Molecular Ecology (2005) **14**, 3909–3919 doi: 10.1111/j.1365-294X.2005.02717.x

INVITED REVIEW

Review of capture–recapture methods applicable to noninvasive genetic sampling

PAUL M. LUKACS* and KENNETH P. BURNHAM†

*Colorado Cooperative Fish and Wildlife Research Unit, Department of Fishery and Wildlife Biology, Colorado State University, Fort Collins, CO 80523, USA, †US Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit, Colorado State University, Fort Collins, CO 80523, USA

Abstract

The use of noninvasive genetic sampling to identify individual animals for capture–recapture studies has become widespread in the past decade. Strong emphasis has been placed on the field protocols and genetic analyses with fruitful results. Little attention has been paid to the capture–recapture application for this specific type of data beyond stating the effects of assumption violations. Here, we review the broad class of capture–recapture methods that are available for use with DNA-based capture–recapture data, noting the array of biologically interesting parameters such as survival, emigration rates, state transition rates and the finite rate of population change that can be estimated from such data. We highlight recent developments in capture–recapture theory specifically designed for noninvasive genetic sampling data.

Keywords: abundance, capture–recapture, individual identification, mark–recapture, microsatellites, sampling, survival

Received 14 March 2005; revision received 17 June 2005; accepted 25 July 2005

Introduction

In recent years, the use of noninvasive genetic sampling for capture–recapture studies has increased rapidly (Fig. 1). The methods have been applied to a diverse array of taxa such as bears (*Ursus* spp., Woods *et al*. 1999), African elephants (*Loxodonta cyclotis*, Eggert *et al*. 2003), coyotes (*Canis latrans*, Kohn *et al*. 1999), humpback whales (*Megaptera novaeangliae*, Palsbøll *et al*. 1997), and painted turtles (*Chrysemys picta*, Pearse *et al*. 2001) among others. Large advances have been made in field protocols (Woods *et al*. 1999) and genetic analyses (Paetkau 2003).

Most DNA-based capture–recapture studies follow the same basic principles. First, samples containing DNA are collected at several points in time. Often the samples are collected noninvasively. Hair, feathers, faeces, or other tissues are collected in a way that does not require physical contact of the animal. DNA is extracted from the samples and amplified at several microsatellite loci. Other molecular markers, such as single nucleotide polymorphisms (SNP), may be used as well (Budowle 2004). Matching genotypes are considered to arise from the same individual and classified as recaptures. The data are then analysed in a capture–recapture framework.

Little attention has been placed on the capture–recapture analysis of DNA-based data. A few studies have examined the effects of assumption violations on estimators of population abundance (Mills *et al*. 2000; Waits & Leberg 2000). Analyses focused on simple estimators of abundance. Recently, capture–recapture models been specifically developed for use with the special set of circumstances that arise with DNA-based sampling (Lukacs 2005; Lukacs & Burnham 2005).

Here we present a review of the literature on DNA-based capture–recapture studies. Thorough review papers dealing with genetics and field protocols already have been written (Taberlet & Luikart 1999; Taberlet *et al*. 1999; Mills *et al*. 2000; Waits 2004); therefore we choose to point out the advances that occurred since then and concentrate more heavily on data analysis issues. First, we provide an overview of the state of the genetics work in individual identification. Second, we focus on the unique features of DNA-based capture–recapture that separate it from standard tagging

Correspondence: Paul M. Lukacs, US Geological Survey, Patuxent Wildlife Research Center, 1201 Oakridge Dr., Suite 250, Fort Collins, CO 80525, USA. Fax: 970-225-3597; E-mail: paul_lukacs@partner.nps.gov

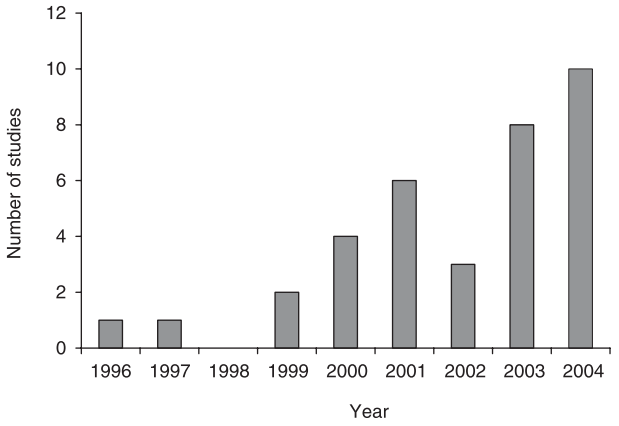

Fig. 1 The number of published DNA-based capture–recapture field studies by year from the first appearance in 1996 through 2004.

studies. Third, we present an overview of the broad array of analysis options available within capture–recapture theory that go far beyond simple estimates of abundance. Within this overview, we present recent developments in capture–recapture theory specifically designed for use with DNA-based data.

Genetics for individual identification

Much progress has been made in the genetic techniques for identifying individual animals. The work done in ecology builds off a strong foundation of research in human forensics. The two fields together offer a lot of knowledge to a researcher pursuing a DNA-based capture–recapture study.

In the ecological literature, the potential for problems to arise in genotyping microsatellite loci has been demonstrated repeatedly. Creel *et al*. (2003) show high error rates in genotyping wolf (*Canis lupus*) faecal DNA. Kohn *et al*. (1999) also demonstrated errors occurring in genotypes from field-collected faecal samples of coyotes. Similar results occur in laboratory-test situations of plucked hair samples (Goosens *et al*. 1998).

The problems faced in genetic analysis of noninvasive samples fall into three categories: (i) amplification failure, (ii) allelic dropout, and (iii) mutation during amplification. Amplification failure is a failure of the polymerase chain reaction (PCR) to produce copies of the intended piece of the genome. The failure may be caused by a lack of the target DNA or by PCR inhibitors present in the sample. From a capture–recapture standpoint, amplification failure is an easy problem to deal with because it simply results in no data from a sample analogous to never having obtained the sample. In capture–recapture studies, it is typically assumed that capture probability is less than one. Therefore, amplification failure can be absorbed into capture probability. Allelic dropout occurs when one allele at a locus fails to amplify or is not present in the pipetted DNA sample. The result causes the sample to appear to be homozygous at the locus when it may actually be heterozygous. This can be problematic for capture–recapture studies if it results in an incorrect genotype. The problematic effects of allelic dropout can be further confounded if samples contain a mixture of DNA from more than one individual (Roon *et al*. 2005). When samples are mixed and allelic dropout is occurring, it may not be possible to exclude a sample based on the appearance of more than two alleles at a locus. Lastly, it is possible, although rare, for an allele to mutate early in the PCR process and create a spurious allele. This could be problematic if an incorrect genotype results, but the rarity of the occurrence minimizes the risk.

Efforts have been made to limit genotyping error as much as possible. Paetkau (2003) shows that with careful laboratory protocol, including culling poor quality samples, a reliable set of microsatellite loci and experience, genotyping error rates can be greatly decreased and in some cases virtually eliminated. Extensive culling of lesser quality samples may be a double-edged sword for capture–recapture studies. On one hand, the final genotypes produced are more reliable. On the other hand, it has been demonstrated in human forensics that individual humans leave varying amounts of DNA. Some individuals, termed 'shedders', tend to leave much more DNA than nonshedders (Alessandrini *et al*. 2003; Lowe *et al*. 2002). Probably, animals also vary individually in their cell-shedding rate. Therefore, additional heterogeneity in capture probability is likely added as low quality samples are removed. Heterogeneity in capture probability is the most difficult problem to address in a capture–recapture analysis. Therefore, it may not always be the best option to use a laboratory protocol that completely eliminates genotyping errors. In some situations, a researcher may benefit from allowing a small amount of genotyping error in order to more evenly sample the entire population.

Explicit tests have been developed to determine whether error is present in a set of genotypes and the relative amount of error (McKelvey & Schwartz 2004a). First, the examining-bimodality (EB) test looks for an over-abundance of genotypes observed only once (McKelvey & Schwartz 2004a). When errors arise in the genotyping process, they typically lead to unique observed genotypes. Second, the difference-in-capture history (DCH) test examines the rate at which new individuals are recognized by looking at more loci (McKelvey & Schwartz 2004a). If the rate of adding more individuals exceeds that expected by increasing resolution among individuals alone, then genotyping error is likely. Combining these tests provides a researcher with useful information about the quality of the genotypes. The two tests assume equal capture probability among individuals. The effectiveness of the tests when capture probability

is heterogeneous has not been determined. In addition, the use of hypothesis testing to detect genotyping error can be problematic because the power of the test will decrease as the goal of reducing genotyping error is achieved.

An additional way to increase amplification rate and reduce error is multiplex pre-amplification (Piggot *et al*. 2004). For this approach, an initial large-volume PCR run is performed on the original sample with primers for the entire PCR panel. Then, a portion of the result of the first run is used in a second PCR with each primer individually. Piggot *et al*. (2004) demonstrate that the multiplex pre-amplification method greatly increases amplification rate for samples nearing the lower bound of template DNA available.

In human forensics, tetranucleotide microsatellites are used because they are less likely to produce errant genotypes (Eckert *et al*. 2002). Development of tetranucleotide markers for most wildlife species may be impractical because the increased cost and effort of locating polymorphic loci. Future developments in techniques for finding microsatellites and developing microsatellite primers will likely make using tetranucleotides more practical.

Issues facing DNA-based capture–recapture

The unique issues confronted in a DNA-based capture–recapture study can be split into two general categories. First, the list of marks in the population is unknown. Second, the concept of a sampling occasion is poorly defined. This sounds daunting as marks and sampling occasions are the major pillars of capture–recapture studies; but both of these issues can be overcome when properly addressed.

In a standard capture–recapture study, the identifying marks placed on animals captured are known to the investigator. In a DNA-based capture–recapture study, every animal in the population is self-marked prior to the study's beginning because every individual has a genotype. Unfortunately, the researcher does not know what genotypes (marks) are in the population. Not knowing the list of marks in the population leads to two problems: (i) it is difficult to differentiate between correctly and incorrectly read marks, and (ii marks may not be unique given only a subset of an animal's genome is used for identification.

When it is difficult to determine if a mark is read correctly, identification errors are more likely to be introduced into the data. In standard capture–recapture studies, when a mark is read that does not match a mark from the known list, the researcher rejects or corrects the observation. Thus few errors occur. In rare situations where one mark is misread as an already existing mark, there is little impact on the resulting parameter estimates because little change occurs in the sufficient statistics of the capture–recapture model. In DNA-based capture–recapture, no list of marks exists and therefore, incorrect genotypes are added to the sample as new individuals. Even small amounts of this error can have substantial impacts on parameter estimates (Wait & Leberg 2000; Creel *et al*. 2003). The basic form of a capture–recapture estimator of abundance is

$$\hat{N} = \frac{M}{\hat{p}^*}$$

where N is abundance, M is the total number of unique animals marked, and p^* is the probability that an animal is encountered at least once. When errors occur, M is larger than it should be and p^* is underestimated. Therefore, the ratio of the two quantities is greatly overestimated. In some cases, errors can largely be eliminated through strict protocols in the DNA analysis (Paetkau 2003), while in other cases errors persist (Creel *et al*. 2003). Methods have been developed to account for genotyping error in capture–recapture models and will be addressed in the next section of this review.

Marks may not be unique when the list of marks in the population is self-assigned as it is in DNA-based identification. When animals share a common mark, they appear to be only one animal. This has been dubbed the 'shadow effect' (Mills *et al*. 2000). The shadow effect results in underestimation of abundance and overestimation of survival. To a large extent, this problem can be overcome by using a set of microsatellite loci with high power to resolve individuals. Unfortunately, the use of a larger set of loci results in an increased probability of genotyping error. Nonetheless, lack of uniqueness of marks is a problem that should be dealt with by using molecular markers with more resolution rather than by modelling the problem in a capture–recapture framework.

In most DNA-based capture–recapture studies, animals are passively detected through observations of hair, scat or feathers. This introduces some uncertainty as to when the animal left the sample. It also allows multiple samples to be observed within a single sampling occasion. Both of these issues result in the concept of a sampling occasion to be blurred from the traditional definition of a sampling occasion as a short, discrete event.

When the time of deposition of a sample is unknown, the sampled population may be poorly defined. Typically, researchers set criteria for the age of a sample. In other cases, the design determines when a sample was deposited. For example, hair left on a snag must have been left after the snag was set. This helps better define the population. Even so, the estimated population remains as the number of animals that have used the sampled area within the time since the cut-off of sample decay. Therefore, the number of individuals that left samples may exceed the number of individuals currently in the population if animals are moving off the study area. This results in a type of geographical closure violation.

In addition to uncertain timing of when samples are left, multiple samples from the same individual may be

encountered within an occasion. This situation is largely avoided with active sampling because trapped animals generally cannot be trapped again until released by the researcher and visually detected animals are ignored if seen again before the next sampling occasion. Conversely, with DNA-based methods, multiple samples from the same individual are often collected and the identity is not known until a considerable investment of money and time is put into the sample. Through this large investment, additional data are available in these sampling schemes that have not been previously dealt with in capture–recapture analysis. Until recently, multiple detections of an individual within a sampling occasion were condensed down to a single detection for standard capture–recapture estimation. Now those additional detections can be used to help estimate abundance. The new method will be addressed more in the next section of this review.

The concept of a capture occasion can be unclear in studies that repeat sampling across space rather than time. For example, Eggert *et al*. (2003) sample African elephant dung in different sections of their study area for each occasion. Sampling across space rather than time requires the investigator to assume that all of the animals move throughout the entire study area. An advantage of sampling across space rather than time is that sampling can be done more efficiently. A disadvantage of this sampling scheme is it could add additional individual heterogeneity in capture probability. When samples are taken across space rather than time, some animals may be unavailable for sampling in some locations.

Some basic designs have been established that help to better define an occasion. Woods *et al*. (1999) describe a hair snag protocol for bears. They suggest placing the hair snags in the field and checking at regular intervals, much as would be done with a standard trap except the interval tends to be longer. Other options include sampling a set of transects or quadrats for dung. Transects themselves could be considered the occasions or could be sampled multiple times. Using transects as occasions leads to a lack of uniqueness in the 'time' ordering of the encounter history. Another design used is to sample continuously and set the number of occasions as the maximum number of times a single genotype is observed (Øystein *et al*. 2004). This method is not desirable because it greatly limits analysis options. Without a predefined number of sampling occasions, only the jackknife (Burnham & Overton 1979) and frequency of capture methods (Chao 1988; Chao *et al*. 1992) are appropriate analysis methods.

One problem where genetic identification has a clear advantage over traditional marking is that of tag loss (McDonald *et al*. 2003). Capture–recapture models assume that marks are permanent once they are placed on an animal. Sometimes tags fall off of animals or are no longer readable. Tag loss is problematic for capture–recapture analyses because the formal inference made is to the population of tags rather than the population of animals to which the tags are attached. Therefore, when tags are lost the direct link between tags and animals is lost also. Tag loss is not possible when a genotype is used as a mark, and thus the problem is avoided completely.

Options available in capture–recapture analysis

Capture–recapture theory has been a field of intense research for the past century. Advances and extensions of capture–recapture models allow a vast array of biologically interesting parameters to be estimated. Due to the vast array of possibilities available in capture–recapture analysis, we restrict the discussion to the types of questions that can be answered with DNA-based capture–recapture data. Currently, only a small piece of capture–recapture theory has been exploited in DNA-based studies (Table 1).

Animal abundance in a demographically and geographically closed population at one point in time is the target parameter of most DNA-based capture–recapture studies. A vast theory exists on how to estimate population size under different sets of assumptions (Table 2; Otis *et al*. 1978; Huggins 1989; Schwarz & Seber 1999; Pledger 2000; Borchers *et al*. 2002; Williams *et al*. 2002). Much of the current literature applying capture–recapture analysis to

Table 1 Five notable studies using DNA-based capture–recapture data collected in the field

Citation	Objective	Estimation method	Unique feature
Palsbøll *et al*. 1997	Estimate abundance of humpback whales	Not described	Application to a marine mammal
Boulanger *et al*. 2002	Estimate abundance of grizzly bears and explore assumption violations	Finite mixtures and Pradel models	Meta-analysis of seven studies
Creel *et al*. 2003	Estimate abundance of wolves when genotyping errors are present	Focused on estimating error rate	Proposed a method for handling errors
Eggert *et al*. 2003	Estimate abundance of forest elephants	Jackknife and accumulation curves	Uses space as the sampling occasion
Paetkau 2003	Determine the extent of genotyping error	Focused on estimating error rate	Meta-analysis of error rate

Table 2 Closed population capture–recapture models to estimate animal abundance from noninvasive genetic data. Along with each model type is if and how individual heterogeneity in capture probability is handled, what types of covariates can be used to model capture probability, whether the model has been extended to incorporate genotyping error, and whether it has been used in the robust design

Model type	Heterogeneity	Covariates	Genotyping error	Robust design
Full likelihood	None	Group	Yes	Yes
Finite mixtures	As a finite mixture	Group	Yes	Yes
Conditional likelihood	As an individual covariate	Group and individual	Yes	Yes
Continuous mixtures	As a continuous mixture	Group	Yes	No
Conditional likelihood with finite mixtures	As a covariate and finite mixture	Group and individual	Yes	Yes
Jackknife	By jackknifing	None	No	No
Frequency of capture	Bias adjustment	None	No	No

DNA-based data restricts analysis to simple cases of the likelihood-based models in Otis *et al*. (1978) and the jackknife estimator (Burnham & Overton 1979). Individual heterogeneity in capture probability, the variation among individuals in their probability of being detected, is the most difficult problem facing the estimation of animal abundance. Advances in closed population capture–recapture theory allow heterogeneity in capture probability to be modelled in more flexible frameworks. The conditional likelihood models of Huggins (1989, 1991) allow capture probability to be modelled as a function of individual covariates. For example, the distance from the point where an individual is detected to the nearest road may be a useful covariate as individuals that spend more time near roads are often more likely to be detected. Finite-mixture models approximate individual heterogeneity in capture probability with unobservable group differences in capture probability (Norris & Pollock 1996; Pledger 2000). Finite-mixture models may be particularly useful for DNA-based studies, because there are frequently differences in capture probability between unobservable groups, such as young and adult animals. Continuous-mixture models allow individual heterogeneity to be modelled on a continuous scale (Dorazio & Royle 2003). In general, model-based estimation of individual heterogeneity can be sensitive to the choice of model form (Link 2004). The most effective way to reduce bias caused by individual heterogeneity while adding the fewest assumptions is to sample a large portion of the population.

Two extensions to the closed population capture–recapture models have been developed to specifically address issues faced with DNA-based data. First, an extension to the models of Otis *et al*. (1978), Huggins (1989), and Pledger (2000) has been developed that includes a parameter to estimate genotyping error rate (Lukacs 2005; Lukacs & Burnham 2005). The method uses the disproportionate number of genotypes only observed once relative to genotypes seen more than once to estimate genotyping error. This model allows abundance to be properly estimated in the presence of genotyping error. The method provides a major advance because it eliminates the debate over whether genotyping error is problematic in DNA-based estimates of abundance. The debate has been raging in recent literature (McKelvey & Schwartz 2004a, b; Paetkau 2004) without clear resolution. The models from Lukacs & Burnham (2005) allow standard capture–recapture models to be compared directly to models expanded to estimate genotyping error in an information-theoretical framework. Therefore, the data guide the decision as to whether genotyping error is important or not. The results can then be model averaged for a further reduction in bias (Burnham & Anderson 2002).

A second method was developed to take advantage of the additional data available when multiple dung samples are collected from the same individual within an occasion. Lukacs (2005) and Lukacs *et al*. (in prep.) developed an estimator which uses these data to help account for the individual heterogeneity in capture probability caused by varying numbers of dung piles available. They assume numbers of dung piles deposited by animals follow either a Poisson or negative binomial distribution. Counts of dung piles per individual come from a combination of the numbers per individual and the probability of detecting a dung pile. Based on this, the authors fit mixture models to estimate the total unobserved abundance of dung, detection probability per dung pile, and abundance of animals. They demonstrate the effectiveness of the method through simulation and an example using African elephant dung data.

Miller *et al*. (2005) present an alternative method for using multiple encounters of individuals within a sampling occasion. The authors recast the problem into one of sampling with replacement; the statistical ball-in-urn type model. The method shows promising results in simulation and when applied to read data.

Accumulation or rarefaction curve analysis has been used to estimate abundance from noninvasive DNA data (Kohn *et al*. 1999; Eggert *et al*. 2003). Accumulation-curve

analysis fits a curve to the total number of unique genotypes seen after each sampling occasion. There should be fewer and fewer new genotypes observed after each sampling occasion. The curve eventually reaches an asymptote. The asymptote represents an estimate of population size. Capture–recapture analysis has several advantages over accumulation curves. First, accumulation curves do not account for the sampling design used to obtain the data (Cam *et al*. 2002). Accumulation curves are merely designed to approximate the appearance of the data not the process that generates the data. Capture–recapture models directly estimate detection probability. Second, accumulation curves do not efficiently use the data collected. Accumulation curves only use the first detection of an individual, whereas capture–recapture methods can use all detections. Finally, accumulation curves cannot account for variation in detection probability. Detection probability is known to vary widely in many situations across time, space and individuals. Estimation methods need to be able to account for these differences in order to appropriately estimate abundance. Therefore, accumulation curves are not recommended for estimating abundance.

Estimating abundance from DNA-based capture–recapture methods is likely to be most useful for relatively small populations, up to a few thousand individuals. Beyond this size, a very large number of samples would have to be collected and analysed making the study cost prohibitive. For large populations, other methods such as double sampling (Thompson 2002), mark–resight sampling (White 1996), or distance sampling (Buckland *et al*. 2001) are likely to be more efficient.

Abundance is just one parameter that can be estimated from capture–recapture data. Models have been developed to estimate survival, emigration rates, movement or transition rates, fecundity, and population growth (Nichols 1992). As more surveys continue across time, the desire to estimate more population parameters will increase. The Jolly–Seber (Jolly 1965; Seber 1965) model provides the basis from which much of open capture–recapture theory is built upon. While the Jolly–Seber model is rarely applied in its original form, it provides the underlying basis of many other models. Much of the work in this field is presented in the proceedings from the EURING technical conferences (North & Nichols 1995; Baillie & North 1999; Morgan & Thomson 2002; Senar *et al*. 2004). Table 3 provides an overview of open population models available for use with genetic data. The models are described in more detail below.

The Cormack–Jolly–Seber model (CJS, Cormack 1964; Jolly 1965; Seber 1965) allows estimates of apparent survival, the probability of surviving and remaining in the study area, of individuals conditioned on the individual being captured and marked at least once. The CJS model is useful for DNA-based capture–recapture when genotyping error is not present and the population is only sampled once per year or other relevant time interval. An advantage of the CJS model is that the estimates of apparent survival from this model are robust to individual heterogeneity in capture probability. A limitation is that apparent survival from a CJS model will underestimate true survival if animals permanently emigrate from the study area.

The robust design is a powerful way to bridge the gap from single-year abundance estimates to estimates of survival and emigration rates (Box 1, Pollock 1982; Kendall *et al*. 1997; Schwarz & Stobo 1997). The robust design uses two types of sampling periods. Primary sampling periods are separated by long intervals across which the population is assumed to be demographically and geographically open. Survival and temporary emigration rates are estimated across primary periods. Secondary periods occur at

Table 3 Open population capture–recapture models for the analysis of noninvasive genetic data. Along with each model type are the parameters that can be estimated and whether the model can be extended to estimate genotyping error

	Parameters estimated					
Model type	Survival	Movement rate	State transition rate	Abundance	Population growth rate	Genotyping error
Jolly–Seber	Yes	No	No	Yes, but not reliable	No	No
Cormack–Jolly–Seber	Yes	No	No	No	No	No
Robust design*	Yes	Temporary emigration	No	Yes	As a derived parameter	Yes
Multistate/-strata	Yes	Among observable strata	Yes	No	No	No
Robust design Multistate/-strata*	Yes	Among observable strata and to an unobservable stratum	Yes	Yes	As a derived parameter	Possible, but model not yet derived
Pradel	Yes	No	No	No	Yes	No

*See Table 2 for closed population submodels.

Box 1 The Robust Design

The robust design (Pollock 1982; Kendall *et al.* 1997) is one of the most useful sampling designs for capture–recapture studies. The robust design refers both to a sampling design and the class of models used for its analysis. The design consists of two levels of sampling periods. The primary sampling occasions are separated by a relatively long period of time over which the population is assumed to be demographically and geographically open, this means that birth, death, immigration and emigration all may occur. At each primary sampling occasion, there are several secondary sampling occasions. Secondary sampling occasions occur over a very short time interval. The population is assumed to be closed between secondary sampling occasions, although generalizations to this exist. The robust design allows abundance, temporary emigration/immigration rates, and survival to be estimated. Capture probability is estimated by secondary sampling occasion, therefore allowing a robust estimate of abundance for each primary occasion.

Across secondary sampling occasions within a primary occasion, it is best to actively try to give all individuals a chance of being detected. For example, it could be useful to change the type of scent lure used on different occasions to appeal to a broader range of the population. In addition, it is useful to have multiple sampling locations within an individuals home range to increase the probability of detecting an individual.

the transition between primary periods. The secondary periods are short and used for better estimating capture probability. The secondary periods may assume a geographically closed population (Kendall *et al*. 1997) or open population (Schwarz & Stobo 1997; Kendall & Bjorkland 2001). The robust design allows abundance to be estimated at each primary period. When the secondary periods are assumed to be geographically closed, the closed population abundance models described previously are used to estimate capture probability and abundance (Table 2). Survival and temporary emigration rates can be estimated between primary periods. The most important advantage of the robust design is that it allows animals to come into and leave the sampled population.

The robust design has been extended to handle situations where genotyping error is present. Lukacs (2005) developed a robust-design model that estimates genotyping error rate at each primary sampling occasion. By doing so, the model properly estimates abundance and survival. If genotyping error is present and not taken into account, abundance would be biased high and survival would be biased low. The robust design is required to extend the CJS model to estimate genotyping error because the information needed to estimate genotyping error is in the repeated detections of individuals within a primary occasion.

Another particularly useful advance in capture–recapture theory is that of multistate and multistrata models (Schwarz *et al*. 1993; Lebreton & Pradel 2002). The multistate or strata models allow one to estimate transition rates among conditions or locations. For example, one might wish to estimate the rate at which individual animals move among subpopulations. Movement among population can be particularly interesting when examined with molecular techniques because it is possible to determine both the rate of movement of animals, through capture–recapture, and the rate at which emigrating animals are contributing to the breeding population, through population genetics techniques. In addition, combining capture–recapture estimates of dispersal with molecular measures of dispersal can allow hypotheses about population structuring to be tested (Vandewoestijne & Baguette 2004). The multistrata models have been expanded to fit into a robust design sampling scheme as well (Kendall & Nichols 2002).

An application of capture–recapture theory of particular interest to ecologists is the direct estimation of population growth rate. Typically, this would be done by computing the dominant eigenvalue of a matrix projection model, such as a Lefkovitch or Leslie matrix (Caswell 2001). Such a method requires age and/or stated specific estimates of survival and fertility. With the method developed by Pradel (1996), population growth rate can be directly estimated from capture–recapture data. With this method, there is no need to separately estimate survival and fertility and assume a stable age distribution. One important sampling component to consider when using a Pradel model is that the area sampled must remain constant across time. If the area changes, the population growth rate loses interpretability. An exciting step beyond estimating population growth rate is to estimate an individual animal's contribution to population growth. This opens up capture–recapture theory to answer exciting evolutionary biology questions. Link *et al*. (2002) present a Bayesian approach combining capture–recapture theory and matrix modelling to estimate the latent individual fitness of an animal.

A recent extension of capture–recapture theory is the class of site occupancy models (MacKenzie *et al*. 2002). These models change the sampling unit from the individual animal to the plot. The data consist of presence or absence (more properly detection–nondetection) of an animal at each plot across sampling occasions. The models allow for imperfect detection of animals. The model estimates the proportion of plots occupied during the study.

These data are particularly useful for large-scale studies or surveys of rare species. The method is also useful for examining habitat usage and as a way to fit resource selection functions (Manly *et al*. 2002) while accounting for imperfect detection. MacKenzie *et al*. (2003) extended the method to a robust-design framework which allows colonization to be estimated in addition to occupancy rate. MacKenzie *et al*. (2004) further extended the model to handle more than one species and investigate patterns of co-occurrence. An advantage of site occupancy estimation is that it is typically less expensive to estimate occupancy rather than abundance. In most cases, only species must be determined from the molecular data, rather than individual. In addition, for species with very high variation in abundance across time and space, proportion of area occupied may be a more useful metric of the status of the species than abundance. A disadvantage of using site occupancy estimates for species with low variation in abundance is that population size may drop dramatically before any decline in the proportion of area occupied is detected.

A major advance in capture–recapture theory for application to answering ecological questions is placing the models in a general-linear-model (GLM) framework (Lebreton *et al*. 1992). The GLM framework allows parameters to be modelled as functions of covariates. The linear model is then tied to the model parameters with a link function. This allows biological hypotheses to be examined and complicated models to be built in parsimonious ways. For example, a 20-year time variation in survival could be modelled as a function of a climate variable with only 2 parameters in a GLM framework, rather than 20 parameters if each annual survival is estimated separately. Environmental, physiological or molecular covariates can be used at the individual or group level to address hypotheses in the GLM framework.

Multiple sources of data can be easily combined in capture–recapture analyses. For example, a segment of a population being sampled with DNA-based methods may also be radio-marked. Combining the radio telemetry data with a robust design DNA-based survey could provide more information about survival and abundance than either survey could provide alone. Multiple sources of data also arise from laboratory and field data in DNA-based studies. Laboratory estimates of genotyping error can help better estimate abundance in the wild (Lukacs 2005; Lukacs & Burnham in review). In addition, harvest data can be used as a source of dead recovery data and a known DNA source.

Laboratory estimates of genotyping error can be particularly useful in reducing sampling variance in DNA-based capture–recapture estimates. The use of a joint likelihood with a binomial estimator of error rate and the capture–recapture likelihood as is demonstrated in Lukacs & Burnham (in review) and Lukacs (2005). The joint likelihood produces standard errors on estimated abundance that are 20% smaller than without the additional data for a five-occasion case with a capture probability of 0.2 and 75 laboratory trials. Little additional reduction occurs from adding more laboratory samples because the variance from the detection process constitutes the vast majority of the remaining variation.

Software exists for a wide range of capture–recapture analyses (Box 2). Most capture–recapture theory builds off a reparameterization of a multinomial model (Burnham 1991); therefore, software can be designed to analyse a wide variety of capture–recapture data within a common framework. Nearly all model types included in this review are available in program MARK (White & Burnham 1999). Program M-SURGE is available for analysis of multistate problems (Choquet *et al*. 2003). Program PRESENCE performs site-occupancy estimation (MacKenzie *et al*. 2002). For unique problems, models can be coded in SAS (SAS Institute 2004), R (Ihaka & Gentleman 1996) or similar programmable statistics software for greater flexibility.

Capture–recapture analysis does not end at selecting the model type for your data and producing estimates. Once a

Box 2 Software available for capture–recapture analysis

MARK — Software for open- and closed-population capture–recapture and patch occupancy analysis. MARK has taken the place of many older programs such as CAPTURE, RELEASE, and BROWNIE. MARK includes models specifically designed to estimate genotyping error rate. It is freely available at www.cnr.colostate.edu/~gwhite/mark/mark.htm.

M-SURGE — Software for multistate or multistrata problems or other open-population capture–recapture for which the multistrata model is a generalization. M-SURGE is freely available at ftp://ftp.cefe.cnrs-mop.fr/biom/Soft-CR/.

POPAN — Software for the analysis of Jolly–Seber model problems. POPAN is freely available at www.cs.umanitoba.ca/~popan/ and also available within MARK.

PRESENCE — Software for the analysis of patch occupancy data. PRESENCE includes variations on patch occupancy models such as multispecies occupancy, robust design and latent abundance. It is freely available at www.mbr-pwrc.usgs.gov/software.html.

Analysis Software Forum — An online forum and other electronic resources for capture–recapture software and analysis are available at www.phidot.org/forum.

model type is selected, there is a process of determining which parameterizations of the model and which hypotheses are supported by the data. Thus capture–recapture analysis is a model-selection problem. The software presented above provides tools for model selection.

Sampling design is an important consideration in capture–recapture studies. Standard sampling theory only applies to capture–recapture at the broad study area scale. Sampling theory can help guide where to place effort when only a fraction of the study area can be sampled. At a more localized scale, sampling design is dictated by the biology of the study organism. It is important to provide the animals a reasonable chance of being detected. The 'big law' of capture–recapture design states that the best results occur when capture probability is high. Burnham *et al*. (1987; p. 315) show the trade-off between number of animals marked and capture probability to maintain a common level of precision. Far more individuals must be marked if capture probability is low. It is also important to sample in a way that provides the entire population of interest a chance of being detected, not just a subset. For example, if the males of a species are commonly found along streams and females with young are found away from streams, it is important to sample both areas if inference is to be made to the entire population of the species.

Sample size is also an important consideration for capture–recapture studies. Sample size is a combination of the number of detections and the number of sampling occasions. Traditional power analysis is typically not useful for capture–recapture studies because the investigator is not testing a null hypothesis. Determining the sample size needed to achieve a desired level of precision is more useful. This can be done through simulation. Program MARK provides the capability for simulating most analysis models.

Conclusion

The stage is set for a fruitful expansion of DNA-based capture–recapture studies to explore deeper ecological questions. Effective field protocols have been designed for a variety of organisms. Technology and techniques exist for reliable genotyping from samples that contain low numbers of copies of the organism's DNA. Capture–recapture theory exists to estimate an assortment of biologically important parameters and to cope with the situations faced with DNA-based sampling. The merging of the three areas is occurring rapidly.

While abundance is an important starting point for many investigations, especially for rare and endangered species, it should not be the end point of a study. Now it is time to move to broader ecological questions. Examining survival, population growth and state transition rates can provide insight into more details of the demographic process. When the molecular genetics and capture–recapture theory combine fully, many interesting questions will be answered.

Acknowledgements

The Colorado Division of Wildlife, Great Outdoors Colorado and Dr Tanya Shenk provided support for this review. M. Antolin, L. Bernatchez, and an anonymous referee provided comments on the manuscript.

References

Alessandrini F, Cecati M, Pesaresi M, Turchi C, Carle F, Tagliabracci A (2003) Fingerprints as evidence for a genetic profile: morphological study on fingerprints and analysis of exogenous and individual factors affecting DNA typing. *Journal of Forensic Science*, **48**, 1–7.

Baillie SR, North PM (1999) Large-scale studies of marked birds: proceedings of the EURING97 conference. *Bird Study*, **46** (supplement).

Borchers DL, Buckland ST, Zucchini W (2002) *Estimating Animal Abundance: Closed Populations*. Springer–Verlag, New York.

Boulanger J, White GC, McLellan BN, Woods J, Proctor M, Himmer S (2002) A meta-analysis of grizzly bear DNA mark–recapture projects in British Columbia, Canada. *Ursus*, **13**, 137–152.

Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L (2001) *Introduction to Distance Sampling*. Oxford University Press, Oxford.

Budowle B (2004) SNP typing strategies. *Forensic Science International*, **146S**, S139–S142.

Burnham KP (1991) On a unified theory for release-resampling of animal populations. In: *Proceedings 1990 Taipei Symposium in Statistics* (eds Chao MT, Cheng PE), pp. 11–35. Institute of Statistical Science, Academia Sinica: Taipei, Taiwan.

Burnham KP, Anderson DR (2002) *Model Selection and Multimodel Inference: An Information Theoretic Approach*. Springer–Verlag, New York.

Burnham KP, Overton WS (1979) Robust estimation of population size when capture probabilities vary among individuals. *Ecology*, **60**, 927–936.

Burnham KP, Anderson DR, White GC, Brownie C, Pollock KH (1987) Design and analysis methods for fish survival experiments based on release–recapture. *American Fisheries Society Monograph*, **5**.

Cam E, Nichols JD, Sauer JR, Hines JE (2002) On the estimation of species richness based on the accumulation of previously unrecorded species. *Ecography*, **25**, 102–108.

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

Chao A (1988) Estimating animal abundance with capture frequency data. *Journal of Wildlife Management*, **52**, 295–300.

Chao A, Lee SM, Jeng SL (1992) Estimation of population size for capture–recapture data when capture probabilities vary by time and individual animal. *Biometrics*, **48**, 201–216.

Choquet R, Reboulet A-M, Pradel R, Gimenez O, Lebreton J-D (2003) M-SURGE — *New software for multistate recapture models. Animal Biodiversity and Conservation*, **27**, 207–215.

Cormack RM (1964) Estimates of survival from the sighting of marked animals. *Biometrika*, **51**, 429–438.

Creel S, Spong G, Sands JL, Rotella J, Zeigle J, Joe L, Murphy KM, Smith D (2003) Population size estimation in Yellowstone

wolves with error-prone noninvasive microsatellite genotypes. *Molecular Ecology*, **12**, 2003–2009.

Dorazio RM, Royle JA (2003) Mixture models for estimating the size of a closed population when capture probabilities vary among individuals. *Biometrics*, **59**, 351–364.

Eckert KA, Mowery A, Hile SE (2002) Misalignment-mediated DNA polymerase β mutations: comparison of microsatellite and frame-shift error rates using a forward mutation assay. *Biochemistry*, **41**, 10490–10498.

Eggert LS, Eggert JA, Woodruff DS (2003) Estimating population sizes for elusive animals: the forest elephants of Kakum National Park, Ghana. *Molecular Ecology*, **12**, 1389–1402.

Goosens B, Waits LP, Taberlet P (1998) Plucked hair samples as a source of DNA: reliability of dinucleotide microsatellite genotyping. *Molecular Ecology*, **7**, 1237–1241.

Huggins RM (1989) On the statistical analysis of capture experiments. *Biometrika*, **76**, 133–140.

Huggins RM (1991) Some practical aspects of a conditional likelihood approach to capture experiments. *Biometrics*, **47**, 725–732.

Ihaka R, Gentleman R (1996) R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, **5**, 299–314.

Jolly GM (1965) Explicit estimates from capture–recapture data with both death and immigration-stochastic model. *Biometrika*, **52**, 225–247.

Kendall WL, Bjorkland R (2001) Using open robust design models to estimate temporary emigration from capture–recapture data. *Biometrics*, **57**, 1113–1122.

Kendall WL, Nichols JD (2002) Estimating state–transition probabilities for unobservable states using capture–recapture/resighting data. *Ecology*, **83**, 3276–3284.

Kendall WL, Nichols JD, Hines JE (1997) Estimating temporary emigration using capture–recapture data with Pollock's robust design. *Ecology*, **78**, 563–578.

Kohn MH, York EC, Kamradt DA, Haught G, Sauvajot RM & Wayne R (1999) Estimating population size by genotyping faeces. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **266**, 657–663.

Lebreton J-D, Burnham KP, Clobert J, Anderson DR (1992) Modelling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological Monographs*, **62**, 67–118.

Lebreton J-D, Pradel R (2002) Multistate recapture models: modelling incomplete individual histories. *Journal of Applied Statistics*, **29**, 353–369.

Link WA (2004) Nonidentifiability of population size from capture–recapture data with heterogeneous detection probabilities. *Biometrics*, **59**, 1123–1130.

Link WA, Cooch EG, Cam E (2002) Model-based estimation of individual fitness. *Journal of Applied Statistics*, **29**, 207–224.

Lowe A, Murray C, Whitaker J, Tully G, Gill P (2002) The propensity of individuals to deposit DNA and secondary transfer of low level DNA from individuals to inert surfaces. *Forensic Science International*, **129**, 25–34.

Lukacs PM (2005) *Statistical aspects of using genetic markers for individual identification in capture–recapture studies*. PhD Thesis, Colorado State University.

Lukacs PM, Burnham KP (in review) Extending the robust design for DNA-based capture–recapture data incorporating genotyping error and multiple data sources. *Ecological Applications*.

Lukacs PM, Burnham KP (2005) Estimating population size from DNA-based closed capture–recapture data incorporating genotyping error. *Journal of Wildlife Management*, **69**, 396–403.

MacKenzie DI, Bailey LL, Nichols JD (2004) Investigating species co-occurrence patterns when species are detected imperfectly. *Journal of Animal Ecology*, **73**, 546–555.

MacKenzie DI, Nichols JD, Hines JE, Knutson MG, Franklin AB (2003) Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology*, **84**, 2200–2207.

MacKenzie DI, Nichols JD, Lachman GB, Droege S, Royle JA, Lantimm CA (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, **83**, 2248–2255.

Manly BK, McDonald LL, Thomas DL, McDonald TL, Erikson WP (2002) *Resource Selection by Animals: Statistical Design and Analysis for Field Studies*. Kluwer Academic Publishers, Boston, Massachusetts.

McDonald TL, Amstrup SC, Manly BF (2003) Tag loss can bias Jolly–Seber capture–recapture estimates. *Wildlife Society Bulletin*, **31**, 814–822.

McKelvey KS, Schwartz MK (2004a) Genetic errors associated with population estimation using non-invasive molecular tagging: problems and new solutions. *Journal of Wildlife Management*, **68**, 439–448.

McKelvey KS, Schwartz MK (2004b) Providing reliable and accurate genetic capture–mark–recapture estimates in a cost-effective way. *Journal of Wildlife Management*, **68**, 453–456.

Miller CR, Joyce P, Waits LP (2005) A new method for estimating the size of small populations from genetic mark–recapture data. *Molecular Ecology*, **14**, 1991–2005.

Mills LS, Citta JJ, Lair KP, Schwartz MK, Tallmon DA (2000) Estimating animal abundance using noninvasive DNA sampling: promise and pitfalls. *Ecological Applications*, **10**, 238–294.

Morgan BJT, Thomson DL (2002) Statistical analysis of data from marked bird populations. *Journal of Applied Statistics*, **29** (1–4).

Nichols JD (1992) Capture–recapture models. *Bioscience*, **42**, 94–102.

Norris JL, Pollock KH (1996) Nonparametric MLE under two closed capture–recapture models with heterogeneity. *Biometrics*, **52**, 639–649.

North PM, Nichols JD (1995) Statistics and ornithology. *Journal of Applied Statistics*, **22** (5–6).

Otis DL, Burnham KP, White GC, Anderson DR (1978) Statistical inference from capture data on closed animal populations. *Wildlife Monographs*, **62**.

Øystein F, Hedmark E, Landa A *et al.* (2004) Colonization history and noninvasive monitoring of a reestablished wolverine population. *Conservation Biology*, **18**, 676–688.

Paetkau D (2003) An empirical exploration of data quality in DNA-based population inventories. *Molecular Ecology*, **12**, 1375–1387.

Paetkau D (2004) The optimal number of markers in genetic capture–mark–recapture studies. *Journal of Wildlife Management*, **68**, 449–452.

Palsbøll PJ, Allen J, Bérubé M *et al.* (1997) Genetic tagging of humpback whales. *Nature*, **388**, 767–769.

Pearse DE, Eckerman CM, Janzen J, Avise JC (2001) A genetic analogue of 'mark–recapture' methods for estimating population size: an approach based on molecular parentage assessments. *Molecular Ecology*, **10**, 2711–2718.

Piggot MP, Belleman E, Taberlet P, Taylor AC (2004) A multiplex pre-amplification method that significantly improves microsatellite amplification and error rates for faecal DNA in limiting conditions. *Conservation Genetics*, **5**, 417–420.

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

Pollock KH (1982) A capture–recapture design robust to unequal probability of capture. *Journal of Wildlife Management*, **46**, 752–757.

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

Roon DA, Thomas ME, Kendall KC, Waits LP (2005) Evaluating mixed samples as a source of error in non-invasive genetic studies using microsatellites. *Molecular Ecology*, **14**, 195–201.

SAS Institute Inc. (2004) *SAS version 9.1*. Cary, North Carolina.

Schwarz CJ, Seber GAF (1999) Estimating animal abundance: review III. *Statistical Science*, **14**, 427–456.

Schwarz CJ, Stobo WT (1997) Estimating temporary migration using the robust design. *Biometrics*, **53**, 178–194.

Schwarz CJ, Schweigert JF, Arnason AN (1993) Estimating migration rates using tag-recovery data. *Biometrics*, **49**, 177–193.

Seber GAF (1965) A note on the multiple-recapture census. *Biometrika*, **52**, 249–259.

Senar JC, Dhondt AA, Conroy MJ (2004) The quantitative study of marked individuals in ecology, evolution, and conservation biology. *Animal Biodiversity and Conservation*, **27**.

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

Taberlet P, Waits LP, Luikart G (1999) Noninvasive genetic sampling: look before you leap. *Trends in Ecology& Evolution*, **14**, 323–327.

Thompson SK (2002) *Sampling*, 2nd edn. John Wiley & Sons., New York.

Vandewoestijne S, Baguette M (2004) Demographic versus genetic dispersal measures. *Population Ecology*, **46**, 281–285.

Waits LP (2004) Using noninvasive genetic sampling to detect and estimate abundance of rare wildlife species. In: *Sampling Rare or Elusive Species: Concepts, Designs, and Techniques for Estimating Population Parameters* (ed. Thompson WL), pp. 211–228. Island Press, Washington, D.C.

Waits JL, Leberg PL (2000) Biases associated with population estimation using molecular tagging. *Animal Conservation*, **3**, 191–199.

White GC (1996) NOREMARK: population estimation from mark–resighting surveys. *Wildlife Society Bulletin*, **24**, 50–52.

White GC, Burnham KP (1999) Program MARK: survival estimation from populations of marked animals. *Bird Study*, **46**, S120–S139.

Williams BK, Nichols JD, Conroy MJ (2002) *Analysis and Management of Animal Populations*. Academic Press, New York.

Woods JG, Paetkau D, Lewis D, McLellan BN, Proctor M, Strobeck C (1999) Genetic tagging of free-ranging black and brown bears. *Wildlife Society Bulletin*, **27**, 616–627.

Paul M. Lukacs is Biometrician for the USGS Patuxent Wildlife Research Center. His current research includes capture-recapture theory and application and design of long-term monitoring programs. Kenneth P. Burnham is Senior Scientist and Assistant Unit Leader at the USGS Colorado Cooperative Fish and Wildlife Research Unit. His interests include study design, sampling and data analysis methodology for wildlife and ecology, in particular population estimation from capture-recapture, distance sampling, population monitoring, and inference about population dynamics. Ken's recent research emphasis focuses on theory and application of information-theoretic data-based model selection and multimodel inference.
