7.3

THE ROBUST DESIGN FOR CAPTURE–RECAPTURE STUDIES: ANALYSIS USING PROGRAM MARK

WILLIAM L. KENDALL

U.S. Geological Survey, Patuxent Wildlife Research Center, 11510 American Holly Drive, Laurel, MD 20708-4017, USA, William_Kendall@usgs.gov

Abstract: Collecting capture–recapture data under Pollock's robust design provides an additional source of information on capture probability that can be used to provide less biased and more efficient estimates of population dynamics parameters. In addition, it can be used to estimate the probability of being available for capture, which in some cases (e.g., breeding proportion) has ecological significance. This phenomenon can be modeled as a completely random process, Markovian, or with temporary trap dependence. Analysis of this type of data is an option in program MARK. By using MARK, the relationship between parameters and covariates can be modeled, and various approaches to goodness of fit, model selection, and model averaging can be implemented.

Key words: closed population, estimation, mark–recapture, open population, program MARK, program RDSURVIV, temporary emigration, trap response

---

Capture–recapture statistical methods have long been used by scientists for studying the abundance or dynamics of wildlife and fish populations. These methods are useful because they adjust estimates of demographic parameters for detection probabilities. Nevertheless, capture studies can be expensive and time-consuming, and therefore study design and statistical methods should result in efficient results.

CLASSIC DESIGN

Traditionally, long-term capture–recapture data from studies that include natality, immigration, mortality, and emigration have been analyzed using the Jolly-Seber (JS) method (see Pollock et al. 1990) or some variant of it. This method is based on 1 capture occasion per period of interest. Nevertheless, a single sample in many studies actually consists of several sampling sessions from which the data are pooled for the period into whether an individual animal was captured at least once in those sessions. In addition, an assumption of the JS method is that samples are collected instantaneously. Although this is rarely achieved in wildlife science, collecting a sample over a longer period of time does not bias parameter estimators if the population dynamics are static for the duration of that sampling effort period. Alternatively, a closed-population capture–recapture study—where multiple samples taken over a sufficiently short time period to assume closure—requires another set of methods for estimating abundance (or detection probability; see Otis et al. 1978, Rexstad and Burnham 1991).

POLLOCK'S ROBUST DESIGN

Pollock (1982) proposed that closed-population methods be used to analyze data from sessions within a period and JS methods be used for data from multiple periods. The JS method for estimating survival rate tends to be robust to heterogeneity in detection probability, but its abundance and recruitment estimators are not. Closed-population methods include abundance estimators that are robust to this heterogeneity. Therefore, by using closed-population estimators for abundance and JS estimators for survival rate (and a combination of these for recruitment), the overall analysis is more robust to heterogeneity.

Kendall et al. (1995) built models that exploit within- and between-period information simultaneously, instead of piecemeal as in the ad hoc approach above, for the cases where detection probability varies only by time or by whether the animal has been detected in the past (i.e., a trap-happy or trap-shy response). They showed that even survival rate estimators under this approach are more precise than under the JS method.

Kendall and Nichols (1995) and Kendall et al. (1997) modified these models to allow for an animal in the population to be unavailable for capture at a given time (i.e., a temporary emigrant). They allowed for 3 cases: where the probability of availability (1) is completely random, or depends on whether the animal was (2) available for capture or (3) actually captured in the previous period. This phenomenon can have several interpretations. (1) The study area could exclude part of the home range of each animal (Fig. 1a). (2) The activities of a subset of the population might preclude their detection in a given period; e.g., when mammals are in torpor (Fig. 1b; Kendall et al. 1997). (3) In the case of breeding populations, availability for detection may be equivalent to being a breeder (Kendall and Nichols 1995).

The models based on the robust design described above are multinomial statistical models that include the following parameters: (1) ϕ_t = probability that a member of the population in period t survives and is still a member of the population in period $t+1$; (2) γ''_t, γ'_t = probability that a member of the population in period t is unavailable for detection (e.g., outside the study area), given that it was available or unavailable, respectively, for detection in period $t-1$; (3) p_{ts}, c_{ts} = probability that an animal that is available for detection during period t and has not or has, respectively, previously been detected during period t is detected in sample s. These models condition on the total number of animals captured during a period (i.e., do not include population size N_t as a parameter). Instead, an estimate of this parameter is derived from n_t = total
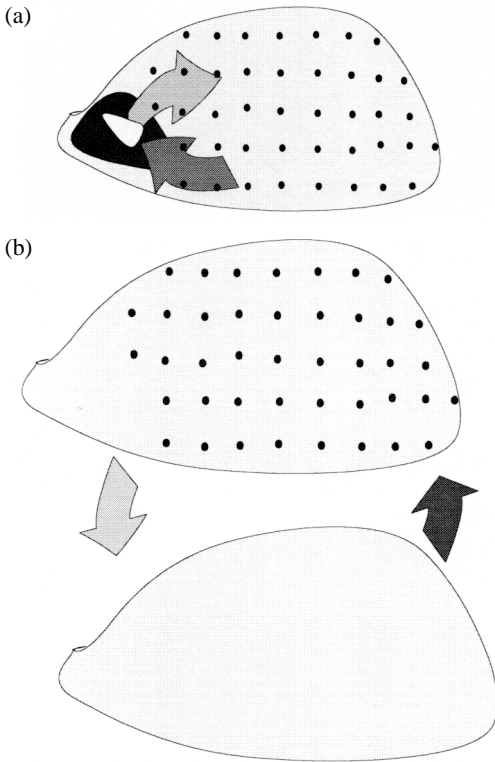
(a)

(b)

Fig. 1. (a) Population where some animals are absent from the study area because it does not encompass the entire range of the population. (b) Population where some animals are absent because they are in dens below the ground and therefore cannot be accessed.

number of individuals detected during period t, and the estimated pooled detection probability for period i, $\hat{p}^*_i = 1 - \prod_{s=1}^{S_t} (1 - p_{ts})$, where S_t is the number of sessions in period t. Abundance is then estimated as $\hat{N}_t = n_t/p^*_t$.

The parameterization above lends itself to computing estimates using program SURVIV (White 1983), or more specifically program RDSURVIV (Kendall and Hines 1999), a version of SURVIV customized for the robust design. Program MARK provides options to use a variation on this approach or a model that includes N_t.

Estimation of unavailability for capture (γ''_t, γ'_t) cannot be accomplished using the classic design. This is easy to illustrate if we assume that this phenomenon is completely random ($\gamma''_t = \gamma'_t = \gamma_t$). In this case, the JS estimator for capture probability is biased: $E(\hat{p}^{*js}_t) = (1 - \gamma_t)p^*_t$. This indicates that γ_t and p^*_{tb} are confounded. Using the unbiased estimator that comes from closed population methods, p^{*cl}_t, simple algebra yields an estimator for γ_t: $\hat{\gamma}_t = 1 - \hat{p}^{*js}_t / \hat{p}^{*cl}_t$ (Kendall and Nichols 1995). When availability for capture is Markovian, no ad hoc estimators of this kind are available.

ADVANTAGES OF THE ROBUST DESIGN

Advantages of the robust design alluded to above include (1) estimates of p^*_t, and thus N_t and recruitment are less biased by heterogeneity in capture probability; and (2) temporary emigration can be estimated assuming completely random, Markovian, or temporarily trap dependent availability for capture (Kendall and Nichols 1995, Kendall et al. 1997). (3) If temporary emigration does not occur, abundance, survival, and recruitment can be estimated for all time periods (e.g., in a 4-period study, half the parameters are unestimable using the JS method; Kendall and Pollock 1992). (4) Precision tends to be better using the formal robust design models of Kendall et al. (1995). (5) Because there is information on capture for the youngest catchable age class, estimation of recruitment into the second age class can be separated into in situ recruitment and immigration when there are only 2 identifiable age classes. Using the classic design, 3 identifiable age classes are required (Nichols and Pollock 1990). (6) The robust design's 2 levels of sampling allow for finer control over the relative precision of each parameter (Kendall and Pollock 1992).

ASSUMPTIONS OF ANALYSIS UNDER THE ROBUST DESIGN

For the most part, the assumptions under the robust design are a combination of the assumptions for closed-population methods and the JS method.

(1) The population is assumed closed to additions and deletions across all sampling sessions within a period. Kendall (1999) identified 3 scenarios where estimation of p^*_t would still be unbiased when closure was violated. (a) If movement in and out of the study area is completely random during the period, then the estimator for p^*_t remains unbiased. The other 2 exceptions require that detection probability vary only by time and might apply most with migratory populations. (b) If the entire population is present at the first session within a period but begins to leave before the last session, then the estimator is unbiased if detection histories are pooled for sessions 2 to S_t. This creates a new 2-session detection history within period. (c) Conversely, if sampling begins before all animals in the population have arrived but they are all present in the last session, then sessions 1 to S_t should be pooled.

(2) If temporary emigration between periods occurs and is Markovian then variation in detection probability is limited to time dependence and trap-happy or trap-shy response to detection. If temporary emigration is completely random or absent, then there can be other heterogeneity in detection probability. However, an ad hoc approach to estimation must be taken. Program MARK could still be used in this case. The recaptures-only option could be used with pooled detection histories to estimate survival. One could use it to invoke program CAPTURE to estimate p^*_t and N_t and then solve for $\hat{\gamma}_t$ as described above.

(3) Temporary emigration is assumed to be either completely random, Markovian, or based on a temporary response to first capture.

(4) Survival rate is assumed to be the same for all animals in the population, regardless of availability for capture. This is a strong assumption, especially in the Markovian availability case.

ANALYSIS USING PROGRAM MARK

Program MARK includes analysis based on the robust design as 1 of its options. It goes beyond program RDSURVIV in 2 ways. First, as with all of the options for analyzing long-term data sets, it uses a parameterization of the model that allows for modeling parameters as functions of individual covariates. Second, it offers 2 different parameterizations of the closed population part of the model. The first includes N_t as a parameter in the model, as is done with the closed-population model option in MARK. The advantages of this approach include the direct estimation of abundance, its variance and covariances with other parameters, and the ability to include parsimonious models where N_t is set equal across time. The disadvantage of this approach is that neither p_{ts} nor N_t can be modeled as a function of individual covariates because the covariate cannot be measured for those that are never detected within the period. The second option is to incorporate the parameterization of Huggins (1991) for the closed-population part of the model. This approach assumes each animal is captured in at least 1 session of a given period and thus allows p_{ts} to be modeled as a function of individual covariates. The disadvantage, as with the approach of Kendall et al. (1997), is that N_t is not included in the model and its derived estimator is not quite as precise.

As with any statistical software, the data must be in the correct format and the attributes of the program must be set for the appropriate model. Other papers in this proceedings describe the general attributes and features of program MARK. A few aspects of setup for the robust design merit discussion. First, one has to account for the fact that the program is a hybrid of open- and closed-population models. I illustrate this with the example data set provided with MARK. The study consists of 5 years. There are 2, 2, 4, 5, and 2 sessions, respectively, within periods 1–5, totaling 15 sessions. Program MARK still reads the capture history as a string of 0 and 1 numbers. This is accounted for by the user in the option to set time intervals. In this case, there are 14 intervals that separate the 15 sessions. For this example, the setting in MARK would be:

<div align="center">0 1 0 1 0 0 0 1 0 0 0 0 1 0</div>

A 0 indicates there is no time between sessions and the population is closed, and a 1 indicates 1 time period and coincides with the last session in each period.

Second, the default robust design model in MARK includes both time variation and trap response in detection probability within each period. With some exceptions discussed below, one cannot estimate N_t nor p_t^* with both these sources of variation unless there is a relationship assumed between p_{ts} and c_{ts}. This is also true for the purely closed-population case. In MARK, this is accomplished by the design matrix. Cooch (2001) introduces the topic of design matrices in MARK. I compare design matrices below that have been simplified to include only the p_{ts}'s and c_{ts}'s from a 3-session period:

	β_{t1}	β_{t2}	β_{t3}	β_{t4}	β_{t5}
p_{t1}	1	0	0	0	0
p_{t2}	0	1	0	0	0
p_{t3}	0	0	1	0	0
c_{t2}	0	0	0	1	0
c_{t3}	0	0	0	0	1

	β_{t1}	β_{t2}	β_{t3}	β_{t4}
p_{t1}	1	0	0	0
p_{t2}	0	1	0	0
p_{t3}	0	0	1	0
c_{t2}	0	1	0	1
c_{t3}	0	0	1	1

The first is the identity matrix, which indicates a unique effect for each of the 5 parameters. The second matrix indicates that the effect for c_{ts} is the effect for p_{ts} plus an additional effect that is common for all s. Assuming a logit link, an example can be expressed as

$$logit(c_{t2}) = \beta_{t2} + \beta_{t4},$$

where β_{t2} is the effect for initial detection probability in session 2 and β_{t4} is the additive (within-period) trap response effect.

The general issue of estimability of parameters is worthy of attention here. Although MARK has an algorithm for determining which parameters are estimable, this can be a problem with the robust design because problems are not always due to simple confounding between parameters. When the phenomenon of temporary emigration is not present, then ϕ_t, p_{ts}, and B_t (the number of recruits that enter the population between periods t and $t+1$) are estimable for every period. In addition, for all but periods 1 and T, c_{ts} is estimable without making any assumptions about its relationship to p_{ts}. This is due to the second source of information on p_{ts} coming from between-period information.

When temporary unavailability for detection is present, and it is a completely random process (i.e., $\gamma_t'' = \gamma_t' = \gamma_t$), then $(1 - \gamma_T)$ is confounded with ϕ_{T-1}. When this unavailability is Markovian, bias in other parameters will occur if some constraints are not made (Kendall et al. 1997). Parameters γ_T'' and γ_T' must be set constrained over time (e.g., $\gamma_T'' = \gamma_{T-1}''$, $\gamma_T' = \gamma_{T-1}'$). In addition, for trap response within period, a relationship

between p_{ts} and c_{ts}, such as described in the design matrices above, must be included. An exception to this is when survival rate or availability for capture is set equal across periods.

DISCUSSION

There are many advantages to using Pollock's robust design in capture–recapture studies. The benefits described are based on the fact that there is a second source of information on detection probability (from the sessions within period). In addition to its benefits in setting up studies of population dynamics, it also is useful in studying community dynamics, where detection implies detecting any member of a population (see Nichols and Conroy 1996, Nichols et al. 1998). The robust design usually is needed in this case because of great heterogeneity in detection probability among species.

Schwarz and Stobo (1997) presented an approach to analyzing data under the robust design where animals enter the study area once and leave once during the sampling period. They were able to estimate breeding probability of grey seals (*Halichoerus grypus*) while relaxing the assumption of closure within periods.

There really are no disadvantages to conducting a study using the robust design. In fact, as stated above, most studies are in some sense already conducted using this design. The only case where estimation methods based on the robust design would not be an improvement over use of the JS method is if very little sampling effort is devoted to each session. In that case, detection probability for a given session can be so small that the estimation routine becomes unstable (Kendall et al. 1995). However, in that event, the data can be pooled and methods like the JS method used.

Program MARK provides a convenient tool for analyzing data under the robust design. As with any statistical software, the user should be cautious that he/she knows what model he/she is using to produce estimates and its assumptions and limitations. Given that knowledge, these models can be easily implemented in MARK or in RDSURVIV (Kendall and Hines 1999). Program MARK provides the ability to model parameters as functions of covariates, and has the added advantage of fit testing, model selection, and model averaging features described elsewhere in this volume (Cooch 2001, White et al. 2001). I encourage the reader to think of his/her next capture–recapture study in the context of Pollock's robust design.

Acknowledgments. I thank G. White for his initial instruction in the use of MARK and E. Cooch for much assistance in the fine points of the program's use.

LITERATURE CITED

COOCH, E. 2001. First steps with program MARK: linear models. Pages 343–349 *in* R. Field, R. J. Warren, H. Okarma, and P. R. Sievert, editors. Wildlife, land, and people: priorities for the 21st century. Proceedings of the Second International Wildlife Management Congress. The Wildlife Society, Bethesda, Maryland, USA.

HUGGINS, R. M. 1991. Some practical aspects of a conditional likelihood approach to capture experiments. Biometrics 47:725–732.

KENDALL, W. L. 1999. Robustness of closed capture-recapture methods to violations of the closure assumption. Ecology 80:2517–2525.

———, AND J. E. HINES. 1999. Program RDSURVIV: an estimation tool for capture–recapture data collected under Pollock's robust design. Bird Study 46 (suppl):S32–38.

———, AND J. D. NICHOLS. 1995. On the use of secondary capture–recapture samples to estimate temporary emigration and breeding proportions. Journal of Applied Statistics 22:751–762.

———, ———, AND J. E. HINES. 1997. Estimating temporary emigration using capture–recapture data with Pollock's robust design. Ecology 78:563–578.

———, AND K. H. POLLOCK. 1992. The robust design in capture–recapture studies: a review and evaluation by Monte Carlo simulation. Pages 31–43 *in* D. R. McCullough and R. H. Barrett, editors. Wildlife 2001: populations. Elsevier, London, United Kingdom.

———, ———, AND C. BROWNIE. 1995. A likelihood-based approach to capture–recapture estimation of demographic parameters under the robust design. Biometrics 51:293–308.

NICHOLS, J. D., T. BOULINIER, J. E. HINES, K. H. POLLOCK, AND J. R. SAUER. 1998. Estimating rates of local species extinction, colonization, and turnover in animal communities. Ecological Applications 8:1213–1225.

———, AND M. J. CONROY. 1996. Estimation of species richness. Pages 226–234 *in* D. E. Wilson, F. R. Cole, J. D. Nichols, R. Rudran, and M. Foster, editors. Measuring and monitoring biological diversity. Standard methods for mammals. Smithsonian Institution Press, Washington, D.C., USA.

———, AND K. H. POLLOCK. 1990. Separate estimation of recruitment from immigration versus in situ reproduction using Pollock's robust design. Ecology 71:21–27.

OTIS, D. L., K. P. BURNHAM, G. C. WHITE, AND D. R. ANDERSON. 1978. Statistical inference for capture data on closed animal populations. Wildlife Monographs 62.

POLLOCK, K. H. 1982. A capture–recapture design robust to unequal probability of capture. Journal of Wildlife Management 46:757–760.

———, J. D. NICHOLS, J. E. HINES, AND C. BROWNIE. 1990. Statistical inference for capture–recapture experiments. Wildlife Monographs 107.

REXSTAD, E., AND K. P. BURNHAM. 1991. User's guide for interactive program CAPTURE. Colorado Cooperative Fish and Wildlife Research Unit, Fort Collins, USA.

SCHWARZ, C. J., AND W. T. STOBO. 1997. Estimating temporary migration using the robust design. Biometrics 53:178–194.

WHITE, G. C. 1983. Numerical estimation of survival rates from band-recovery and biotelemetry data. Journal of Wildlife Management 47:716–728.

———, K. P. BURNHAM, AND D. R. ANDERSON. 2001. Advanced features of program MARK. Pages 368–377 *in* R. Field, R. J. Warren, H. Okarma, and P. R. Sievert, editors. Wildlife, land, and people: priorities for the 21st century. Proceedings of the Second International Wildlife Management Congress. The Wildlife Society, Bethesda, Maryland, USA.