Ch. 7

Introduction to mark-recapture models

"Man is the only kind of varmint who sets his own trap, baits it, and then steps in it."
-- John Steinbeck

Questions to ponder:

- *What is the difference between closed and open populations?*
- *What is meant by 'encounter probability'?*
- *How do catch-per-unit-effort methods compare to mark-recapture methods?*

Mark-recapture data and analyses

The use of mark-recapture data has a long history in wildlife management and ecological studies. The placement of marks on animals will continue to be a staple method in the toolkit of a wildlife biologist.   Mark-recapture data allows us to answer questions such as: "how many animals are there in this population?" Or, "Are some animals at a higher risk of mortality than other animals?" Or, "what is the probability that animals in the population will remain in the population?"

As technology has changed, biologists have added more type of marks to their toolkit—for example, the use of Passive Integrated Transponders (PIT) and satellite telemetry. But, the basic approach to analysis of mark-recapture data has not changed in decades. Here, we will explore some newer methodologies and analysis models, but we begin with the standard: a "closed" mark-recapture model (Chapter 8) to consider animal abundance questions. Chapters 9-13 will explore "open" mark-recapture models to address questions of mortality and movement.

The common denominator of all mark-recapture methodology is that an animal is captured in some manner and marked. The animal is then released and captured at a subsequent time period. Some marking methods provide unique identifications (e.g. "pheasant #133"), but other marking methods only show than an animal is marked (e.g., a splotch of white paint on the scales of an iguana).

Samples and encounter probability

We know that when we capture animals in most populations, we are not able to capture the entire population at one time. Thus, by definition, we are sampling the population. And, we will try to make inferences about the population from our sample.

What if you were told that a biologist went to three study sites, placed some sort of net or trap, and recorded the capture of 100 unique individuals at each site? Could you make an inference about the size of the population without knowing any other information? The inference, certainly, is limited—the best inference is that there are at least 100 individuals in each population.

But, how many animals are really in the population? What additional information about our trapping effort do we need to know?

We must know something about the capture, or encounter, probability. That is, what is the probability that an individual living in the population is captured? In mathematical language, we have our sample, n, but we want to estimate the population size, N. We know that if capture probability is p:

$$\hat{N} = \frac{n}{\hat{p}}$$

So, if we told you that in the first population, $p = 0.30$, you could tell me that the population size is 333. We know that because we captured 30 percent of the population (another way of saying that an animal had a 30% chance of being captured in our trapping program). So, 100 animals represented 30 percent of the population. And, by math we can see that $\hat{N} = 100/0.30 = 333$.

> ## Types of encounters
>
> Specific types of research often have different terms to indicate the type of encounter that is involved:
>
> **"Capture probability"** typically refers to the probability that an individual is captured during a given time period in an actual trap or net.
>
> **"Recovery probability"** is defined as the probability that a hunter will, during a given time period, shoot a marked animal—shot and recovered animals cannot be released to the wild because they are dead. Thus, "recoveries" are a special type of mark/recapture data.
>
> **"Re-sighting probabilities"** are defined as the probability that an animal marked with colored or numbered tags will be seen by a person and recorded (i.e., not in a trap) during a given time period.

Can you estimate the population size if $\hat{p} = 0.50$? What if $\hat{p} = 0.70$?

You should find that $\hat{N} = 100/0.50 = 200$ and $\hat{N} = 100/0.70 = 142$.

So, capturing 100 animals in each population is only our sample, and we cannot infer much about the population size until we know the encounter probability, p.

NOTE: You may see, especially in older literature, the term 'encounter rate' or 'capture rate'. Most scientists (and editors of journals!) now prefer the term "encounter probability" as a true "rate" has a time-specific element attached, such as a flow rate of a stream (meters/second). Although it is true that 'survival rates' also have a time element that should be specified (annual, monthly, daily, etc.), the stochastic nature of binomial-type events is emphasized with the use of the term "probability".

Catch-per-unit-effort

There are implications to standard methodologies that assume equal catchabilities—fisheries biologists and small mammal biologists, for example, historically calculated "Catch-per-unit-effort" (CPUE) statistics. CPUE might be expressed as 350 fish/net or 1300 small mammals per trap night.

But, if a net captures 350 fish of size >=700mm, 350 fish of size between 200 and 500mm, and 350 fish of size <200mm, we are in the same situation as we were with our 100 animals captured at each of three study sites. The size of the holes in the net, perhaps, might favor the capture of fish in the 200-500 mm size—larger fish bounce off and small fish swim through it. So, if capture probability is highest for the 200-500mm fish and lowest for the other two size classes, our CPUE cannot be used to infer anything about population size. In fact, there would be many more of the smallest- and largest-sized fish, in our example, than the middle size class. Why? Because capture probability of the smallest and largest size classes was very low. For example, if $p=0.10$ for those size classes, $\hat{N} = 350/0.10 = 3500$ fish. And, if $p=0.9$ for the middle size class, $\hat{N} = 350/0.9=389$ fish.

Despite the original inference from CPUE of similar numbers of each size category, we now see that there are many more small and large fish than mid-sized fish. For this reason, most fisheries biologists use CPUE results with caution because they know that their nets' capture efficiencies are different for individuals of different sizes.

Some logic

Before we launch into the mathematics of the simple estimator for population size, let's take a look at the logic behind the estimator. In Figure 7.1, you see two populations (one at top, the other at the bottom). A sample of 100 animals is taken during the first capture occasion in both populations. Things are looking fairly even at the moment. But, we should be aware that we can't infer much about population size, yet. We only know that each population has at least 100 animals!

Next, a second capture occasion is held and another 100 animals are captured (perhaps our traps or nets only hold 100 animals?). But, some interesting differences arise, now. In the first population only 20 of the 100 animals captured in time 2 are marked. And, in the second population, 80 of the 100 animals captured in time 2 are marked (Figure 7.1).

Can you suggest which population is larger than the other?

You should have answered "Population #1" (Figure 7.1). Why? Because we marked 100 animals in both populations in time 1. And, in time 2, we appear to have marked almost all (80% in fact) of Population #2, while we appear to have only marked 20% of Population #1. If 100 animals are only 20% of the population in Population #1, it would seem to be very big— much bigger than Population #2. In a few pages, we will put some equations to this logic. For now, you should be able to look at this example and see we can make a much better inference about population size with a very simple mark-recapture experiment.
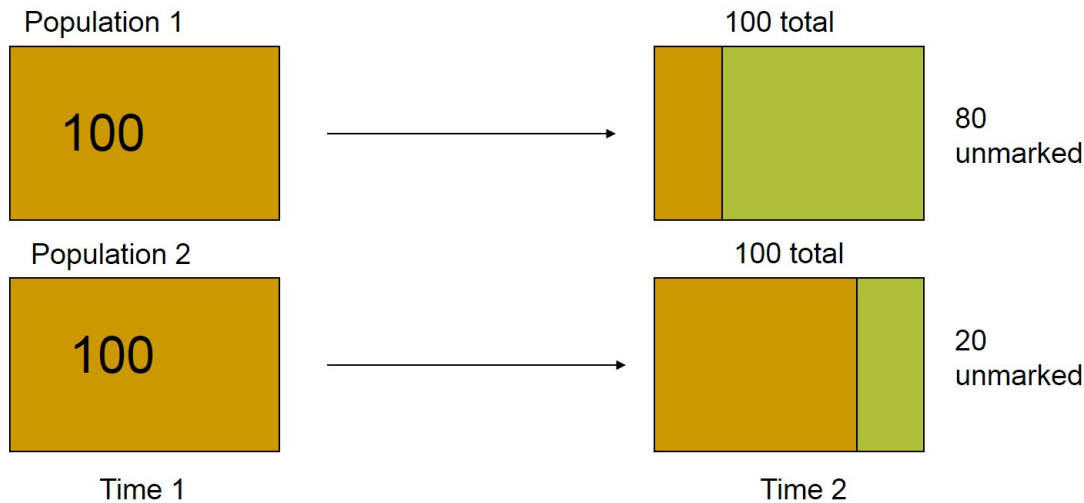

Figure 7.1: *Two populations with very different sizes (Population 1 and 2) are sampled with 100 animals marked in each population during Time 1. During Time 2 another 100 animals are sampled from each population. The number of marked animals captured varies between populations in Time 2. The presence of many unmarked individuals in Population 1, relative to the number of unmarked individuals in Population 2 suggests the Population 1 is bigger than Population 2.*

The cohort

An important concept in the use of mark-recapture analyses is the "cohort". A cohort is the number of marked animals released during time period *t*. We use cohorts to establish capture histories, which are the basis for all analyses of mark-recapture data. Capture histories are records that show when an individual has been captured during a study. For example, if a cohort of individuals is captured during time period 1 and then released immediately, those animals can have four different capture histories during a study with three capture occasions: "111", "110", "101", and "100". For simple capture histories, we use a "1" to indicate that an animal has been captured (a 'success'). And, we use a "0" to indicate that an animal was not captured (a 'failure').

This type of data is called **"binomial"**—two forms (1/0, or success/failure).

But, the released cohort is critical to our analyses—all subsequent work is conditioned (or statistically based) on that initial sample of animals that are captured and released. Mark-recapture analyses, at their core, are an attempt to estimate something you don't know (e.g., population size). To accomplish that goal, you must have something you do know—and when

we release a cohort of animals, we know how many we released. And, that will be very important, as you will see.

For example, if you were to decide to study ducks in the state of New York, USA, you could not start by collecting bands, or rings, or ducks recovered by hunters in New York. Those bands, or rings, would provide useless data unless you knew how many birds had been banded, or ringed, initially. That is, without knowledge of the cohort, all you have is a shiny pile of bands in your hand!

Closed or open?

Mark-recapture analyses can be divided into two main types of analyses: analyses of 'closed' populations and analyses of 'open' populations.

Remember the basic equation of population growth (Chapter 2) is a function of births (B), deaths (D), immigration (I), and emigration (E): $\Delta N = B + I - D - E$.

We refer to 'closed' populations as having $\Delta N = 0$. That is, no births, deaths, or movements of animals in or out. Although we know that no species exists in closed populations, populations may be considered 'closed' if we study them over very short time periods. This is necessary to estimate population size. And, it is important to note that 'very short time period' is relative to each species. The scale over which animals are born and die is short for small rodents, relative to elephants. So, a month-long study of small rodents might be considered to be too long for our closure assumption (many small rodents might die or be born during a month). But, a 4-month study of elephants might be considered very acceptable (very few deaths and births are likely during such a period).

We will come back to questions of closure of populations later, but this is a key consideration when you start to design a study for your species. When are the birth pulses? How often do individuals die? After a given time period (e.g., day, week, month), will animals have left your study site?

Alternatively, models of "open" populations are often more realistic. When we study 'open' populations, we assume that deaths are occurring. Mark-recapture methods treat movements in unique ways, but many types of analyses of open populations assume movements are occurring as well.

Therefore, studies of open populations are usually longer in duration than studies of closed populations (however, remember that a closed study of elephants might be longer than an open study of mice!). In fact, we want deaths to occur during studies of open populations. If animals do not die (or, for more complex models, leave or enter the study site), our study will not be able to estimate a useful survival (or movement) probability.

Models matter

As we start to envision our analyses as models of systems, it may be useful to compare simple open and closed model structures. As you begin working with mark-recapture data, you must start to think in terms of probability statements. So, let's start with a simple example:

Let's consider the situation where you captured a sample of individuals and released them as a cohort during time 1. Then, during time 2 you recapture some of that cohort. Given that you captured the animals during time 1, they will either have a "11" or "10" capture history (obviously, other animals not captured during time 1 could have a "01" or a "00" capture history as well, correct?).

So, what is the probability of having a capture history of "11"? It depends completely on the 'model' under which we are working. Is it a <u>closed system</u>? If so, the only thing that had to happen for our animal released in time 1 to have a "11" capture history is…to be captured again. So, we can represent that as a probability statement: p_2—the probability of capture during time 2.

What is the probability of having a 10 capture probability? It is $1\text{-}p_2$, or the probability of NOT being captured during time 2. It is as simple as that.

Now, let's change our model. Let's say, instead, that we are doing an analysis of animals in an <u>open population</u> (specifically, a "Cormack-Jolly-Seber," or CJS-type, open population—more on that in Chapter 8). We have the

Type of model	Capture history	Probability of animal having capture history
Closed population	11 10	p_2 $1\text{-}p_2$
Open population, CJS	11 10	$S_1 p_2$ $S_1(1\text{-}p_2)+(1\text{-}S_1)$

Note: probabilities for capture histories are conditioned on capture and release in time 1

same animals released in time period 1. Now…what has to happen for them to have a capture history of "11"? It's going to be more complicated than the probability in our closed model—because there is something else happening: *some animals are dying.*

For an animal to have a capture history of "11" in an open population, the animal must first survive the interval in which they are released, and then they must be captured in the next interval. So, the probability would be: $S_1 p_2$. We multiply the two probabilities, S_1 and p_2, together, as the animal must do both—survive and be captured. We will cover this again when we discuss open populations, but we assume in analyses of open populations that the cohort of animals is released simultaneously at the beginning of time 1. So, they must survive through time period 1, and hence we used the S_1 designation to refer to survival during time period 1. Capture happens at the beginning of the next time period, so we use the p_2 to designate capture probability in time 2.

What is the probability of having a capture history of "10"? You might simply look at the closed model and suggest: $S_1(1\text{-}p_2)$. And, you would be partially correct. That is, an animal *can* have a capture history of "10" by surviving time 1 and then avoiding capture in time 2.

But, there is *another alternative* that also results in a capture history of "10"—the animal might die during time 1. Yes, as the famous Alfred, Lord Tennyson wrote, Nature is *"red in tooth and claw."* And, that affects our statement of capture probability—we must account for both possibilities. The total probability for having capture history of "10" is the *sum* of the two alternatives: capture avoidance and death. So, we write it as: $S_1(1-p_2) + (1-S_1)$. For a further discussion of CJS-type capture histories with more than two capture periods see Chapter 10.

Conclusion

Biologists use mark-recapture models to estimate parameters for populations, and the models are based on probability statements that require certain assumptions about the population in question. To estimate population size, we use models that assume a 'closed population' (no deaths or movement in/out). To estimate survival, we use models that are based on 'open populations'. Capture histories (e.g., "1001100") are the typical type of input data for these analyses, and our analyses are conditioned on the released cohort: animals released at the same time.

For more information on topics in this chapter

Amstrup, S. C., T. L. McDonald, and B. F. J. Manly. 2005. Handbook of capture-recapture analysis. Princeton Univ. Press: Princeton, NJ.

Conroy, M. J., and J. P. Carroll. 2009. Quantitative Conservation of Vertebrates. Wiley-Blackwell: Sussex, UK.

Cooch, E., and G. White. 2014. Chapter 1: First Steps. *In* Program MARK: a gentle introduction, 12th edition, Cooch, E. and G. White, eds. Online: http://www.phidot.org/software/mark/docs/book/pdf/chap1.pdf

Nichols, J. D. 1992. Capture-recapture models. BioScience 42: 94-102.

White, G. C., D. R. Anderson, K. P. Burnham, and D. L. Otis. 1982. Capture-recapture and removal methods for sampling closed populations. Los Alamos National Laboratory. LA-8787-NERP. 235 pp.

Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. Analysis and management of animal populations. Academic Press, San Diego.

A female white-tailed deer (Odocoileus virginianus) is marked with a numbered ear tag and a radio-telemetry collar in western Iowa, USA. Photo by Greg Clements, University of Nebraska-Lincoln.