{43}$$

The block elements of  $\mathcal{I}$  are easily derived. Some preliminary notes:

1. Let  $D$  be a random variable with density  $g(d; \theta) / \int_0^w g(u; \theta) du$  for  $0 \leq d \leq w$ .
2. The distances  $d_i$  follow the same distribution as  $D$ , conditional on  $n$ .
3.  $E(g(d_i; \theta)^{-1} h(d_i; \theta) | n) = \int_0^w g(d; \theta)^{-1} h(d; \theta) g(d; \theta) dd / \int_0^w g(u; \theta) du = \bar{h}(\theta) / \bar{g}(\theta)$ .

For the remainder of the proof, I will write  $g(u)$ ,  $\mathbf{h}(\mathbf{u})$ ,  $\bar{g}$  and  $\bar{\mathbf{h}}$  for readability, rather than  $g(u; \boldsymbol{\theta})$  etc. Using (a), (b) and (c), we obtain:

$$\mathcal{I}_{\theta\theta} = \text{var}(\partial l / \partial \boldsymbol{\theta}) = E\{\text{var}(\partial l / \partial \boldsymbol{\theta} | n)\} + \text{var}\{E(\partial l / \partial \boldsymbol{\theta} | n)\} \tag{44}$$

$$= E(n \text{var}(\mathbf{h}(D) / g(D))) + \text{var}\{n \bar{\mathbf{h}} \bar{g}^{-1} - (N - n)(1 - P\bar{g})^{-1} P \bar{\mathbf{h}}\} \tag{45}$$

$$= E[n \text{var}(\mathbf{h}(D) / g(D))] + \text{var}\{n \bar{\mathbf{h}} \bar{g}^{-1} (1 - P\bar{g})^{-1} (1 - P\bar{g} + P\bar{g}) + \text{const.}\} \tag{46}$$

$$= E[n \text{var}(\mathbf{h}(D) / g(D))] + \text{var}\{n \bar{\mathbf{h}} \bar{g}^{-1} (1 - P\bar{g})^{-1}\} \tag{47}$$

$$= NP\bar{g} \text{var}(\mathbf{h}(D) / g(D)) + \text{var}(n \bar{\mathbf{h}} \bar{\mathbf{h}}^T \bar{g}^{-2} (1 - P\bar{g})^{-2}) \tag{48}$$

$$= NP\bar{g} \text{var}(\mathbf{h}(D) / g(D)) + NP\bar{g} (1 - P\bar{g}) \bar{\mathbf{h}} \bar{\mathbf{h}}^T \bar{g}^{-2} (1 - P\bar{g})^{-2} \tag{49}$$

$$= NP\bar{g} \text{var}(\mathbf{h}(D) / g(D)) + NP \bar{\mathbf{h}} \bar{\mathbf{h}}^T \bar{g}^{-1} (1 - P\bar{g})^{-1} \tag{50}$$

$$\mathcal{I}_{N\theta} = \text{cov}(\partial l / \partial N, \partial l / \partial \boldsymbol{\theta}) \tag{51}$$

$$= E \text{cov}(\partial l / \partial N, \partial l / \partial \boldsymbol{\theta} | n) + \text{cov}\{E(\partial l / \partial N | n), E(\partial l / \partial \boldsymbol{\theta} | n)\} \tag{52}$$

$$= 0 - \text{cov}\{\log(N - n), n \bar{g}^{-1} \bar{\mathbf{h}} + n(1 - \bar{g})^{-1} \bar{\mathbf{h}}\} \tag{53}$$

$$= -\bar{g}^{-1} (1 - \bar{g})^{-1} \bar{\mathbf{h}} \text{cov}\{\log(N - n), n\} \tag{54}$$

As  $N \rightarrow \infty$   $n \xrightarrow{p} NP\bar{g}$ , so the right hand side of Eq (54) can be approximated using a first order Taylor Series of  $n$  about  $E[n] = NP\bar{g}$ :

$$\mathcal{I}_{N\theta} \approx -\bar{g}^{-1} (1 - P\bar{g})^{-1} P \bar{\mathbf{h}} \cdot \text{cov}\{\log(N - NP\bar{g}) - (N - NP\bar{g})^{-1} (n - NP\bar{g}), n\} \tag{55}$$

$$= \bar{g}^{-1} (1 - P\bar{g})^{-1} \bar{\mathbf{h}} N^{-1} (1 - P\bar{g})^{-1} \text{var}(n) \tag{56}$$

$$= \bar{g}^{-1} (1 - P\bar{g})^{-2} \bar{\mathbf{h}} N^{-1} NP\bar{g} (1 - NP\bar{g}) \tag{57}$$

$$= (1 - P\bar{g})^{-1} P \bar{\mathbf{h}} \tag{58}$$

The top-left element of  $\mathcal{I}$ ,  $\mathcal{I}_{NN}$ , can also be approximated by a first order Taylor Series about

$$n = N\bar{g}:$$

$$\mathcal{I}_{NN} = \text{var}(\partial l / \partial N) = \text{var}(\log(N - n)) \quad (59)$$

$$\approx \text{var}\{-\log(N - NP\bar{g}) + (N - NP\bar{g})^{-1}(n - NP\bar{g})\} \quad (60)$$

$$= N^{-2}(1 - P\bar{g})^{-2}\text{var}(n) \quad (61)$$

$$= N^{-2}(1 - P\bar{g})^{-2}NP\bar{g}(1 - P\bar{g}) \quad (62)$$

$$= N^{-1}(1 - P\bar{g})^{-1}P\bar{g} \quad (63)$$

## Supporting Information

**S1 Fig. Simulation results for P = 0.2 with E[n] = 100.**  
(EPS)

**S2 Fig. Simulation results for P = 0.25 with E[n] = 100.**  
(EPS)

**S3 Fig. Simulation results for P = 0.3 with E[n] = 100.**  
(EPS)

**S4 Fig. Simulation results for P = 0.5 with E[n] = 100.**  
(EPS)

**S1 Code. R program used to produce simulation results.**  
(R)

## Acknowledgments

I am grateful to Geoff Robertson and Michelle Dawson who provided background on sampling of brumbies and to Alan Welsh for his insights and encouragement. Anonymous referees made comments which improved the paper. All views are the author's own.

## Author Contributions

Conceived and designed the experiments: RC. Performed the experiments: RC. Analyzed the data: RC. Contributed reagents/materials/analysis tools: RC. Wrote the paper: RC.

## References

1. Walter MJ, Hone J. A comparison of 3 aerial survey techniques to estimate wild horse abundance in the Australian Alps. *Wildlife Society Bulletin*. 2003;p. 1138–1149.
2. Kruger J, Reilly B, Whyte I. Application of distance sampling to estimate population densities of large herbivores in Kruger National Park. *Wildlife Research*. 2008; 35(4):371–376. doi: [10.1071/WR07084](https://doi.org/10.1071/WR07084)
3. Oppel S. Using distance sampling to quantify Odonata density in tropical rainforests. *International Journal of Odonatology*. 2006; 9(1):81–88. doi: [10.1080/13887890.2006.9748265](https://doi.org/10.1080/13887890.2006.9748265)
4. Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L. *Advanced Distance Sampling*. Oxford: Oxford University Press; 2004.
5. Hedley SL, Buckland ST. Spatial models for line transect sampling. *Journal of Agricultural, Biological, and Environmental Statistics*. 2004; 9(2):181–199. doi: [10.1198/1085711043578](https://doi.org/10.1198/1085711043578)

6. Johnson DS, Laake JL, Ver Hoef JM. A model-based approach for making ecological inference from distance sampling data. *Biometrics*. 2010; 66(1):310–318. doi: [10.1111/j.1541-0420.2009.01265.x](https://doi.org/10.1111/j.1541-0420.2009.01265.x) PMID: [19459840](https://pubmed.ncbi.nlm.nih.gov/19459840/)
7. Miller DL, Burt ML, Rexstad EA, Thomas L. Spatial models for distance sampling data: recent developments and future directions. *Methods in Ecology and Evolution*. 2013; 4(11):1001–1010. doi: [10.1111/2041-210X.12105](https://doi.org/10.1111/2041-210X.12105)
8. Oedekoven C, Buckland S, Mackenzie M, King R, Evans K, Burger L Jr. Bayesian methods for hierarchical distance sampling models. *Journal of Agricultural, Biological, and Environmental Statistics*. 2014;p. 1–21.
9. Buckland S, Anderson D, Burnham K, Laake J, Borchers D. *Introduction to Distance Sampling: Estimating Abundance of Biological Populations*. Oxford: Oxford University Press; 2001.
10. Barry SC, Welsh A. Distance sampling methodology. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2001; 63(1):23–31. doi: [10.1111/1467-9868.00274](https://doi.org/10.1111/1467-9868.00274)
11. Melville G, Welsh A. Line transect sampling in small regions. *Biometrics*. 2001; 57(4):1130–1137. doi: [10.1111/j.0006-341X.2001.01130.x](https://doi.org/10.1111/j.0006-341X.2001.01130.x) PMID: [11764253](https://pubmed.ncbi.nlm.nih.gov/11764253/)
12. Fewster RM, Laake JL, Buckland ST. Line transect sampling in small and large regions. *Biometrics*. 2005; 61(3):856–859. doi: [10.1111/j.1541-0420.2005.00413\\_1.x](https://doi.org/10.1111/j.1541-0420.2005.00413_1.x) PMID: [16135038](https://pubmed.ncbi.nlm.nih.gov/16135038/)
13. Melville GJ, Tracey JP, Fleming PJ, Lukins BS. Aerial surveys of multiple species: critical assumptions and sources of bias in distance and mark–recapture estimators. *Wildlife Research*. 2008; 35(4):310–348. doi: [10.1071/WR07080](https://doi.org/10.1071/WR07080)
14. Melville GJ, Welsh AH. Model-based prediction In ecological surveys including those with incomplete detection. *Australian & New Zealand Journal of Statistics*. 2014; 56(3):257–281. doi: [10.1111/anzs.12084](https://doi.org/10.1111/anzs.12084)
15. Burnham KP, Anderson DR, Laake JL. Efficiency and bias in strip and line transect sampling. *The Journal of Wildlife Management*. 1985;p. 1012–1018. doi: [10.2307/3801387](https://doi.org/10.2307/3801387)
16. Azhar B, Zakaria M, Yusof E, Leong PC. Efficiency of fixed-width transect and line-transect-based distance sampling to survey red junglefowl (*Gallus gallus spadiceus*) in Peninsular Malaysia. *Journal of Sustainable Development*. 2009; 1(2):63–73. doi: [10.5539/jsd.v1n2p63](https://doi.org/10.5539/jsd.v1n2p63)
17. Welsh A. Incomplete detection in enumeration surveys: whither distance sampling? *Australian & New Zealand Journal of Statistics*. 2002; 44(1):13–22. doi: [10.1111/1467-842X.00204](https://doi.org/10.1111/1467-842X.00204)
18. Burnham KP, Anderson DR. *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*. New York: Springer-Verlag; 2002.
19. Fewster R, Jupp P. Inference on population size in binomial detectability models. *Biometrika*. 2009; 96(4):805–820. doi: [10.1093/biomet/asp051](https://doi.org/10.1093/biomet/asp051)
20. Thomas L, Buckland ST, Rexstad EA, Laake JL, Strindberg S, Hedley SL, et al. Distance software: design and analysis of distance sampling surveys for estimating population size. *Journal of Applied Ecology*. 2010; 47(1):5–14. doi: [10.1111/j.1365-2664.2009.01737.x](https://doi.org/10.1111/j.1365-2664.2009.01737.x) PMID: [20383262](https://pubmed.ncbi.nlm.nih.gov/20383262/)
21. Marques FF, Buckland ST. Incorporating covariates into standard line transect analyses. *Biometrics*. 2003; 59(4):924–935. doi: [10.1111/j.0006-341X.2003.00107.x](https://doi.org/10.1111/j.0006-341X.2003.00107.x) PMID: [14969471](https://pubmed.ncbi.nlm.nih.gov/14969471/)
22. Marques TA, Thomas L, Fancy SG, Buckland ST, Handel C. Improving estimates of bird density using multiple-covariate distance sampling. *The Auk*. 2007; 124(4):1229–1243. doi: [10.1642/0004-8038\(2007\)124%5B1229:IEOBDU%5D2.0.CO;2](https://doi.org/10.1642/0004-8038(2007)124%5B1229:IEOBDU%5D2.0.CO;2)
23. Buckland ST, Oedekoven CS, Borchers DL. Model-based distance sampling. *Journal of Agricultural, Biological, and Environmental Statistics*. 2015;p. 1–18.
24. Borchers DL, Buckland ST. *Estimating Animal Abundance: Closed Populations*. London: Springer-Verlag; 2002.
25. Lehmann EL. *Elements of Large-Sample Theory*. New York: Springer; 1999.
26. Clark RG. *Efficiency and Robustness in Distance Sampling*; 2015. National Institute for Applied Statistics Research Australia (NIASRA) Working Paper 15–15. Available from: <http://niasra.uow.edu.au/publications/UOW185981.html>.
27. Harville DA. *Matrix Algebra from a Statistician's Perspective*. New York: Springer; 1997.
28. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria; 2013. <http://www.R-project.org/>.
29. Johnson NL, Kotz S, Balakrishnan N. *Discrete Multivariate Distributions*. Wiley New York; 1997.
30. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian data analysis*. 3rd ed. Taylor & Francis; 2014.

31. Miller DL. Distance: a simple way to fit detection functions to distance sampling data and calculate abundance/density for biological populations; 2014. R package version 0.9. Available from: <http://CRAN.R-project.org/package=Distance>.
32. Clark RG. distance.sample.size: Calculates study size required for distance sampling; 2016. R package version 0.0. Available from: <http://CRAN.R-project.org/package=distance.sample.size>.