Bias in estimation of adult survival and asymptotic population growth rate caused by undetected capture heterogeneity

David Fletcher[1]*, Jean-Dominique Lebreton[2], Lucile Marescot[2], Michael Schaub[3], Olivier Gimenez[2], Steve Dawson[4] and Elisabeth Slooten[5]

[1]*Department of Maths and Statistics, University of Otago, P.O. Box 56, Dunedin, New Zealand;* [2]*Centre d'Ecologie Fonctionnelle et Evolutive, UMR 5175, CNRS, 1919 Route de Mende, 34293 Montpellier Cedex 5, France;* [3]*Swiss Ornithological Institute, 6024 Sempach, Switzerland;* [4]*Department of Marine Science, University of Otago, P.O. Box 56, Dunedin, New Zealand; and* [5]*Department of Zoology, University of Otago, P.O. Box 56, Dunedin, New Zealand*

Summary

1. Mark–recapture studies are often used to estimate adult survival probability (ϕ), which is an important demographic parameter for long-lived species, as it can have a large impact on the population growth rate. We consider the impact of variation in capture probability among individuals (capture heterogeneity) on the estimation of ϕ from a mark–recapture study and thence on estimation of the asymptotic population growth rate (λ).

2. We review the mechanisms by which capture heterogeneity arises, methods of allowing for it in the analysis, and use simulation to assess the power of detecting three types of capture heterogeneity (two-group heterogeneity, trap-response and temporary emigration) using standard mark–recapture lack-of-fit tests.

3. We use simulation to assess the bias that can arise in the estimation of ϕ from a mark–recapture study when we do not allow for capture heterogeneity. Using a generic population model, we assess the effect this bias has on estimation of λ.

4. We use our results on the power of the lack-of-fit tests, together with a measure of the size of the bias relative to the standard error of the estimate of ϕ, to assess which situations might lead to an important level of undetected bias. Our results suggest that undetected bias is not likely to be an issue when there is trap-response, owing to the lack-of-fit tests having sufficient power to detect any trap-response that could lead to non-negligible bias. For two-group heterogeneity, the worst bias generally occurs when the difference between the capture probabilities for the two groups is moderate and both capture probabilities are low. For temporary emigration, the worst bias generally occurs when the rate of emigration and the capture probability are both low.

5. We illustrate the issues for conservation management using data from studies of Hector's dolphin (*Cephalorhynchus hectori*) in New Zealand and wolves (*Canis lupus*) in France.

6. Previous studies have suggested that capture heterogeneity will generally lead to a relatively small bias in the estimate of ϕ. However, given the high sensitivity of the asymptotic population growth rate to adult survival, a small bias in ϕ might lead to nontrivial bias in the estimate of λ.

Key-words: bias, capture heterogeneity, mark–recapture, maximum population growth rate, population growth rate, population model, survival

Introduction

For a long-lived species, reliable estimation of adult survival probability (ϕ) can be crucial to understanding population dynamics and hence to effective management of a population (Russell 1999; Caswell 2001). Typically, ϕ is estimated from data collected in a mark–recapture study (e.g. Lebreton *et al.* 1992). Although we might sometimes be interested in annual estimates, for long-lived species the between-year variation in adult survival is generally low, so we focus attention on the case where $\hat{\phi}$ comes from fitting the Cormack–Jolly–Seber

*Correspondence author. E-mail: dfletcher@maths.otago.ac.nz
Correspondence site: http://www.respond2articles.com/MEE/

(CJS) model (ϕ, p_t), in which ϕ is constant and capture probability is time dependent.

To provide useful information for the management of a population, $\hat{\phi}$ can be used in a population model, together with estimates of other demographic parameters. Alternatively, a population model might be fitted directly to all the available data, typically mark–recapture data, census data and data on reproductive success (Buckland *et al.* 2007; Schaub *et al.* 2007). The type of population model used will vary according to the species, the available data and the management context. For example, it might be deterministic and density independent or it might include demographic stochasticity, environmental stochasticity and/or density dependence. Likewise, the outcomes from the population model that are of interest will vary; we might focus on the asymptotic growth rate, the probability of quasi-extinction over some period or the time taken for the population to reach a specified fraction of its putative carrying capacity. Although much of what follows could easily be extended to more complex cases, for the sake of simplicity we focus on the case where $\hat{\phi}$ will be used in a deterministic, density-independent population model to estimate the asymptotic population growth rate (λ), where all the demographic parameters are estimated separately.

It has long been recognized that the reliability of $\hat{\phi}$ can be affected by variation among individuals in capture probability ('capture heterogeneity', which we denote as CH). Generally, CH has been found to induce a small negative bias in ϕ, although the bias can sometimes be positive (Nichols & Pollock 1983; Hwang & Chao 1995). The bias increases as mean capture probability decreases and as CH increases (Carothers 1979; Hwang & Chao 1995; Pledger, Pollock & Norris 2003). If CH is strong and some individuals have very low capture probabilities, it is also possible to detect apparent age effects in ϕ (Prévot-Julliard, Lebreton & Pradel 1998), in a fashion similar to that induced by transients (Pradel, Cooch & Cooke 1995). Given that CH is likely to occur for many populations, we can expect some bias in $\hat{\phi}$ if our model does not allow for it (Cormack 1972; Carothers 1973, 1979; Gilbert 1973; Nichols and Pollock 1983; Stromborg *et al.* 1988; Pollock *et al.* 1990; Hwang & Chao 1995; Manly, McDonald, & McDonald 1999; Pledger, Pollock & Norris 2003; Devineau, Choquet & Lebreton 2006; Cubaynes *et al.* 2010).

Carothers (1973) suggested that the bias in $\hat{\phi}$ caused by CH will often be small, especially when compared with the bias CH causes to estimates of abundance. However, if $\hat{\phi}$ is used to estimate λ, the high sensitivity of λ to ϕ (Lebreton & Clobert 1991; Russell 1999; Caswell 2001) means that an apparently small bias in $\hat{\phi}$ might lead to an important effect on $\hat{\lambda}$. The motivation for this work came from our desire to consider those levels of undetected CH for which the bias in $\hat{\phi}$ has practical consequences for management. We therefore consider the amount of bias in $\hat{\phi}$ that arises when the power of the standard lack-of-fit (LOF) tests is low. In addition, we ignore those cases where the undetected bias is small relative to the standard error of $\hat{\phi}$ (Carothers 1973).

Several important, related issues are outside the scope of this paper. First, if marked and unmarked individuals are equally likely to be captured, we can use mark–recapture data to obtain an overall estimate of the realized population growth rate (λ_e) (Pradel 1996). The type of CH present appears to determine whether there is any significant bias in the estimate of λ_e (Hines & Nichols 2002; Marescot *et al.* in press). Second, we do not consider the bias in $\hat{\phi}$ caused by heterogeneity in the survival probabilities (Pledger, Pollock, & Norris 2003, 2010) nor by transience (Pradel *et al.* 1997). Third, we do not consider the impact of CH on estimation of the precision of $\hat{\phi}$; CH may lead to underestimation of the standard error of $\hat{\phi}$, and this can be allowed for to some extent by the use of $\hat{c}$, the standard correction for overdispersion (Anderson, Burnham & White 1994). As we are considering only constant survival, we do not address the issue of the potential variation in bias over time (Lebreton 1995).

We use simulation to assess the power of standard LOF tests for the global CJS model (ϕ_t, p_t) and the amount of bias in $\hat{\phi}$ when undetected CH is present. We then consider the impact such bias has on $\hat{\lambda}$. We provide two illustrative examples, discuss our findings and make recommendations. In the Appendix S1, we provide an overview of possible biological mechanisms for CH and a review of methods available for dealing with CH in the context of open population mark–recapture models.

Power of LOF tests for the (ϕ_t, p_t) model

Given that we want to assess the amount of undetected bias, it is natural to first consider the power of standard LOF tests of the global CJS model (ϕ_t, p_t). In doing so, we focus on the tests available in U-CARE, Tests 2.CL, 2.CT, 3.SM and 3.SR (Choquet *et al.* 2009), as they deal more effectively with sparse data than those in the RELEASE option within Program Mark (White & Burnham 1999); see Pradel, Gimenez & Lebreton (2005). Pollock, Nichols & Hines (1985) used simulation to assess the power of several LOF tests when the study population consists of two groups of individuals, with all individuals in a group having the same capture probability (hereafter referred to as 'two-group CH'), but did not consider other types of CH. One standard approach is to assess the significance of the overall test statistic, i.e. the sum of the 2.CL, 2.CT, 3.SM and 3.SR statistics. Manly, McDonald & McDonald (1999) used simulation to assess the power of this approach for a range of situations, including two-group CH and temporary emigration, and concluded that in general, the power was very low. We therefore consider the power of individual tests to detect particular forms of CH.

We focus on Tests 2.CT and 3.SR (Pradel, Gimenez & Lebreton 2005), as 2.CL and 3.SM are often difficult to interpret and, in our experience, are less sensitive to heterogeneity. 2.CT tests the hypothesis that all individuals known to be alive in years i and $i+1$ have the same probability of recapture in year $i+1$, regardless of whether or not they were captured

in year i; 3.SR tests the hypothesis that all individuals captured in year i have the same probability of being recaptured later, regardless of their prior capture history. For simplicity of presentation, we focus on the results for the Pearson's X^2 versions of the tests, as those for the likelihood ratio versions (G^2 in U-CARE) were very similar.

Given our overview of possible mechanisms for CH (Appendix S1), we considered the following settings:

1. Two-group CH in which all individuals in group i had capture probability p_i ($i = 1, 2$). For simplicity, we assumed that the two groups were of equal size. Carothers (1973) provides a discussion of the motivation for considering just two groups. We considered all combinations of $p_i = 0.1, 0.2, ..., 0.9$ ($i = 1, 2$).

2. Trap-response, in which an individual not captured in the previous year had capture probability p_0, while one captured in the previous year had capture probability p_1. We considered all combinations of $p_i = 0.1, 0.2, ..., 0.9$ ($i = 0, 1$).

3. Temporary emigration, in which individuals were assumed to be in either area 1 or 2, with a capture probability of p in area 1 and of zero in area 2. A similar situation arises when individuals skip breeding, e.g. as a consequence of the death of their partner. The probability of movement from area 1 to 2 (a) was assumed to be equal to the probability of movement from area 2 to 1. We assumed that initially half the population was in area 1 and half in area 2. We considered all combinations of $p = 0.1, 0.2, ..., 0.9$ and $a = 0.1, 0.2, ..., 0.9$. For an individual that is alive in year i, the probability of being captured that year is $\pi_i p$, where $\pi_i = (1 - a)\pi_{i-1} + a(1 - \pi_{i-1})$ is the probability that the individual is in the study area in year i. For the special case $a = 0.5$, $\pi_i = 0.5$ for all individuals (and all years), and there is no CH.

We used simulation to estimate the power of each test as follows. Given a true ϕ (assumed to be the same for each year), length of study, number of individuals marked each year (R) and a specified form of CH, we generated 2000 sets of mark–recapture data. The individual power of Tests 2.CT and 3.SR was estimated by the proportion of data sets in which that test was significant (at the 5% level). Use of 2000 simulations meant that the standard error of each of these estimates would be at most 0.011.

Figure 1 provides a visual summary of the estimated power of Test 2.CT to detect LOF caused by each type of CH, for a 10-year study in which $R = 50, 100$ or 500 and $\phi = 0.90$. For all three types of CH, the power increases with R, as expected. For two-group CH, the power increases with $|p_1 - p_2|$ and with p_1 and p_2 for a fixed value of $|p_1 - p_2|$, again as expected. For trap-response, the power increases very quickly with $|p_0 - p_1|$ and is generally high, even when $R = 500$. Interestingly, for $R = 50$ or 100 and a fixed value of $|p_0 - p_1|$, the power is generally higher when p_0 and p_1 are lower. For temporary emigration, the power is close to 0.05 when $a = 0.5$, as there is no CH. For other values of a, the power increases with p. For $a > 0.5$, the power increases with a, while for $a < 0.5$, the pattern is less clear-cut. Overall, it appears that Test 2.CT has good power to detect trap-response, but may have low

power to detect moderate two-group CH or temporary emigration, especially when R and/or the capture probabilities are low.

For Test 3.SR, the power was uniformly low for both trap-response and temporary emigration; for simplicity, we therefore present the results for just two-group CH. Figure 2 provides a summary of these results, again for a 10-year study in which $R = 50, 100$ or 500 and $\phi = 0.90$. As for Test 2.CT, the power increases with both R and $|p_1 - p_2|$. Unlike Test 2.CT, however, the power decreases with p_1 and p_2 for a fixed value of $|p_1 - p_2|$. Thus, Test 3.SR might detect LOF caused by two-group CH in those cases where Test 2.CT does not, and vice versa.

Bias in adult survival

For each type of study discussed above, and for each type of CH, the bias B_ϕ was estimated by the difference between the mean of $\hat{\phi}$ (over the simulated data sets) and the true value of ϕ. For all three types of CH, B_ϕ is largely unaffected by the value of R; for simplicity of presentation, we only present the results for $R = 500$. Figure 3 provides a summary of these results. For two-group CH, the bias is generally negative and gets worse as $|p_1 - p_2|$ increases. In addition, it becomes worse as p_1 and p_2 decrease, for a fixed value of $|p_1 - p_2|$. For trap-response, the bias becomes worse as $|p_0 - p_1|$ increases. When $p_1 > p_0$, the bias is negative, and vice versa. For temporary emigration, the bias is negative for $a < 0.5$, positive for $a > 0.5$ and becomes worse as $|a - 0.5|$ increases. Interestingly, there appears to be very little effect of p on the bias. Overall, the worst bias occurs when there is trap-response, $p_1 > p_0$ and p_0 is low.

To allow for the fact that the bias is likely to be of concern only when the power of Tests 2.CT and 3.SR is low, we then considered only those cases in which the power of both tests was < 0.5. Henceforth, we use the term 'undetected bias' as shorthand for 'bias that is likely to go undetected', in the sense implied by this criterion. We further reduced the number of cases by allowing for the fact that the standard error of $\hat{\phi}$ (SE_ϕ) might be large enough to make the amount of bias irrelevant (Carothers 1973). Thus, we additionally restricted attention to those cases for which $|B_\phi|/SE_\phi \geq 0.5$ (Burnham *et al.* 1987).

For trap-response, none of the cases satisfied all three criteria; for two-group CH and temporary emigration, Figure 4 shows the bias for the cases satisfying the criteria. For two-group CH, these cases correspond to moderate differences between p_1 and p_2, as a small difference leads to $|B_\phi|/SE_\phi < 0.5$ and a large difference to the power of the LOF tests being high (Figs 1 and 2). The bias is always negative and is worst (-0.038), when $R = 50$ and the capture probabilities are 0.2 and 0.6. For $R = 100$, the worst bias is -0.033, when one of the capture probabilities is 0.1 and the other 0.3; for $R = 500$, it is -0.013, when one of the capture probabilities is 0.1 and the other 0.2. For temporary emigration, the cases satisfying the three criteria generally correspond to lower values of a and p, particularly for higher values of R. The bias can be

Fig. 1. Estimates of the power of Test 2.CT to detect LOF for three types of CH, for a 10-year study in which the number of individuals marked each year (R) is 50, 100 or 500, and adult survival probability (ϕ) is 0·9.
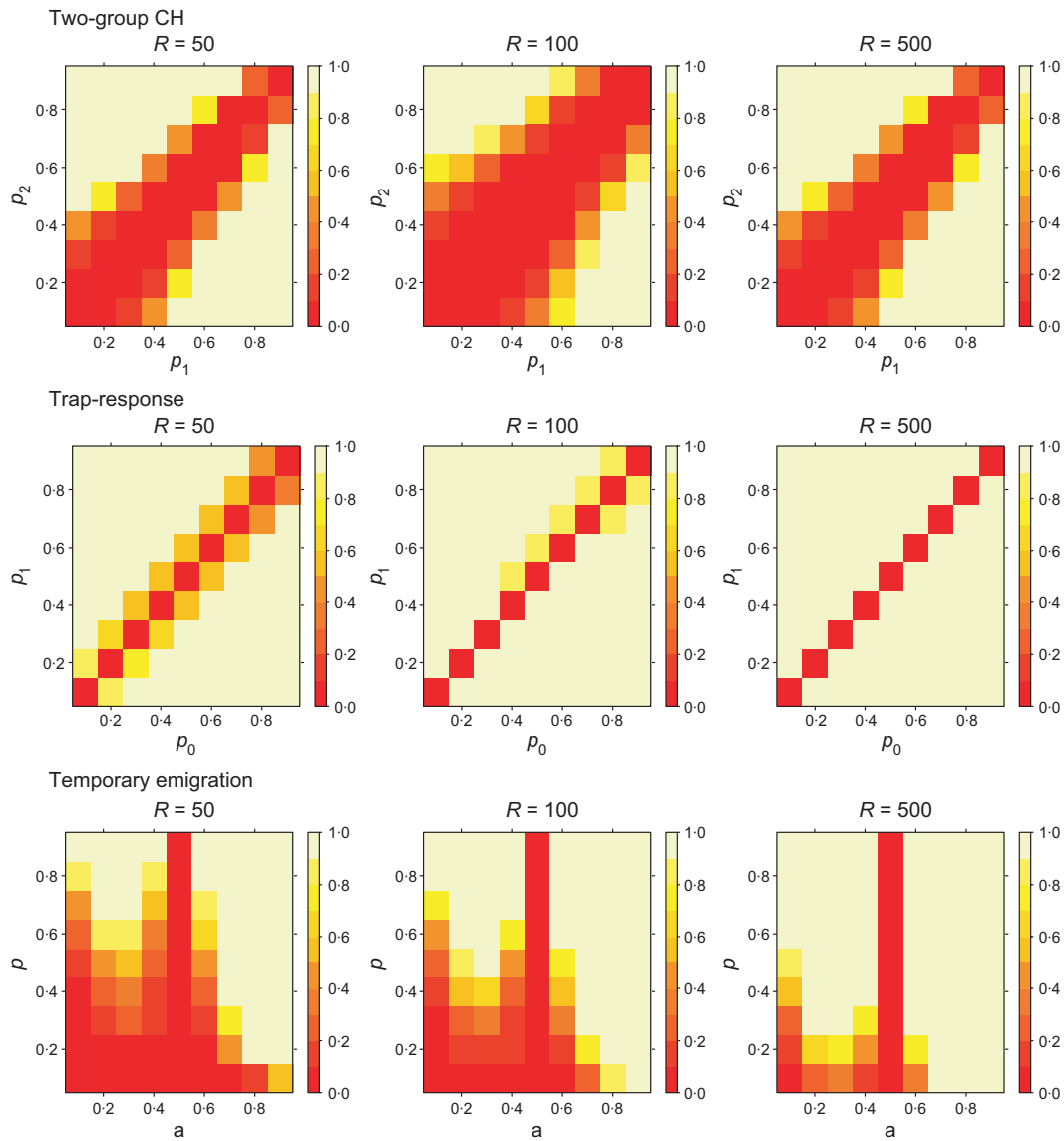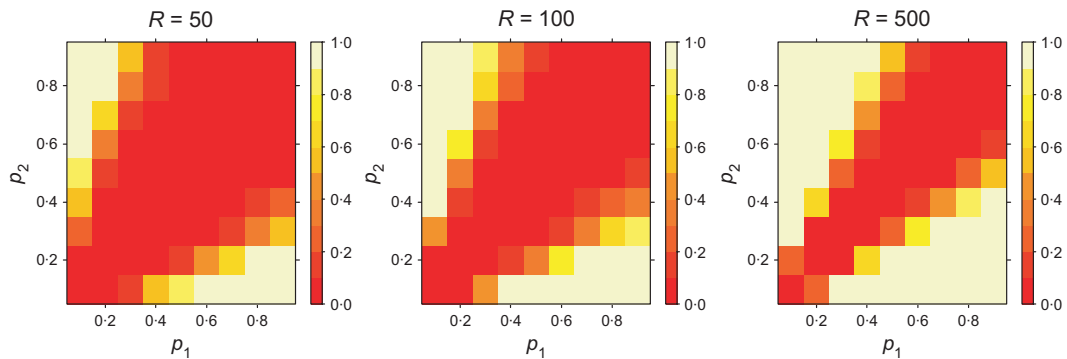

Fig. 2. Estimates of the power of Test 3.SR to detect LOF for two-group CH, for a 10-year study in which the number of individuals marked each year (R) is 50, 100 or 500 and adult survival probability (ϕ) is 0·9.
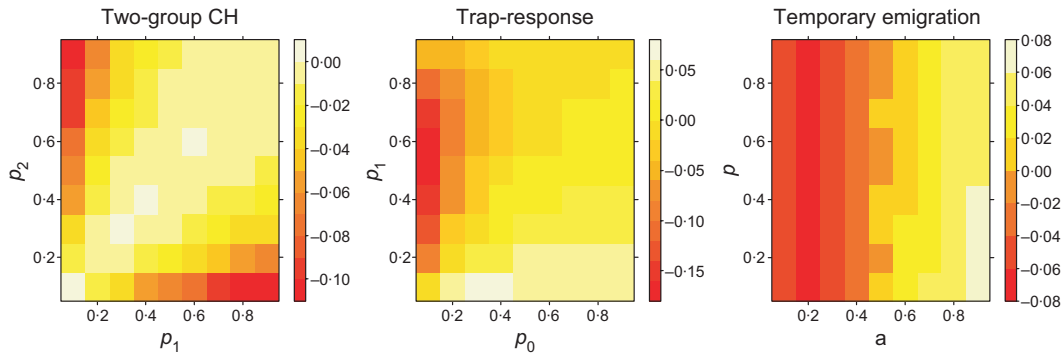
Fig. 3. Estimates of B_ϕ for three types of CH, for a 10-year study in which the number of individuals marked each year (R) is 500 and adult survival probability (ϕ) is 0·9.
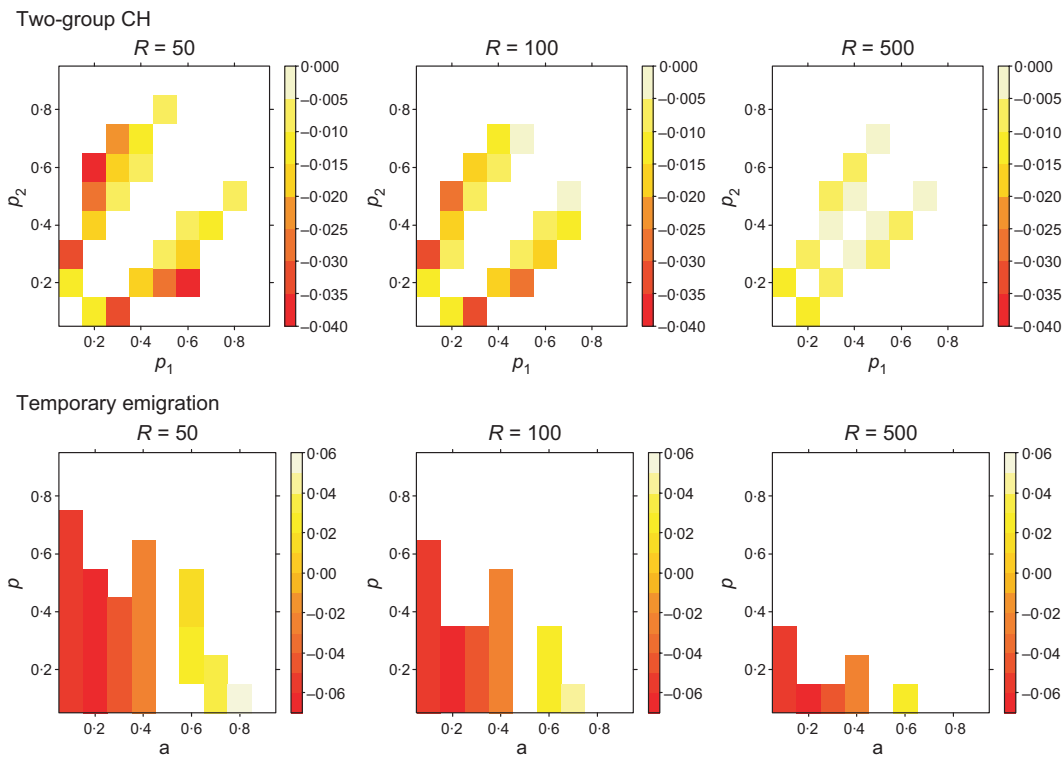

Fig. 4. Estimates of B_ϕ for two types of CH, for a 10-year study in which the number of individuals marked each year (R) is 50, 100 or 500 and adult survival probability (ϕ) is 0·9, for thoses cases where the estimated power of both Test 2.CT and Test 3.SR is $< 0·5$, and $|B_\phi| / SE_\phi > 0·5$ (white indicates cases where these criteria are not satisfied).

positive or negative and is worst ($-0·066$) when $a = 0·2$ and p is low, the exact values of p depending on R. For $R = 50$, even when $p = 0·5$, the bias can be as much as $-0·064$ ($a = 0·2$), while for $R = 100$, it can be $-0·057$ ($a = 0·1$).

Bias in asymptotic population growth rate

In general, the bias in $\hat{\lambda}$ (B_λ) caused by bias in $\hat{\phi}$ will depend on the form of the population model and the estimates of the other demographic parameters in the model, such as reproductive success, juvenile survival and age at first reproduction. To provide general guidance, we focus on a generic model for

which it is possible to obtain an analytical approximation to B_λ that depends only on the age at first reproduction (α), the ratio ϕ/λ and B_ϕ. The model involves a prebreeding census and contains $\alpha - 1$ juvenile stages and an adult stage. Both adult survival and fertility are assumed to be independent of age, and fertility includes survival to age 1. For many long-lived species, dispersal of juveniles means that it is difficult to obtain age-specific estimates of survival prior to reaching maturity. The model therefore involves a single overall survival probability for the newborn and juvenile stages, which we denote as ϕ_J. The $\alpha \times \alpha$ projection matrix for the model is given by

$$A = \begin{bmatrix} 0 & 0 & \cdots & \cdots & 0 & b\phi_J \\ \phi_J & 0 & \cdots & \cdots & 0 & 0 \\ 0 & \phi_J & 0 & \cdots & 0 & 0 \\ \vdots & 0 & \phi_J & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 & 0 \\ 0 & 0 & \cdots & 0 & \phi_J & \phi \end{bmatrix} \qquad \text{eqn 1}$$

where b is the reproductive rate (newborns per adult). The dominant eigenvalue of A is our estimate of λ.

Assuming that the estimates of the other demographic parameters are unbiased, a first-order Taylor series approximation to B_λ is given by (Daley 1979)

$$B_\lambda \approx s_\phi B_\phi \qquad \text{eqn 2}$$

where $s_\phi = \partial\lambda/\partial\phi$ is the sensitivity of λ to ϕ (Caswell 2001), evaluated at the true values of the parameters. It is worth noting that $B_\lambda \neq 0$ even when $B_\phi = 0$ (Boyce 1977; Daley 1979), as λ is a nonlinear function of the demographic parameters, and eqn 2 ignores higher-order terms. However, as these terms are usually much smaller than $s_\phi B_\phi$, use of eqn 2 should be reliable (Houllier, Lebreton, & Pontier 1989).

For the model in eqn 1, we can write (Gaillard *et al.* 2005, Heppell *et al.* 2000)

$$s_\phi = \frac{\lambda}{\phi}\left(1 - \frac{\alpha}{T}\right) = \left(1 + (\alpha - 1)\left(1 - \frac{\phi}{\lambda}\right)\right)^{-1} \qquad \text{eqn 3}$$

where $T = \alpha + \frac{\phi}{\lambda - \phi}$ is generation time. Figure 5 shows the values of s_ϕ for values of α and ϕ/λ that cover the range likely to apply to long-lived species. As can be deduced from the form of eqn 3, s_ϕ is always positive, meaning that B_λ will have the same sign as B_ϕ. In addition, s_ϕ, and therefore B_λ, increases as α decreases and as ϕ/λ increases.
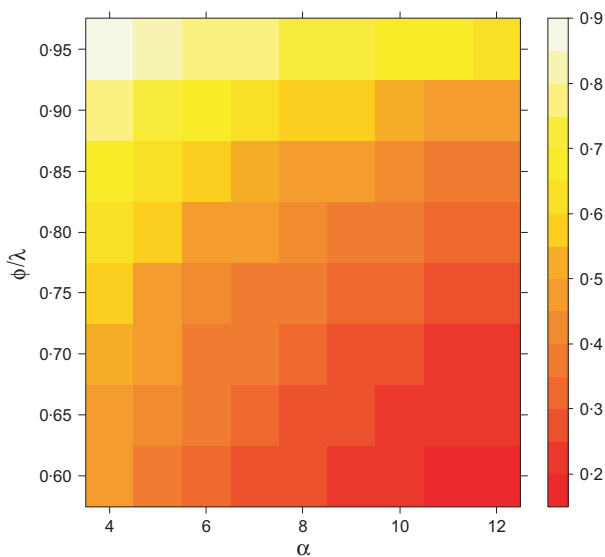

Fig. 5. Sensitivity (s_ϕ) of asymptotic population growth rate (λ) to the value of adult survival probability (ϕ), versus age at first reproduction (α) and ϕ/λ, using the population model in Eq. 1.

We can also use s_ϕ to gauge the difference between the estimates of λ obtained from alternative estimates of ϕ, i.e.

$$\hat{\lambda}_1 - \hat{\lambda}_2 \approx s_\phi\left(\hat{\phi}_1 - \hat{\phi}_2\right) \qquad \text{eqn 4}$$

where $\hat{\lambda}_i$ is the estimate of λ obtained using $\hat{\phi}_i (i = 1, 2)$.

Examples

Undetected bias in $\hat{\phi}$ might lead to non-negligible bias in $\hat{\lambda}$, which in turn could affect the management of a species. For example, we might think that the population is in decline ($\lambda < 1$) when it is stable or growing ($\lambda \geq 1$), and vice versa. We now consider two examples that illustrate these issues.

HECTOR'S DOLPHINS IN NEW ZEALAND

We illustrate the potential for bias in the presence of trap-response using a study of Hector's dolphin (*Cephalorhynchus hectori*). This species has been studied extensively around Banks Peninsula, New Zealand. The data we consider come from a long-term photo-ID survey of individually identifiable dolphins, i.e. captures are photographic (Slooten, Dawson, & Lad 1992; Dawson & Slooten 1993; Cameron *et al.* 1999). For individuals considered to be readily identifiable, the capture histories obtained over an eleven-year period are summarized in the m-array shown in Table 1.

Table 2 summarizes the LOF test results from U-CARE. Test 2. CT is clearly significant, and the overall test with this component removed shows no lack-of-fit, suggesting that we should fit a trap-response model. None of the other tests is close to significant. Note that the overall test is not quite significant at the 5% level. There appears to be no need to adjust for overdispersion: after removal of 2.CT, we have $\hat{c} = 22 \cdot 6/24 = 0.94$ (Pradel, Gimenez, & Lebreton 2005). The fact that use of a trap-response model is suggested does not necessarily imply a physical trap-response mechanism in this study (Appendix S1).

Table 1. M-array for the Hector's dolphin study, where zeros have been removed for ease of presentation

Year	Number captured	Year of first recapture									
		2	3	4	5	6	7	8	9	10	11
1	8	3	1	1		1					
2	35		26	1			2	1		1	
3	46			31	2	3	3	1			
4	47				9	5	6	4	1	1	1
5	15					5	3	1			1
6	18							10	1	1	1
7	31							17	2	1	1
8	37								15	4	5
9	23									11	4
10	19										10

Table 2. Summary of U-CARE LOF tests for the Hector's dolphin study

	df	X^2	P
2.CL	8	5·4	0·712
2.CT	8	21·5	0·006
3.SM	8	7·8	0·450
3.SR	8	9·4	0·311
Total	32	44·1	0·075
2.CT removed	24	22·6	0·542

Table 3. Comparison of models for the Hector's dolphin study

Model	ΔAICc	AICc Weight	np	Deviance
(ϕ, p_{t+n})	0·0	0·991	12	655·7
(ϕ_t, p_{t+n})	9·4	0·009	19	649·3
(ϕ, p_{t+m})	34·0	0	12	689·7
(ϕ_t, p_{t+m})	43·7	0	19	683·7
(ϕ, p_t)	52·6	0	11	710·5
(ϕ_t, p_t)	63·0	0	19	703·0

Following Pradel (1993), we consider a trap-response model in which the probability of capture in year *i* for individual *j* is modelled using

$$\text{logit}(p_{ij}) = \alpha_i + \beta x_{j,i-1}$$

where α_i is a year-effect and $x_{j,i-1}$ is set to 1 or 0 according to whether individual *j* was or was not captured in year $i-1$ ($i = 2, ..., 11$). We fitted the model in Program Mark (White & Burnham 1999) by regarding the capture histories for the first 10 years as individual covariates. We denote this model (ϕ_t, p_{t+m}) or (ϕ, p_{t+m}), according to whether ϕ is time dependent or not.

The photo-ID survey is such that an individual can be captured more than once in a given year. It might be expected that an individual sighted many times within a year (when sighted at least once) has a higher capture probability than one sighted less often. This motivated Fletcher (1994) to suggest a modification of the above model, in which $x_{j,i-1}$ is the number of times individual *j* was captured in year $i-1$, which we denote as (ϕ_t, p_{t+n}) or (ϕ, p_{t+n}), according to whether ϕ is time dependent or not.

We consider a total of six models, as shown in Table 3. It is clear that allowing for trap-response improves the fit to the data and that the model (ϕ, p_{t+n}) provides the best fit. This gives $\hat{\phi} = 0.907$ (95% CI: 0·847–0·945), compared with $\hat{\phi} = 0.863$ (95% CI: 0·818–0·898) from the model (ϕ, p_t), a difference of 0·044. Based on the overall LOF test statistic, we might have concluded there was no LOF and set $\hat{\phi} = 0.863$.

Using eqn 3 with $\alpha = 8$ (Slooten, Fletcher & Taylor 2000), $\phi = 0.907$ and assuming that the true value of λ lies between 0·95 and 1·05, we find that s_ϕ lies between 0·512 and 0·759. Use of eqn 4 then suggests that allowing for CH in the mark–recapture model will have increased $\hat{\lambda}$ by between 0·023 and 0·033. If $\hat{\lambda}$ is close to 1, this could mean estimating that the population is stable or increasing ($\hat{\lambda} \geq 1$) rather than in decline ($\hat{\lambda} < 1$).

WOLVES IN FRANCE

We provide a further illustration of the potential for bias in the presence of CH with a study of wolves (*Canis lupus*). After near-extinction in western Europe at the end of the 19th century (Breitenmoser, 1998), wolves are recovering in areas with high human density, causing interactions with agricultural activities (Linnell 2001) and therefore raising management issues. We focus on data from a noninvasive study carried out in the French Alps by the ONCFS, the French National Game and Wildlife Agency. Identification of individuals was based on the genotyping of faecal samples (Taberlet & Luikart 1999). The data we consider consist of capture histories for 160 individuals, partitioned into 35 3-month intervals (from spring 1995 to autumn 2003), as summarized in the m-array in Table 4.

Table 5 summarizes the results of the LOF test results from U-CARE. Both Test 2.CT and 3.SR are clearly significant. There appears to be no need to adjust for overdispersion: after removal of 2.CT and 3.SR, we have $\hat{c} = 50.9/55 = 0.93$. As for the dolphin study, use of a trap-response model does not imply a physical trap-response mechanism. When dispersal of individuals is not random, CH can be induced by transients (Pradel, Cooch & Cooke 1995). Dispersing wolves travel long distances to find a mate (Mech & Boitani 2003), which could explain why test 3SR is significant. When Tests 2.CT and 3.SR are both significant, individual CH is suspected (Péron *et al.* 2010).

To allow for group-effect CH, we fitted a mixture model (Pledger 2000; Pledger, Pollock & Norris 2003) using the multi-event modelling framework developed by Pradel (2005; see also Pradel 2009) and implemented in E-SURGE (Choquet *et al.* 2009). We assume that there are two groups because dominant individuals are more likely to use faeces to mark their territories and are therefore more detectable than subordinates or dispersers. We also consider the possibility of an additive effect of season on capture probability (Cubaynes *et al.* 2010).

For simplicity of presentation, we consider the four time-independent models in Table 6. There is clearly a need to allow for an effect of both two-group CH and season on capture probability. The best model, (ϕ, p_{h+s}), gives $\hat{\phi} = 0.839$ (95% CI: 0·717–0·987), compared with $\hat{\phi} = 0.680$ (95% CI: 0·615–0·880) from the model that considers only seasonal effects on capture probability, (ϕ, p_s), a difference of 0·159. Based on the overall LOF test statistic, we might have concluded that there was no LOF and set $\hat{\phi} = 0.680$.

Studies of North American populations suggest a mean age at first reproduction of 2 years and population growth rates between 0·90 and 1·58 (Fuller, Mech & Cochrane 2003). Using eqn 3 with $\alpha = 2$, $\phi = 0.839$ and λ between 0·90 and 1·58, we find that s_ϕ lies between 0·936 and 1·644. Use of eqn 4 then suggests that allowing for CH in the mark–recapture model will have increased $\hat{\lambda}$ by between 0·149 and 0·261. This is an even more striking example than the dolphin study of a situation in which we might estimate that the population is stable/increasing ($\hat{\lambda} \geq 1$) rather than in decline ($\hat{\lambda} < 1$).

Table 4. M-array for the wolves study, where zeros have been removed for ease of presentation

Year	Season	Occasion	Number captured	Occasion of first recapture																																	
				2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
1995	Spring	1	1	1																																	
	Summer	2	3		1	1																															
	Autumn	3	1			1																															
	Winter	4	5				3																														
1996	Spring	5	6					2																													
	Summer	6	5						1	1	1																										
	Autumn	7	1								1																										
	Winter	8	8								6				1																							
1997	Spring	9	15									6	1	2	3																						
	Summer	10	8											4						1																		
	Autumn	11	2																					1														
	Winter	12	16												10	2	1																1					
1998	Spring	13	18													6	2			1							2											
	Summer	14	9														3	1	3																			
	Autumn	15	10															5	1								1			1								
	Winter	16	12																6		1	1					1											
1999	Spring	17	15																	4	2	3		1	2		1						1			1	
	Summer	18	7																		3	1								1							
	Autumn	19	12																			4	1	1	3								1			1	
	Winter	20	15																				3	3	3	1											
2000	Spring	21	6																					4		1	1										
	Summer	22	10																						2	1	2		1	2							
	Autumn	23	8																							5			2		1						
	Winter	24	15																								8				1		1				
2001	Spring	25	16																									8		1	1		1		1		
	Summer	26	13																										4	4		2					
	Autumn	27	14																											10				3			2
	Winter	28	33																												12	2	2	2	1		
2002	Spring	29	22																													2	8	5	1	1	
	Summer	30	10																														3	2	2	1	
	Autumn	31	31																															12	3	1	
	Winter	32	37																																13	1	
2003	Spring	33	24																																	9	1
	Summer	34	19																																		4

Table 5. Summary of U-CARE LOF tests for the wolf study

	df	X^2	P
2.CL	30	27·9	0·576
2.CT	31	64·4	0·0004
3.SM	25	23·0	0·579
3.SR	29	65·4	0·0001
Total	115	180·7	0·0001
3.SR removed	55	50·9	0·631

Table 6. Comparison of models for the wolves study. For all models, the survival probability is constant. For the capture probability, p indicates use of a single capture probability, p_h indicates a two-group capture heterogeneity (CH) effect, p_s indicates a season effect and p_{h+s} indicates an additive effect of two-group CH and season

Model	ΔAICc	AICc Weight	np	Deviance
(ϕ, p_{h+s})	0·0	0·999	7	1228·9
(ϕ, p_h)	25·4	0	4	1260·3
(ϕ, p_s)	127·9	0	5	1360·8
(ϕ, p)	147·1	0	2	1386·1

Discussion

Previous work on the potential for CH to cause bias in the estimate of survival probability obtained from a CJS model (Carothers 1979) has suggested that this bias will be small, especially compared with that which occurs when estimating abundance. One of the aims of this paper was to reconsider this issue in the context of management decisions for long-lived species, for which estimation of the asymptotic population growth rate is highly sensitive to the estimate of adult survival probability.

For both of our examples, ignoring CH leads to the underestimation of adult survival probability and therefore of the asymptotic population growth rate. In both cases, the amount of underestimation was such that we could have reached different conclusions about the status of the population, compared with an analysis that allowed for CH. Our results suggest that undetected CH will generally lead to the asymptotic growth rate being underestimated, except when there is a high level of temporary emigration and the capture probability is low (Fig. 4), when it will tend to be overestimated.

We have focussed on the case where the estimate of adult survival is used in a deterministic, density-independent population model to estimate the asymptotic population growth rate. We would expect our results to still apply if we were to fit a population model, similar to the one used here, directly to all the available data.

Both of our examples involve noninvasive sampling (photography and DNA samples), a situation that can lead to misidentification of individuals. For the dolphin example, this issue was addressed by considering only the data for individuals that satisfied criteria concerning their identifiability (Slooten, Dawson & Lad 1992). A similar approach was used in the wolf example; genotypes with a mean 'quality index' less than a specified threshold were not used in the analysis. In addition, when using DNA samples, it is possible to allow for misidentification using modelling techniques (Lukacs & Burnham 2005; Wright *et al.* 2009). A related issue is that CH can sometimes be avoided by the use of an appropriate sampling design; Ebert *et al.* (2010) provided an example in which intense sampling effort on a small study area with few sampling occasions led to CH being less likely to be detected. It is worth noting that in the context of abundance estimation, use of a mixture model to allow for individual CH (Pledger, Pollock & Norris 2003) can sometimes lead to identifiability issues (Holzmann, Munk & Zucchini 2006).

For three common types of CH, our results have highlighted those situations where undetected bias might be high enough to cause a substantial bias in the estimate of asymptotic population growth rate. We could have used a higher threshold value for the power criterion, such as 0·8, which would have led to even more cases and more extreme levels of bias. Likewise, we chose to consider only those cases where the bias in the estimate was at least half its standard error. If we had chosen a higher threshold, the results would be similar in terms of the extreme levels of bias. With the chosen criteria, our results suggest that trap-response will cause little problem, as the power of Test 2.CT to detect it is generally high, even for smaller studies, and fitting a trap-response model should lead to an estimate that is almost unbiased. Two-group CH can lead to substantial undetected bias when the capture probabilities are low. Likewise, temporary emigration can do so when the rate of emigration is low, and for smaller studies this can occur even when the capture probability is reasonably high.

Throughout the paper, we have focussed on well-established LOF tests for detecting CH. More recently, directional components of Tests 2.CT and 3.SR have been developed (Pradel, Gimenez & Lebreton 2005). These show promise and are currently being investigated to derive new procedures for detecting heterogeneity. One could also consider detecting group-effect CH directly by comparing the fit of a mixture model (Pledger, Pollock & Norris 2003; Cubaynes *et al.* 2010; Marescot *et al.* in press) with several groups to one with a single group.

Acknowledgements

References

Anderson, D.R., Burnham, K.P. & White, G.C. (1994) AIC model selection in overdispersed capture-recapture data. *Ecology*, **75**, 1780–1793.

Boyce, M.S. (1977) Population growth with stochastic fluctuations in the life table. *Theoretical Population Biology*, **12**, 366–373.

Breitenmoser, U. (1998) Large predators in the Alps: the fall and rise of man's competitors. *Biological Conservation*, **83**, 279–289.

Buckland, S.T., Newman, K.B., Fernández, C., Thomas, L. & Harwood, J. (2007) Embedding population dynamics models in inference. *Statistical Science*, **22**, 44–58.

Burnham, K.P., Anderson, D.R., White, G.C., Brownie, C. & Pollock, K.H. (1987) Design and analysis methods for fish survival experiments based on release-recapture. *American Fisheries Society Monograph* Number 5 Bethesda, Maryland, USA, p. 215.

Cameron, C., Barker, R., Fletcher, D., Slooten, E. & Dawson, S. (1999) Modelling Survival of Hector's Dolphins around Banks Peninsula, New Zealand. *Journal of Agricultural, Biological and Environmental Statistics*, **4**, 1–10.

Carothers, A.D. (1973) The effects of unequal catchability on Jolly-Seber estimates. *Biometrics*, **29**, 79–100.

Carothers, A.D. (1979) Quantifying unequal catchability and its effect on survival estimates in an actual population. *Journal of Animal Ecology*, **48**, 863–869.

Caswell, H. (2001) *Matrix Population Models*, 2nd edn. Sinauer Associates, Massachusetts.

Choquet, R., Lebreton, J.-D., Gimenez, O., Reboulet, A.-M. & Pradel, R. (2009) U-CARE: Utilities for performing goodness of fit tests and manipulating CApture-REcapture data. *Ecography*, **32**, 1071–1074 (Version 2.2).

Cormack, R.M. (1972) The logic of capture-recapture estimates. *Biometrics*, **28**, 337–343.

Cubaynes, S., Pradel, R., Choquet, R., Duchamp, C., Gaillard, J.-M., Lebreton, J.-D., Marboutin, E., Miquel, C., Reboulet, A.-M., Poillot, C., Taberlet, P. & Gimenez, O. (2010) Importance of accounting for detection heterogeneity when estimating abundance: the case of French wolves. *Conservation Biology*, **24**, 621–626.

Daley, M.S. (1979) Bias in estimating the Malthusian parameter for Leslie matrices. *Theoretical Population Biology*, **15**, 257–263.

Dawson, S.M. & Slooten, E. (1993) Conservation of Hector's dolphins, The case and process which led to the establishment of the Banks Peninsula Marine Mammal Sanctuary. *Aquatic Conservation*, **3**, 207–221.

Devineau, O., Choquet, R. & Lebreton, J.-D. (2006) Planning capture–recapture studies: straightforward precision, bias and power calculations. *Wildlife Society Bulletin*, **34**, 1028–1035.

Ebert, C., Knauer, F., Storch, I. & Hohmann, U. (2010) Individual heterogeneity as a pitfall in population estimates based on non-invasive genetic sampling: a review and recommendations. *Wildlife Biology*, **16**, 225.

Fletcher, D. (1994) A mark-recapture model in which sighting probability depends on the number of sightings on the previous occasion. *Statistics in Ecology and Environmental Monitoring* (eds D. Fletcher & B.F.J. Manly), pp. 105–110. Otago University Press, Dunedin.

Fuller, T.K., Mech, L.D. & Cochrane, J.F. (2003) Wolf population dynamics. *Wolves: Behavior, Ecology and Conservation* (eds L.D. Mech & L. Boitani), pp. 161–191. University of Chicago Press, Chicago.

Gaillard, J.M., Yoccoz, N.G., Lebreton, J.-D., Bonenfant, C., Devillard, S., Loison, A., Pontier, D. & Allaine, D. (2005) Generation time: a reliable metric to measure life-history variation among mammalian populations. *American Naturalist*, **166**, 119–123.

Gilbert, R.O. (1973) Approximations of the bias in the Jolly-Seber capture-recapture model. *Biometrics*, **29**, 501–526.

Heppell, S., Caswell, H. & Crowder, L.R. (2000) Life histories and elasticity patterns: perturbation analysis for species with minimal demographic data. *Ecology*, **81**, 654–665.

Hines, J.E. & Nichols, J.D. (2002) Investigations of potential bias in the estimation of λ using Pradel's (1996) model for capture- recapture data. *Journal of Applied Statistics*, **29**, 573–587.

Holzmann, H., Munk, A. & Zucchini, Z. (2006) On identifiability in capture-recapture models. *Biometrics*, **62**, 934–935.

Houllier, F., Lebreton, J.-D. & Pontier, D. (1989) Sampling properties of the asymptotic behaviour of age- or stage-grouped population models. *Mathematical Biosciences*, **95**, 161–177.

Hwang, W.-H. & Chao, A. (1995) Quantifying the effects of unequal catchabilities on Jolly-Seber estimators via sampling coverage. *Biometrics*, **51**, 128–141.

Lebreton, J.-D. (1995) The future of population dynamic studies using marked individuals: a statistician's perspective. *Journal of Applied Statistics*, **22**, 1009–1030.

Lebreton, J.-D. & Clobert, J. (1991) Bird population dynamics, management and conservation: the role of mathematical modelling. *Bird population studies: their relevance to conservation and management* (eds C.M. Perrins, J.-D. Lebreton & G.J.M. Hirons), pp. 105–125. Oxford University Press, Oxford.

Lebreton, J.-D., Burnham, K.P., Clobert, J. & Anderson, D.R. (1992) Modeling survival and testing biological hypothesis using marked animals: a unified approach with case studies. *Ecological Monographs*, **62**, 67–118.

Linnell, J. (2001) Predators and people: conservation of large carnivores is possible at high human densities if management policy is favourable. *Animal Conservation*, **4**, 345–349.

Lukacs, P.M. & Burnham, K.P. (2005) Estimating population size from dna-based closed capture-recapture data incorporating genotyping error. *Journal of Wildlife Management*, **69**, 396.

Manly, B.F.J., McDonald, L.L. & McDonald, T.L. (1999) The robustness of mark-recapture methods: A case study for the Northern Spotted Owl. *Journal of Agricultural, Biological, and Environmental Statistics*, **4**, 78–101.

Marescot, L., Pradel, R., Duchamp, C., Cubaynes, S., Marboutin, E., Choquet, R., Miquel, C. & Gimenez, O. (2011) Capture-recapture population growth rate as a robust tool against detection heterogeneity for population management. *Ecological Applications* (doi:10.1890/10-2321.1).

Mech, L.D. & Boitani, L. (2003) Wolf social ecology. *Wolves: Behavior, Ecology and Conservation* (eds L.D. Mech & L. Boitani), pp. 1–34. University of Chicago Press, Chicago.

Nichols, J.D. & Pollock, K.H. (1983) Estimation methodology in contemporary small mammal capture-recapture studies. *Journal of Mammalogy*, **64**, 253–260.

Péron, G., Crochet, P.-A., Choquet, R., Pradel, R., Lebreton, J.-D. & Gimenez, O. (2010) Capture–recapture models with heterogeneity to study survival senescence in the wild. *Oikos*, **119**, 524–532.

Pledger, S. (2000) Unified maximum likelihood estimates for closed capture-recapture models using mixtures. *Biometrics*, **56**, 434–442.

Pledger, S., Pollock, K.H. & Norris, J.L. (2003) Open capture-recapture models with heterogeneity: I. Cormack-Jolly-Seber model. *Biometrics*, **59**, 786–794.

Pledger, S., Pollock, K.H. & Norris, J.L. (2010) Open capture-recapture models with heterogeneity: II. Jolly-Seber model. *Biometrics*, **66**, 883–890.

Pollock, K.H., Nichols, J.D. & Hines, J.E. (1985) Goodness of fit tests for open capture-recapture models. *Biometrics*, **41**, 399.

Pollock, K.H., Nichols, J.D., Brownie, C. & Hines, J.E. (1990) Statistical inference for capture-recapture experiments. *Wildlife Monographs*, **107**, 1–97.

Pradel, R. (1993) Flexibility in Survival Analysis from Recapture Data: Handling Trap-Dependence. *Marked Individuals in the Study of Bird Population* (eds J.-D. Lebreton & P. North), pp. 29–37. Birkhäuser, Basel.

Pradel, R. (1996) Utilization of capture-mark-recapture for the study of recruitment and population growth rate. *Biometrics*, **52**, 703–709.

Pradel, R. (2005) Multievent: an extension of multistate capture-recapture models to uncertain states. *Biometrics*, **61**, 442–447.

Pradel, R. (2009) The stakes of capture–recapture models with state uncertainty. *Modeling Demographic Processes in Marked Populations* (eds D.L. Thomson, E.G. Cooch & M.J. Conroy), pp. 781–795. Springer, New York.

Pradel, R., Cooch, E.G. & Cooke, F. (1995) Transient animals in a resident population of snow geese: local emigration or heterogeneity? *Journal of Applied Statistics*, **22**, 695–710.

Pradel, R., Gimenez, O. & Lebreton, J.-D. (2005) Principles and interest of GOF tests for multistate capture-recapture models. *Animal Biodiversity and Conservation*, **28**, 189–204.

Pradel, R., Hines, J.E., Lebreton, J.-D. & Nichols, J.D. (1997) Capture-Recapture survival models taking account of transients. *Biometrics*, **53**, 60–72.

Prévot-Julliard, A.C., Lebreton, J.-D. & Pradel, R. (1998) Re-evaluation of adult survival of Black-headed Gulls (*Larus ridibundus*) in presence of recapture heterogeneity. *The Auk*, **115**, 85–95.

Russell, R.W. (1999) Comparative demography and life history tactics of seabirds: implications for conservation and marine monitoring. *Life in the slow lane: ecology and conservation of long-lived marine animals*. (ed. J.A. Musick). *American Fisheries Society*. Bethesda, Maryland, USA, 51–69.

Schaub, M., Gimenez, O., Sierro, A. & Arlettaz, R. (2007) Use of integrated modeling to enhance estimates of population dynamics obtained from limited data. *Conservation Biology*, **21**, 945–955.

Slooten, E., Dawson, S.M. & Lad, F. (1992) Survival rates of photographically identified Hector's dolphins from 1984 to 1988. *Marine Mammal Science*, **8**, 327–343.

Slooten, E., Fletcher, D. & Taylor, B. (2000) Taking Account of Uncertainty in Management: Is Hector's Dolphin at Risk due to Mortality in Gillnet Fisheries? *Conservation Biology*, **14**, 1264–1270.

Stromborg, K.L., Grue, C.E., Nichols, J.D., Hepp, G.R., Hines, J.E. & Bourne, H.C. (1988) Postfledging survival of European starlings exposed as nestlings to an organophosphorus insecticide. *Ecology*, **69**, 590–601.

Taberlet, P. & Luikart, G. (1999) Non-invasive genetic sampling and individual identification. *Biological Journal of The Linnean Society*, **68**, 41–55.

White, G.C. & Burnham, K.P. (1999) Program MARK: survival estimation from populations of marked animals. *Bird Study*, **46**, 120–139.

Wright, J.A., Barker, R.J., Schofield, M.R., Frantz, A.C., Byrom, A.E. & Gleeson, D.M. (2009) Incorporating genotype uncertainty into mark-recapture-type models for estimating abundance using DNA samples. *Biometrics*, **65**, 833–833.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Mechanisms and models for capture heterogeneity.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.