Estimating population size by means of mark-resighting counts: theoretical considerations and empirical results

L. Fattorini (1) , M. Marcheselli (1) , A. Monaco (2) and C. Pisani (1)

(1) *Dipartimento di Metodi Quantitativi, Università di Siena, Siena (Italy)*
(2) *Agenzia Regionale Parchi, Regione Lazio, Roma (Italy)*

Summary

1. For many species and circumstances, mark-resighting procedures constitute valid alternatives to capture-recapture methods. Indeed, resightings are generally cheaper to acquire than physically recapturing and rehandling the animals, especially when radiotelemetry or other tracking devices are available.
2. In order to estimate population abundance, the joint hypergeometric maximum likelihood estimator, the Minta-Mangel estimator and the Bowden estimator are implemented in NOREMARK, software which has become very popular with biologists in the past decade.
3. In this paper, the basic assumptions regarding these widely applied procedures are delineated and discussed. Some shortcomings of the software are outlined, and a simulation study is performed in order to investigate the robustness of the estimators under failure of the assumptions.
4. Theoretical considerations and simulation results motivate the use of the Bowden estimator, which when marks are quite evenly distributed among groups, constitutes the sole reliable method, offering computational simplicity and robustness. An application to a case study is considered.

Keywords: Bowden estimator, Joint hypergeometric estimator, Minta-Mangel estimator, NOREMARK software, simulation studies.

1. Introduction

The estimation of size for a geographically and demographically closed but free-ranging population is a common problem encountered by wildlife biologists. The earliest approaches to this problem were developed in the framework of capture-recapture techniques. For basic reviews on capture-recapture methods, see Otis *et al*. (1978).

Mark-resighting methods constitute technologically advanced approaches to abundance estimation, usually involving radio transmitters or other tracking devices for marking animals. Arnason, Schwarz & Gerrard (1991) first adopted the terms *marking and sighting experiments* to identify techniques in which a sample of marked animals is formed in the population and subsequent resightings of marked and unmarked animals are used to estimate the population size.

More in detail, in mark-resighting experiments, a sample of animals is captured and marked (usually with radios) and resightings (rather than recaptures) are performed during subsequent occasions. Obviously, the advantage of mark-resighting with respect to capture-recapture procedures is that resightings are generally cheaper to acquire than physically recapturing and rehandling the animals and they are also less disruptive to the animals and their environment. Moreover, the use of radios allows for recognition of the marked animals that have died or emigrated from the study area, in such a way that the number of marked animals present in the study region is monitored throughout the survey period. Unfortunately, the decrease of field effort is accompanied by a considerable reduction of the collected information with respect to that obtained using capture-recapture methods. Indeed, mark-resighting methods only involve counting the animals spotted during the resighting occasions and identifying those that are marked. Hence, the huge list of capture-recapture methodologies based on the capture histories of each captured animal cannot be adopted with mark-resighting data, in which only the resighting histories of the marked animals are available.

Mark-resighting methods have been tested with a known population of mule deer (Bartman *et al*. 1987) and subsequently used with white-tailed deer (Rice & Harder 1977), mountain sheep (Furlow, Haderlie & Van den Berge 1981, Neal *et al.* 1993), desert bighorn sheep (Leslie & Douglas 1979 and Leslie & Douglas 1986), black and grizzly bears (Miller, Becker & Ballard 1987), coyotes (Hein & Andelt 1995) and harbour seals (Ries, Hiby & Reijnders 1998) among others.

In order to estimate population abundance by means of mark-resighting data, the joint hypergeometric maximum likelihood estimator, the Minta-Mangel estimator and the Bowden estimator are implemented in NOREMARK (see White 1996), software which has become very popular with biologists in the past decade. Even if widely applied, the assumptions underlying these estimators are not clearly delineated in literature and they are likely not well recognized by biologists. Moreover, no variance estimation is performed in NOREMARK for both joint hypergeometric maximum likelihood and Minta-Mangel estimators. In the first case, confidence intervals are constructed directly from the likelihood function by means of the Hudson (1971) criterion, while in the second case they are constructed from a Monte Carlo distribution. However, several simulation studies highlight the unreliability of these intervals, which usually show coverage levels much smaller than their nominal counterparts.

The aim of this paper is to identify and discuss the basic assumptions actually needed for the three estimators adopted in NOREMARK. The Minta-Mangel procedure is also criticized from both computational and theoretical points of view and an alternative pseudo-maximum likelihood estimator is proposed together with a conservative variance estimator. Moreover, since the assumptions underlying all the estimation criteria are very unrealistic, a simulation study is performed to investigate the robustness of these estimators when mark-resighting data arises from more realistic scenarios. Finally, NOREMARK is checked in a practical application to estimate the size of a population of chamois *Rupicapra rupicapra* (Linnaeus, 1758) females in the Paneveggio - Pale di San Martino Natural Park (Trentino, eastern Italian Alps). The resulting estimates are then interpreted on the basis of the theoretical and empirical findings achieved in the work.

Throughout the paper, N denotes the abundance of a closed population, say $\mathcal{P}$, on a study area and constitutes the target parameter, M denotes the size of the initial portion, say $\mathcal{M}$, of animals captured and marked, $U = N - M$ denotes the size of the unmarked portion, say $\mathcal{U} = \mathcal{P} - \mathcal{M}$, and T denotes the number of resighting occasions performed in the experiment. Moreover, n_t denotes the number of animals resighted at occasion t and m_t denotes the number of marked animals resighted out of the n_t , while $n = n_1 + \ldots + n_T$ represents the overall number of resightings and $m = m_1 + \ldots + m_T$ represents the overall number of resightings for marked animals, in such a way that $n - m$ gives the overall number of resightings for unmarked animals. Finally, x_i denotes the total number of resightings for animal i in T occasions, M_x is the number of marked animals spotted x times ($x = 0,1,\ldots,T$) while $\bar{x}$ and s_x^2 respectively denote the mean and the variance of the x_i's for $i \in \mathcal{M}$. It is worth noting that s_x^2 will be referred to as the *sampling variance* when the sum of squared deviations is divided by M, while it is referred to as the *unbiased sampling variance* if $M - 1$ is adopted instead of M.

2. NOREMARK estimation procedures

2.1 Joint hypergeometric maximum likelihood estimator

The early estimator adopted in NOREMARK is the so-called *joint hypergeometric maximum likelihood estimator* (JHE), proposed first by Bartmann *et al*. (1987), and subsequently investigated by White & Garrot (1990), Neal *et al.* (1993) and White (1993). The JHE is the value $\hat{N}_{JH}$ which maximizes the joint hypergeometric likelihood function of $m_1,\ldots,m_T$, say

$$L(N|m_1,\ldots,m_T)=\prod_{t=1}^{T}\frac{\binom{M}{m_t}\binom{N-M}{n_t-m_t}}{\binom{N}{n_t}} \tag{1}$$

For $T>1$, the maximum likelihood estimate is obtained by means of an iterative numerical procedure. On the other hand, when $T=1$, the estimate reduces to the well-known Lincoln-Peterson index (Le Cren 1965).

It is at once apparent that (1) represents the actual likelihood of $m_1,\ldots,m_T$ if and only if the following assumptions hold:

- *a)* at occasion *t* a pre-fixed number of n_t animals is selected out of the N by means of simple random sampling without replacement ($t=1,\ldots,T$);
- *b)* the *T* selections are performed independently.

Indeed, under assumptions *a*) and *b*), $m_1,\ldots,m_T$ constitute *T* independent hypergeometric random variables each with parameters *N*, *M* and n_t , where *M* and $n_1,\ldots,n_T$ are known quantities established when planning the mark-resighting experiment and *N* is the sole unknown quantity to be estimated. Any violation of one or both these assumptions leads to likelihood deviating from (1). Practically speaking, assumptions *a*) and *b*) entail that, at each time, the animal resightings should be performed as if n_t balls were randomly drawn without replacement from an urn containing M black balls (marked animals) and *U* white balls (unmarked animals) (on this topic, see also Seber, 1982, section 3.2.2). Moreover, after each sampling occasion, the selected balls would be replaced in the urn in order to make the subsequent selections independent from the previous ones.

As to assumption *a*), animals are spread over the study area in such a way that there is no possibility of sampling them by means of SRSWOR. In most situations encounter schemes are adopted, in which the sampled animals are those sighted from *T* paths (transects or observational points) thrown onto the study area in accordance with a planned design. As a consequence, the number of animals selected/resighted at each occasion cannot be pre-fixed but rather it constitutes a random variable. Thus, under encounter schemes, each m_t invariably differs from the hypergeometric distribution in which n_t is a fixed parameter.

Notwithstanding this, most of the literature regarding JHE claims that the crucial assumption underlying (1) is the use of schemes ensuring the same resighting probability for each animal at any occasion. White & Garrot (1990, p 263) state that "*Each individual must have the same probability of capture or sighting as every other individual on a particular occasion*". Subsequently, Neal *et al.* (1993, p.437) emphasize that "*all animals (both marked and unmarked) must have the same, independent probability of being captured on a trapping occasion or of being sighted during an individual sighting occasion*" and White (1993, p.92) outlines the need for "... *extensive simulations of the JHE, evaluating the robustness of the estimator to heterogeneity of sighting probabilities, and to lack of independence of sightings*", thus introducing the assumption of independence among animal sightings, which is unnecessary in this framework. Moreover, when introducing the NOREMARK software, White (1996, p.50) says that "*This estimator assumes that each animal in the population has the same sighting probability on an occasion as every other animal (no individual heterogeneity), but sighting probabilities can vary across occasions*" and that "*Sighting probability is assumed the same for all animals on any particular survey, and animals are assumed to be sampled without replacement*" and White & Shenk (2001, p. 336) state that JHE assumes "... *equal sighting probabilities among animals on a particular occasion. In addition, animals are assumed to be sampled without replacement*", even if the without-replacement sampling is clearly in contrast with the independence previously claimed by White.

These sentences may be highly misleading for a biologist who is not trained in sampling theory. Indeed, many encounter schemes can be constructed in which animals enter the sample without replacement and with the same inclusion/resighting probabilities (see *e.g.* the line transect scheme and the variable circular plot scheme described in Thompson 1992, Chapter 17). Notwithstanding this, all these schemes greatly differ from SRSWOR. Indeed, when SRSWOR is performed at any time *t*, each animal has a pre-fixed inclusion probability equal to n_t / N . On the other hand, in an encounter scheme ensuring the same inclusion/resighting probability for each animal, this probability turns out to be $E(n_t)/N$, where $E(n_t)$ is unknown and depends on many factors (*e.g.* the nature of encounter scheme, detectability conditions, animal density and behaviour, observer ability and so on). Moreover under SRSWOR any couple of animals has a joint inclusion probability equal to $\{n_t(n_t-1)\}/\{N(N-1)\}$, while in an encounter scheme these probabilities obviously depend on the relative locations of the two animals (*e.g.* two animals very far apart cannot be jointly included/resighted from the same observational point). Accordingly, the distribution of m_t may greatly differ from the hypergeometric model even in the presence of schemes ensuring equal sighting probabilities.

As a very simple example, consider an artificial population of $N = 10$ turtles settled in a square region of size 10×10 (see Fig. 1) and suppose that the turtles labelled as 2,3,6 and 9 are marked, for a total of $M = 4$ marked animals. Moreover, the population is supposed to be sampled during a single resighting occasion ($T = 1$) by means of strip sampling in which a point is randomly selected on the baseline, a transect starting from that point perpendicular to the baseline is travelled and all the turtles lying in the strip of width 2 centred at the transect are selected. Suppose also that all the turtles in the strip are spotted and that their movements are so slow relative to the observer's progression down the transect, that they can be considered as immobile units. Table 1*a* and 1*b* respectively show the design (all the possible samples with their corresponding probabilities) arising from this encounter scheme as well as the resulting distribution of m_1. It is at once apparent from Table 1*a* that the inclusion probability of each turtle (obtained as the sum of the probabilities of samples containing that turtle) invariably equals 0.2, while the second-order inclusion probability of each couple of turtles (obtained as the sum of the probabilities of samples containing that couple) strictly depends on the distance between the two animals. For example, the probability of jointly resighting turtles 3 and 4 turns out to be 0.10 while the probability of jointly resighting turtles 1 and 10 turns out to be 0. On the other hand, if a sample of $n_1 = 3$ turtles were selected from the population by SRSWOR, then any of the 240 possible samples would have an equal probability of $1/240$, in such a way that each turtle had an inclusion probability of 0.3 and each couple of turtles had joint inclusion probabilities of $1/15$. In this case, the resulting distribution of m_1 would be hypergeometric with parameter $n_1 = 3, N = 10, M = 4$, *i.e.* $\Pr(m_1 = 0) = 5/30$, $\Pr(m_1 = 1) = 15/30$, $\Pr(m_1 = 2) = 9/30$ and $\Pr(m_1 = 3) = 1/30$. These probabilities deviate from the actual distribution of m_1 conditional to $n_1 = 3$, which on the basis of Table 1b turn out to be $\Pr(m_1 = 1) = 3/4$ and $\Pr(m_1 = 2) = 1/4$. Accordingly, even if strip sampling ensures an equal inclusion probability for each turtle, it greatly differs from SRSWOR.

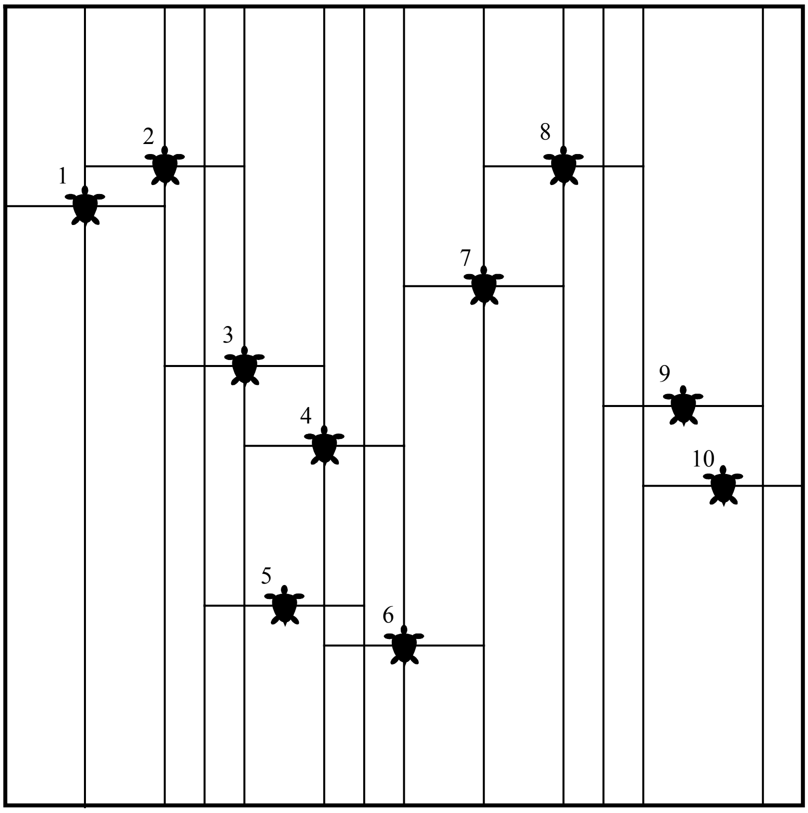
Table 1*a*. Sampling design arising from the strip sampling performed on the artificial population in Fig. 1.

selected turtles	probability	n_1	m_1
1	0.10	1	0
1,2	0.10	2	1
2,3	0.05	2	2
2,3,5	0.05	3	2
3,4,5	0.10	3	1
4,5,6	0.05	3	1
4,6	0.05	2	1
6,7	0.10	2	1
7,8	0.10	2	0
8	0.05	1	0
8,9	0.05	2	1
9,10	0.15	2	1
10	0.05	1	0

Table 1*b*. Probability distribution of n_1 and m_1 arising from the strip sampling performed on the artificial population in Fig. 1.

n_1	m_1	probability
1	0	0.20
2	0	0.10
2	1	0.45
2	2	0.05
3	1	0.15
3	2	0.05

Fig. 1. Spatial distribution and labels for an artificial population of ten turtles spread over a square region of size 10×10.



Finally, as to assumption *b*), the independence between the *T* occasions turns out to be quite unrealistic and/or difficult to obtain when encounter sampling schemes are adopted. Indeed, in any encounter scheme, the probability of selecting/resighting an animal at time *t* obviously depends on its closeness to the path (transects or observational points) travelled by the observer. In turn, the animal position at time *t* may depend on its positions at the previous $t-1$ times. Accordingly, even if the resightings among occasions were independent, conditional on the animal positions (for example this may be ensured by using transects or points independently thrown on the study area at each occasions), they may not be independent with respect to the animal movements during the survey period.

Even this aspect is not clearly recognized in literature regarding JHE. For example, Bartman *et al.* (1987, p.42) say that "*Because individual flights for each pasture were independent, the product of the hypergeometric likelihood function can be numerically* *optimised to estimate N*'. A very similar sentence is also contained in White & Garrot (1990, p.260).

No estimation for the sampling variance of $\hat{N}_{JH}$ is performed in NOREMARK software. Rather, as suggested by Bartman *et al.* (1987), a confidence interval for N is constructed directly from (1) by means of the Hudson (1971) procedure. The method is simply based on the fact that under assumptions *a*) and *b*), the quantity $-2\left\{l(N)-l(\hat{N}_{JH})\right\}$, where $l(N)=\ln L(N\mid m_1,\ldots,m_T)$, is approximately distributed as a $\chi^2_{(1)}$. Accordingly, if $\chi^2_{(1),1-\alpha}$ denotes the $(1-\alpha)$-quantile of the chi-square distribution with one degree of freedom, the set

$$\left\{N: l(\hat{N}_{JH})-l(N)\leq\frac{\chi^2_{(1),1-\alpha}}{2}\right\} \tag{2}$$

constitutes a confidence region for N with approximate coverage $(1-\alpha)$. Practically speaking, for unimodal likelihoods, (2) coincides with the interval obtained by drawing a line across the log-likelihood at the level $l(\hat{N}_{JH})-\chi^2_{(1),1-\alpha}/2$ and projecting the two intersections onto the x-axis. As pointed out by White & Garrot (1990, p.262), an advantage of this technique is that the lower bound of the resulting interval is never less than the minimum number of animals known to exist (*i.e.* the number of marked animals plus the largest number of unmarked animals resighted on any occasion). The obvious shortcoming is that the intervals are based on likelihoods of type (1) which do not constitute the actual likelihoods of mark-resighting data.

2.2 Minta-Mangel estimator

The other estimator adopted in NOREMARK is the so-called *Minta-Mangel estimator* (MME), proposed by Minta & Mangel (1989) and subsequently investigated by Neal *et al.* (1993) and White (1993). In order to introduce MME, denote by $\mathbf{p}=\left[p_0,p_1,\ldots,p_T\right]^{\mathrm{T}}$ the *empirical* resighting distribution observed for the M marked animals, where

$$p_x=\frac{M_x}{M}\quad,\; x=0,1,\ldots,T\,.$$

Then, a sequence of random variables $X_1, X_2,\ldots$ is sequentially generated from the discrete probability distribution defined by $\mathbf{p}$ until their sum (which in this framework constitutes a realization of the total number of resightings) equals or exceeds the actual number of resightings $n-m$ for unmarked animals in the T occasions. Obviously, the corresponding waiting time (*i.e.* the number of generated variables), say u, may be viewed as the number of unmarked animals giving rise to at least n-m resightings. Then, the procedure is repeated B times, obtaining $u_1,\dots,u_B$, and the mode of the resulting Monte Carlo distribution, say $\tilde{U}_{MM}$, is taken as the estimate of U. Obviously, $\tilde{N}_{MM} = M + \tilde{U}_{MM}$ constitutes the final estimate of N. It is worth noting that in the implementation of the procedure adopted in NOREMARK, only the sequences whose sum is exactly equals to n-m are used (White, 1996, p.51).

It is at once apparent that the Monte Carlo procedure leading to MME makes sense if and only if the following assumption holds:

c) $X_1,\dots,X_N$, constitute a set of *iid* random variables.

Indeed, the primary aim of the Monte Carlo procedure proposed by Minta & Mangel (1989) is to determine empirically the probability distribution of the discrete random variable

$$U = \inf\{u : X_1 + \dots + X_u \geq n-m\} \tag{3}$$

with the support $U \geq (n-m)/T$. Alternatively, in the NOREMARK procedure, the random variable of interest is changed into

$$U = \inf\{u : X_1 + \dots + X_u = n-m\} \tag{4}$$

However, when describing their Monte Carlo procedures, neither Minta & Mangel (1989, p.1751) nor White (1996, p.51) refer to any correlation structure adopted for sequentially generating the sequence $X_1, X_2 \dots$ from **p**. Then, it is quite natural to deduce that these random variables are generated independently. However, since in this framework $X_1, X_2 \dots$ represent the number of resightings for the corresponding sequence of unmarked animals $1,2\dots$, then animal resightings are tacitly presumed to be independent events. Moreover, since the X_i's are generated from the same distribution, then it is also tacitly assumed that they are equally distributed for all $i \in \mathcal{U}$. Finally, since the true common distribution, say $\boldsymbol{\pi} = [\pi_0, \pi_1, \dots, \pi_T]^{\mathrm{T}}$ is actually unknown, the empirical distribution **p** observed in the set of marked animals is adopted to generate the resightings of unmarked animals. Obviously, this makes sense only if the M marked animals have the same resighting distribution as the unmarked ones. Practically speaking, *c*) is assumed but never mentioned in both the works by Minta & Mangel (1989) and White (1996).

Moreover, it is worth noting that assumption *c*) also ensures that, for any sub-set $\mathcal{M}$ of M animals, the vector $\mathbf{M} = [M_0, M_1, \dots, M_T]^{\mathrm{T}}$ constitutes a multinomial random vector with parameters M and $\boldsymbol{\pi}$. Thus, assumption *c*) actually ensures that $\mathbf{p}$ turns out to be the minimum variance unbiased estimator of $\boldsymbol{\pi}$, irrespective of the sub-set $\mathcal{M}$ initially selected for capturing and marking the animals.

Any violation of assumption *c*) should lead to Monte-Carlo procedures less simple than the straightforward procedures proposed by Minta & Mangel (1989) or adopted in NOREMARK. Unfortunately, assumption *c*) turns out to be quite unrealistic when encounter sampling schemes are adopted to sample/resight animals. Indeed, even if the same resighting distribution for each animal in the population could be ensured (as already pointed out in section 2.1), the assumption of resighting independence is completely false in the presence of any tendency of the animals to aggregate, to defend territories or to be affected by a patchy environment. Indeed it is quite obvious that if a group is spotted, all the animals in the group are likely to be spotted. In this case the number of resightings for each animal cannot not be generated independently to obtain the empirical distribution of *U*.

Surprisingly, the literature regarding MME completely neglects the need and the role of independence among resightings in the estimation criterion. For example, Minta & Mangel (1989, p.1741) assume that "*The initial captures are a random sample of the population, and marked animals are identical to unmarked animals in detection or observability*", while it is apparent that if *marked animals are identical to unmarked animals in detection or observability*, then the (tacitly assumed) independence among animal resightings suffices to justify the use of $\mathbf{p}$ even if the M marked animals do not constitute a *random sample* from the population. No mention of the independence of resightings is contained in the subsequent articles regarding MME.

Moreover, MME is often presented as a suitable alternative to JHE allowing for different resighting probabilities among animals. Neal *et al.* (1993) emphasize that MME "... *should provide a method robust to heterogeneity of individual sighting probabilities*" and White (1996) states that "*The estimator does not assume that sighting probabilities are the same for each animal on a particular occasion* ..." (see also White & Shenk, 2001, p.333). These sentences may be highly misleading for a biologist not trained in statistical modelling. Indeed, they might be dangerously interpreted as if the Minta-Mangel and NOREMARK simulation procedures remain valid even when animals have different resighting distributions $\boldsymbol{\pi}_1,\ldots,\boldsymbol{\pi}_N$. Rather, these sentences must be correctly interpreted as generalizations of the well-known heterogeneity models (M_h models) which are widely applied in capture-recapture experiments (see *e.g.* Otis *et al*, 1978, p. 11). Thus, generalizing Burnham & Overton (1978), suppose that the random vector $\mathbf{Z}_i = [Z_{1i},\ldots,Z_{Ti}]^{\mathrm{T}}$, with $Z_{ti} = 1$ if animal *i* is spotted at time *t* and $Z_{ti} = 0$ otherwise, has a probability generating function $G_Z(\mathbf{s};\varphi_i)$, where in turn $\varphi_i = [\varphi_{1i},\ldots,\varphi_{Ti}]^{\mathrm{T}}$ is the vector of the resighting probabilities φ_{ti} for animal *i* at occasion *t*. Moreover, suppose that the $\mathbf{Z}_i$'s are mutually independent for given $\varphi_1,\ldots,\varphi_N$ and that $\varphi_1,\ldots,\varphi_N$ are *iid* random vectors from a probability distribution $F(\cdot;\boldsymbol{\theta})$. Accordingly, conditional on φ_i, each $X_i = \mathbf{1}^{\mathrm{T}}\mathbf{Z}_i$ has the probability generating function $G_i(s) = G_Z(s\mathbf{1};\varphi_i)$ which varies with *i*. On the other hand, with respect to the possible realizations of φ_i, the probability generating function of each X_i turns out to be

$$G_X(s;\boldsymbol{\theta}) = \int G_Z(s\mathbf{1};\varphi)dF(\varphi;\boldsymbol{\theta}).$$

Thus, even if the X_i may have heterogeneous distributions conditional on the resighting probability vectors, they have the same distribution when merged with respect to the random mechanism generating these probability vectors.

As previously pointed out, the primary aim of the Minta-Mangel and NOREMARK simulation procedures is to derive empirically the probability distribution of the discrete random variables (3) and (4), respectively. However, under assumption *c*), these variables constitute trivial generalizations of the Pascal waiting time random variable. Thus, their probability distributions can be straightforwardly obtained with no need of simulations.

To this purpose it may be convenient to rewrite the random variable $Y = X_1 + \ldots + X_U$ as $Y = \mathbf{c}^{\mathrm{T}}\mathbf{U}$, where $\mathbf{c} = [0,1,\ldots,T]^{\mathrm{T}}$, $\mathbf{U} = [U_0,U_1,\ldots,U_T]^{\mathrm{T}}$ and U_t denotes the number of animals spotted *t* times out of the *U*. Owing to assumption *c*), $\mathbf{U}$ is a multinomial random vector with parameters *U* and $\boldsymbol{\pi}$. Accordingly, the probability generating function of *Y* turns out to be the polynomial of degree *UT*

$$L_Y(s;U,\boldsymbol{\pi}) = (\pi_0 + \pi_1 s + \ldots + \pi_T s^T)^U = \sum_{l=0}^{UT} a_l(U,\boldsymbol{\pi})s^l$$

in such a way that

$$\Pr(Y = y) = \left[\frac{\partial^y L(s;U,\boldsymbol{\pi})}{\partial s^y}\right]_{s=0} = a_y(U,\boldsymbol{\pi}) \tag{5}$$

where $a_l(U,\boldsymbol{\pi})$ represents the coefficient of the *l* degree term in the polynomial $L_Y(s;U,\boldsymbol{\pi})$. From a practical point of view, it is worth noting that the probability distribution (5) may be readily computed by means of an iterative procedure, starting from the *T*-degree polynomial $L_Y(s;1,\boldsymbol{\pi})$ and adopting the recursive relation

$$L_Y(s;U,\boldsymbol{\pi}) = L_Y(s;U-1,\boldsymbol{\pi})L_Y(s;1,\boldsymbol{\pi}) \tag{6}$$

Moreover, as to the multinomial distribution with parameters U and $\boldsymbol{\pi}$, it is also worth noting that some π_t may be allowed to be 0 providing that 0^0 is set to be 0. In this case, denote by $T_{\max} \le T$ the maximum value of *t* for which $\pi_t > 0$. Thus, in accordance with the previous results, the probability distribution of the random variable (3) turns out to be

$$f(1;n-m,\boldsymbol{\pi}) = \sum_{j=n-m}^{T_{max}} \pi_j$$

for $u = 1$ and $n - m \le T_{\max}$,

$$f(u;n-m,\boldsymbol{\pi}) = \sum_{j=n-m}^{T_{max}} \pi_j \sum_{k=1}^{min(n-m,j)} \Pr(X_1 + \ldots + X_{u-1} = n-m-k) = \sum_{j=n-m}^{T_{max}} \pi_j \sum_{k=1}^{min(n-m,j)} a_{n-m-k}(u-1,\boldsymbol{\pi})$$

for $u \ge 2$ and $u \ge (n-m)/T_{\max}$, and 0 otherwise.

In an analogous way, the probability distribution of the random variable (4) turns out to be

$$f(1;n-m,\boldsymbol{\pi}) = \pi_{n-m}$$

for $u = 1$ and $n - m \le T_{\max}$,

$$f(u;n-m,\boldsymbol{\pi}) = \sum_{j=n-m}^{min(n-m,T_{max})} \pi_j \Pr(X_1 + \ldots + X_{u-1} = n-m-j) = \sum_{j=n-m}^{min(n-m,T_{max})} \pi_j a_{n-m-j}(u-1,\boldsymbol{\pi})$$

for $u \ge 2$ and $u \ge (n-m)/T_{\max}$, and 0 otherwise. Note that (4) is not a proper random variable since it may happen that $\Pr(U = \infty) > 0$.

In accordance with the previous results, MME may be straightforwardly obtained as the mode of the resulting waiting time distribution in which the nuisance parameter $\boldsymbol{\pi}$ is estimated by $\mathbf{p}$. In other words, $\widetilde{U}_{MM}$ may be analytically obtained as

$$\widetilde{U}_{MM} = \underset{u}{\operatorname{argmax}} \{ f(u;n-m,\mathbf{p}) \} \tag{7}$$

where in the case of (4) the distribution of *U* conditional to $U < \infty$ must be considered.

Estimators of type (7) were applied to the four sets of data reported in Minta and Mangel (1989). Both the random variables (3) and (4) were considered. The achieved results show estimates equal or very close to the empirical values obtained by Minta & Mangel (1989) on the basis of $B = 10{,}000$ simulated samples (see Table 2).

Even if Minta & Mangel (1989, p. 1743, p.1745, p.1751) speak about their proposal as the *maximum likelihood estimator* of *U*, it is at once apparent that $\tilde{U}_{MM}$ does not constitute a maximum likelihood estimator. This fact has also been recognized by White & Shenk (2001, p.333). Indeed, in the Minta-Mangel procedure the total number of resightings *n-m* for unmarked animals is held fixed while the random variable is the number *U* of independent trials (animals) generating that number of resightings (or more). Thus, the mode of the resulting waiting time distribution stopping at *n-m* resightings, being the most reliable number of animals generating *n-m* resightings, is taken as an estimate of *U*. However, in real situations, the number of unmarked animals is fixed and constitutes the unknown parameter of interest while the number *Y* of total resightings generated by these animals is the random variable containing *U* as an unknown parameter of its probability distribution. Thus, the actual maximum pseudo-likelihood estimator of *U* is the integer $\hat{U}_{MM}$ maximizing the pseudo-likelihood function of $n-m$, say

$$L(U \mid n-m) = a_{n-m}(U, \mathbf{p}) \tag{8}$$

in which the nuisance parameter $\boldsymbol{\pi}$ is replaced by $\mathbf{p}$. It is worth noting that the maximization of (8) may be effectively performed by means of the recurrence relation (6). Accordingly, there is no reason, neither theoretical nor practical, to adopt $\tilde{U}_{MM}$ instead of the pseudo-maximum likelihood estimator $\hat{U}_{MM}$ (for comprehensive treatment of pseudo-likelihood methods see Gong & Samaniego 1981). Obviously, the final pseudo-maximum likelihood estimate of *N* is obtained by means of $\hat{N}_{MM} = M + \hat{U}_{MM}$.

Even if in most common situations the two estimation criteria give rise to very similar estimates (see *e.g.* Table 2), there may exist situations giving rise to different values. As a trivial example, consider the case in which $T = 1$, in such a way that $\mathbf{p} = [p_0, p_1]^{\mathrm{T}}$, where p_1 denotes the portion of marked animals resighted at the single occasion. In this case *Y* is a binomial random variable with parameter *U* and π_1 and $\hat{U}_{MM}$ is the greatest integer not exceeding $(n-m)/p_1$. On the other hand, the random variable (4) is negative binomial with parameter $n-m$ and π_1 and $\tilde{U}_{MM}$ is the greatest integer not exceeding $(n-m-1)/p_1+1$. Thus, the two estimates may be quite different for small values of p_1, as when, for example, $n-m=3$ and $p_1 = 0.11$, in which case $\hat{U}_{MM} = 27$ and $\tilde{U}_{MM} = 19$.

No estimator for the sampling variance of $\tilde{U}_{MM}$ (or $\tilde{N}_{MM}$ equivalently) is proposed in Minta & Mangel (1989) nor in the NOREMARK software. Rather, in both cases the confidence intervals are obtained directly by means of the appropriate percentiles of the

Monte Carlo distribution. However, several simulation studies (Neal *et al.* 1993, White 1993) show that the resulting coverage is far below the nominal level. As a consequence of this, the authors realize some problems related to the inference performed on the waiting time distribution rather than on the likelihood function. On this subject, Neal *et al.* (1993, p.449) say that "*The procedure used to compute the CIs described by Minta and Mangel (1989) assumes that the variance of the number of sightings of unmarked animals is zero. This assumption is false in that duplications of the experiment would provide a range in the number of sightings of unmarked animals*" while, in a similar way, White & Shenk (2001, p.335) point out that "*the number of unmarked animals seen is taken as a fixed constant rather than a random variable*". Actually, the confidence intervals deriving from the Minta-Mangel estimation criterion arise from the variability of the waiting time distribution rather than from the variability of the total number of resightings.

Moreover, another reason leading to the shortcoming of these intervals is probably due to the fact that the X_is are generated as independent random variables while they are actually dependent variables. Indeed, such a dependence is likely to entail a contagion of resightings which may increase the variance of Y over the level obtained in the case of independence. This fact is not recognised in literature since, as already pointed out, the independence of resightings is not perceived as a basic assumption of the Minta-Mangel model. For example, Neal *et al.* (1993) simulate aggregate populations, thus violating the assumption of independent resightings, but they conclude that the Minta-Mangel confidence intervals provide a coverage lower that the nominal level "*even though the assumption of the estimator were not violated*" (Neal *et al.* 1993, p.449).

On the other hand, if the pseudo-maximum likelihood estimator $\hat{U}_{MM}$ is adopted, an estimator of the variance of $\hat{U}_{MM}$ (or $\hat{N}_{MM}$ equivalently) which tends to be conservative under assumption *c*) may be straightforwardly attempted. Indeed, as a trivial consequence of assumption *c*), it follows that $\mathrm{E}(Y) = U\mu_x$ and $\mathrm{V}(Y) = U\sigma_x^2$, where μ_x and σ_x^2 denote expectation and variance of the X_is. If the moment criterion is adopted instead of the pseudo-maximum likelihood criterion, then, equating $\mathrm{E}(Y)$ to the observed value $n-m$, the moment estimator of U turns out to be

$$\hat{U}_m = \frac{n-m}{\bar{x}}.$$

Moreover, owing to assumption c), $(n-m)$ and $\bar{x}$ are realizations of independent random variables. Thus, by means of the first-order Taylor series approximation, $\hat{U}_m$ is approximately unbiased with approximate variance

$$\mathrm{V}(\hat{U}_m) = \frac{NU}{M}\frac{\sigma_x^2}{\mu_x^2} \tag{9}$$

Accordingly, since $\hat{U}_m$ constitutes the moment estimator of *U*, it should be less efficient than the pseudo-maximum likelihood estimator $\hat{U}_{MM}$. Thus, from (9),

$$\mathrm{V}(\hat{U}_{MM}) \leq \frac{NU}{M}\frac{\sigma_x^2}{\mu_x^2}$$

in such a way that

$$\hat{\mathrm{V}}(\hat{U}_{MM}) = \frac{\hat{N}_{MM}\hat{U}_{MM}}{M}\frac{s_x^2}{\bar{x}^2}$$

should be a conservative estimator for $\mathrm{V}(\hat{N}_{MM})$. Assuming, as customary, the normality of the pseudo maximum likelihood estimator $\hat{N}_{MM}$, the normality-based interval for *N* at the nominal level $1-\alpha$ turns out to be

$$\hat{N}_{MM} \pm z_{1-\alpha/2}\sqrt{\hat{\mathrm{V}}(\hat{U}_{MM})}$$

where $z_{1-\alpha/2}$ denotes the $1-\alpha/2$ quantile of the standard normal distribution function. As shown in column 6 of Table 2, these confidence intervals tend to be much wider than the unreliable intervals arising from the waiting time distribution.

Table 2. Monte Carlo values of $\tilde{N}_{MM}$ derived by Minta and Mangel (1989) on four real data sets compared with the $\tilde{N}_{MM}$ values obtained from the analytical distributions of the waiting times of type (3) and (4) as well as with the pseudo-maximum likelihood estimates $\hat{N}_{MM}$.

Survey	M	Minta-Mangel (1989)	analytical mode of (3)	analytical mode of (4)	maximum likelihood estimate
badgers in Wyoming	15	24 (21-28)	24 (20-29)	23 (19-29)	23 (18.2-27.9)
bisons in Santa Catalina island (1976)	16	172 (162-182)	172 (161-182)	171 (161-182)	171(137.7-204.3)
bisons in Santa. Catalina island (1977)	14	211 (203-221)	211 (201-222)	211(201-221)	211(173.9-248.1)
porcupines on the Negev Desert	7	28 (21-35)	26 (20-35)	26(20-34)	26 (12.8-39.2)

2.3 Bowden estimator

The more recent estimator adopted in NOREMARK is the estimator originally proposed by Bowden (1993) and subsequently investigated by Bowden & Kufeld (1995). The estimator will be referred to as BE and is given by

$$\hat{N}_B = \frac{\frac{n}{\bar{x}} + \frac{s_X^2}{\bar{x}^2}}{1 + \frac{s_X^2}{M\bar{x}^2}} \tag{10}$$

where s_x^2 now denotes the unbiased sampling variance of the x_i 's $(i \in \mathcal{M})$.
Bowden (1993) supposes that

d) the total number of resightings for each animal constitutes a set of fixed values $x_1, \ldots, x_N$;

e) the M animals to be marked are selected from the population by means of SRSWOR.

Under assumption *d*) the total number of resightings

$$n = \sum_{i=1}^{N} x_i$$

constitutes a finite population total which is exactly known after the T occasions. On the other hand, the finite population mean, say $\overline{X} = n / N$, and the finite population variance, say

$$S_x^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \overline{X})^2$$

are unknown parameters. Moreover, under assumption *e*), but also adopting other sampling schemes with inclusion probabilities equal to M / N (*e.g* systematic sampling or stratified sampling with proportional allocation), $\bar{x}$ constitutes a design-unbiased estimator of $\overline{X}$. Accordingly, $n / \bar{x}$ constitutes a very natural estimator for *N*.

It is worth noting that, rewriting *n* as $(n - m) + M\bar{x}$, $n / \bar{x}$ is completely equivalent to $\hat{N}_m$. Thus, an alterative interpretation of the naïve estimator $\hat{N}_m$ arises in the framework of finite population sampling, when the marked animals are selected from the population using a sampling scheme in which the sample mean is an unbiased estimator of the population mean.

Obviously, $n / \bar{x}$ constitutes a biased estimator for *N*, since the expectation of $1 / \bar{x}$ does not equal $1 / \overline{X}$. On the basis of the Taylor series expansion of $n / \bar{x}$ around $\overline{X}$ up to the second order, it follows that

$$\mathrm{E}\left(\frac{n}{\bar{x}} \mid x_1, \ldots, x_N \right) \cong N + N \frac{\mathrm{V}(\bar{x})}{\overline{X}^2}$$

which under SRSWOR reduces to

$$\mathrm{E}\left(\frac{n}{\bar{x}} \mid x_1, \ldots, x_N \right) \cong N \left(1 + \frac{S_x^2}{M \overline{X}^2} \right) - \frac{S_x^2}{\overline{X}^2} . \tag{11}$$

Accordingly, $\hat{N}_B$ is a bias-reduced estimator of *N* when marked animals are selected by means of SRSWOR. It must be noticed that in this framework $\mathrm{E}(\cdot \mid x_1, \ldots, x_N)$ and

$\mathrm{V}(\cdot \mid x_1,\ldots,x_N)$ denote expectation and variance performed with respect to SRSWOR of the animals to be marked but conditional on the values of $x_1,\ldots,x_N$, which are assumed to be fixed values.

As to the basic assumption *d*), it seems far from being adequate in the framework of mark-resighting surveys. In order to justify the fact that x_i is assumed to be a fixed characteristic of animal *i*, Bowden & Kufeld (1995, p.843) point out that "*The sighting period and process should be predetermined, fixed and defined*" and that "*The area searched during the sighting period does not need to be the entire study area*". Probably, these two sentences have the purpose of excluding the use of any sampling design, such as encounter designs, which may be adopted to select/resight the animals after marking. Indeed, if the resightings arise from a random search, as when transects or observation points are randomly thrown onto the study area, the x_i's necessarily constitute realizations of random variables and assumption *d*) has no statistical sense. However, even if no sampling plan is adopted and the paths to be travelled for observing animals are purposively selected, it is quite unrealistic to suppose that the number of resightings is a fixed characteristic of animals, such as their body weight, sex or age.

As to assumption *e*), the use of SRSWOR (as any other fixed-size scheme) for selecting the animals to be marked is unrealistic (see section 2.1). Accordingly, the bias reduction performed in (10) is likely to be poorly effective when the animals to be marked are selected from the population by a scheme which greatly differs from SRSWOR. Once again, some misleading sentences are present in literature regarding the selection of the animals to be marked. For example, Bowden & Kufeld (1995, p.842) state that "*We assume each of the N animals had an equal chance of being selected for marking and the selection for marking were made independently of one another*", while (as already pointed out in section 2.1) many encounter schemes can ensure an equal selection probability for each animal, even if the sample mean no longer constitutes an unbiased estimator of the population mean. Indeed, from the well known Horvitz-Thompson theory, this property necessitates schemes with first-order inclusion probabilities invariably equal to M/N. Moreover, without-replacement sampling adopted to select the animals to be marked (which clearly holds since no animal is marked twice) is in contrast with the independence claimed by the two authors. This is also in contrast with White & Shenk (2001, p.337) who state that when using BE "*sampling can be with or without replacement*" and that BE is suitable when "*sampling is performed with replacement*".

By using the Taylor series expansion of $n/\bar{x}$ up to the first leading terms, it follows that

$$\mathrm{V}\left(\frac{n}{\bar{x}} \mid x_1,\ldots,x_N\right) \cong N^2 \frac{\mathrm{V}(\bar{x})}{\bar{X}^2}$$

which under SRSWOR reduces to

$$\mathrm{V}\left(\frac{n}{\bar{x}} \mid x_1,\ldots,x_N\right) \cong \frac{NU}{M} \frac{S_x^2}{\bar{X}^2}. \tag{12}$$

Note that expression (12) is analogous to expression (9) with $\bar{X}$ and S_x^2 instead of μ_x and σ_x^2. Obviously, a trivial estimator of (12) turns out to be

$$\hat{\mathrm{V}}\left(\frac{n}{\bar{x}}\right) = \frac{\hat{N}_B(\hat{N}_B - M)}{M} \frac{s_x^2}{\bar{x}^2} \tag{13}$$

Thus, in accordance with (13), Bowden (1993) suggests the use of

$$\hat{\mathrm{V}}(\hat{N}_B) = \frac{\dfrac{\hat{N}_B\left(\hat{N}_B - M\right)}{M} \dfrac{s_x^2}{\bar{x}^2}}{\left(1 + \dfrac{s_x^2}{M\bar{x}^2}\right)^2}$$

as an estimator of the variance of $\hat{N}_B$.

Finally, since under SRSWOR, $\bar{x}$ converges to normality as N grows along with M (see *e.g.* Thompson 1992, p.28), the author proposes the use of

$$\hat{N}_B \pm t_{1-\alpha/2,M-1} \sqrt{\hat{\mathrm{V}}\left(\hat{N}_B\right)}$$

as the confidence interval for N at the nominal level $1-\alpha$, where $t_{1-\alpha/2,M-1}$ is the $1-\alpha/2$ quantile of a *t*-distribution with $(M-1)$ degrees of freedom.

Interestingly, assumption *d*) may be relaxed with no detrimental effects on estimation and it may be interpreted more realistically as if the estimation were performed conditional on the resulting values of the X_i's. Indeed, if conditional to $x_1,\ldots,x_N$, $\bar{x}$ is an approximately unbiased estimator of $\bar{X}$ (as ensured by assumption *e*), then from the well known properties on conditional expectation and variance and from (11) and (12), it follows that

$$\mathrm{E}\left(\hat{N}_B\right) = \mathrm{E}_{x_1,\ldots,x_N}\left\{\mathrm{E}\left(\hat{N}_B \mid x_1,\ldots,x_N\right)\right\} \cong \mathrm{E}_{x_1,\ldots,x_N}\{N\} = N \tag{14}$$

and

$$\mathrm{V}\left(\hat{N}_B\right)=\mathrm{E}_{x_1,\dots,x_N}\left\{\mathrm{V}\left(\hat{N}_B \mid x_1,\dots,x_N\right)\right\}+\mathrm{V}_{x_1,\dots,x_N}\left\{\mathrm{E}\left(\hat{N}_B \mid x_1,\dots,x_N\right)\right\}\cong$$

$$\cong \mathrm{E}_{x_1,\dots,x_N}\left\{\frac{\frac{NU}{M}\frac{S_x^2}{\overline{X}^2}}{\left(1+\frac{S_x^2}{M\overline{X}^2}\right)^2}\right\} \tag{15}$$

where now $\mathrm{E}(\cdot)$ and $\mathrm{V}(\cdot)$ denote expectation and variance performed with respect to both the SRSWOR of the animals to be marked and the random variables $X_1,\dots,X_N$.
Practically speaking, relations (14) and (15) ensure that $\hat{N}_B$ and $\hat{\mathrm{V}}\left(\hat{N}_B\right)$, being conditionally unbiased, turn out to be unbiased estimators even if $x_1,\dots,x_N$ constitute realizations of random variables rather than fixed values. Accordingly, if the marks are adequately apportioned among the animals as should be expected under SRSWOR, a good level of robustness should be expected for BE and the related confidence intervals.

3. Simulation results

All the assumptions underlying the NOREMARK estimators turn out to be highly unrealistic. Assumptions *a*) and *e*) refer to a sampling scheme which cannot be performed in animal populations, assumptions *b*) and *c*) entail some forms of independence in time and space which never hold in real situations and assumption *d*) claims an unreliable absence of variability in the number of resightings for each animal. Accordingly, extended simulation studies are necessary to check the robustness of the NOREMARK estimators in more realistic situations in which all the assumptions are supposed to be jointly violated.
The first simulation study was performed by Neal *et al.* (1993) to investigate the robustness of JHE. The authors supposed populations of size $N=50,100,200,500$, a number of occasions $T=5(5)20$, mark proportion $\beta=0.1(0.2)0.5$ and a resighting probability $\pi=0.1(0.2)0.7$. In the first part of the study, for each combination of N,T,β and π, the number of marked animals *M* was generated from a binomial distribution with parameters *N* and β. Then, for each resulting population, the empirical distribution of JHE was determined on the basis of 1,000 independent replicates. At each replication, m_t and n_t-m_t were independently generated for any $t=1,\dots,T$ from two binomial distributions with parameters *M* and π and *N-M* and π, respectively. In the second part of the study, for

the some combinations of N, T and β, the population was randomly structured into groups by generating group sizes from an empirical distribution observed in a field study on mountain sheep at Trickle Mountain (Colorado). Subsequently, the number of marked animals in a group of size g was generated from a binomial distribution with parameters g and β. Then, for each resulting population the empirical distribution of JHE was determined on the basis of 1000 replications. At each replication, each animal in the population was randomly and independently detected at any occasion $t = 1,\ldots,T$ using four different schemes: 1) with probability $\pi = 0.1(0.2)0.7$; 2) with probability increasing with group size; 3) with probability generated from a beta distribution with parameters $a = b = 3$; 4) with probability varying over the occasions from 0.05 to 0.95. A quite similar simulation study was also performed by White (1993) with the purpose of comparing the performance of MME vs JHE. The simulation study repeated exactly the procedures adopted by Neal *et al.* (1993) with the exception that in the second part of the study no group structure was considered and each animal was detected only using scheme 3). In both these studies, the first set of simulations was devoted to checking the performance of JHE in the case in which *all assumptions* [are] *met*, to be compared with the results obtained in the second set where the assumptions underlying JHE were violated. However, it is apparent from the considerations of section 2.1 that the JHE assumptions were always violated in these studies, since the sampling schemes adopted to select/resight the animals always differed from SRSWOR. Moreover, in both studies, it made no sense to adopt sampling scheme 3) to check the robustness of MME under the "*heterogeneity of individual sighting probabilities*". Indeed, on the basis of the considerations in section 2.2, it is apparent that no violation of assumption *c*) occurred under scheme 3). Apart from these inaccuracies, the results of the simulations mainly delineated the possibility of significant if not serious bias together with sometimes poor coverage of the resulting confidence intervals. These turned out to be more marked for the Monte Carlo intervals adopted in MME than for the Hudson intervals adopted in JHE (White 1993, p.96). Finally, a simulation study performed by Bowden & Kufeld (1995) to check the statistical properties of BE should be noted for completeness. However, since assumptions *d*) and *e*) were both met when generating mark-resighting data, the study is of no help in checking the robustness of the method under the failure of assumptions.

In accordance with the previous considerations, a more rigorous simulation study in which all the assumption, from *a*) to *e*) were jointly violated is necessary in order to give more insight into the robustness of mark-resighting estimators. The study was planned to take into consideration quite realistic sampling schemes, several group structures in order to generate several levels of dependence among animal detections, mark effort and mark apportionment among groups in order to generate more-or-less representative marking of animals within populations, sequential movements of animals in order to generate various levels of dependence among resighting occasions and, finally, animal detection probabilities in order to consider more-or-less favourable field conditions. To this purpose, a population of $N = 100$ units was supposed to be spread onto a square region of size 1 and to be surveyed over the course of $T = 10$ resighting occasions. Two group structures were supposed: the first structure involved 30 groups of size 2, 10 groups of size 1 and 10 groups of size 3 for a total of 50 groups with average size $\overline{g} = 2$; the second structure involved 4 groups of size 10, 2 groups of size 9, 2 groups of size 11, 1 group of size 8 and 1 group of size 12 for a total of 10 groups with average size $\overline{g} = 10$. For any population structure an even or aggregated allocation of marks among groups was supposed. More precisely, in the first group structure, the even distribution involved 1 marked animal in 10 groups of size 2 when $M = 10$ and 1 marked animal in all the 30 groups of size 2 when $M = 30$ while, in the second group structure, the even distribution of marks involved 1 or 3 marked animals per group for a total of $M = 10\,(\beta = 0.1)$ or $M = 30\,(\beta = 0.3)$ marked animals, respectively. On the other hand, in the first group structure, the aggregated distribution involved 2 marks in 5 groups of size 2 or 2 marks in 15 groups of size 2 when $M = 10$ or $M = 30$ respectively, while in the second group, the aggregated distribution of marks involved 5 marks in 2 groups of size 10 when $M = 10$ and 5 marks in all the 4 groups of size 10, 5 marks in 1 group of size 9 and 5 marks in a group of size 11 when $M = 30$.

Then, in order to consider animal movements, at each replication of the sampling scheme, the groups were randomly located on the square region at $t = 1$, while in the 9 subsequent occasions, group movements were simulated by randomly changing the spatial location of each group inside a circle of radius $\rho = 0.1,\ 0.5$ centred at the previous location. Finally, at any occasion, a line transect census was simulated by using 4 fixed transects. The transects were constituted by 4 straight lines of length 1 starting on the lower side of the square region at the point 0.125, 0.375, 0.625 and 0.875, in such a way that the perpendicular distance of any group to the nearest transect was bounded by 1/8. As to the resighting process, a group of size g at distance y from the transect was spotted with probability $h(y) = 1 - (8y)^{\theta^{-g}-1}$ where θ was determined in such a way that the final probability of spotting a group turned out to be $\pi = 0.3, 0.5$. More precisely, since owing to the random placements of group, the probability of spotting a group of size g turned out to be

$$\pi(g) = 1 - \int_0^{1/8} (8y)^{\theta^{-g}-1} dy = 1 - \theta^g$$

then if $g_1,\dots,g_K$ denotes the K distinct sizes of groups in the population and if $f_1,\dots,f_K$ denotes their relative frequencies, then, by elementary probabilistic considerations, θ is obtained as the numerical solution of the equation

$$1-\sum_{k=1}^{K}\theta^{g_k} f_k = \pi\,.$$

Finally, when a group was spotted, all the animals in the group were presumed to be spotted, in such a way that π also represented the probability of spotting an animal.

For each of the 32 scenarios resulting from combining the values of $\overline{g},\beta,\rho,\pi$ and the two types of mark distribution, 10,000 replications were performed. For each replication, the estimates $\hat{N}_{JH}$, $\tilde{N}_{MM}$, $\hat{N}_{MM}$ and $\hat{N}_B$ were computed. The values of $\tilde{N}_{MM}$ were analytically computed as the mode of the waiting time distribution (4), avoiding any Monte Carlo procedures. Accordingly, the confidence intervals were constructed via the appropriate quantiles of the same analytical distribution. For each estimator, the empirical values of the relative bias (RB= bias divided by *N*), relative mean error (RME=square root of mean squared error divided by *N*), coverage (C=probability that the 0.95 confidence interval contains *N*) and relative mean interval length (RMIL= expected 0.95 interval length divided by *N*) were computed on the basis of the resulting distributions. The simulation results are reported in Tables 3 to 6.

Table 3. Empirical values of the relative bias (RB), relative mean error (RME), interval coverage (C) and relative mean interval length (RMIL) for the joint hypergeometric maximum likelihood estimator $\hat{N}_{JH}$.

Simulation parameters				RB	RME	C (RMIL)	RB	RME	C (RMIL)
Mark distribution					EVEN			AGGREGATED	
$\overline{g} = 2$	$\beta = 0.1$	$\pi = 0.3$	$\rho = 0.05$	.128	.308	.816 (.780)	.223	.609	.637 (.959)
			$\rho = 0.5$	.095	.193	.941 (.719)	.131	.312	.808 (.775)
		$\pi = 0.5$	$\rho = 0.05$	.077	.199	.770 (.459)	.123	.353	.585 (.513)
			$\rho = 0.5$	.057	.118	.933 (.434)	.077	.182	.79 (.453)
	$\beta = 0.3$	$\pi = 0.3$	$\rho = 0.05$	.078	.134	.796 (.347)	.103	.222	.606 (.366)
			$\rho = 0.5$	.074	.103	.898 (.341)	.091	.156	.736 (.352)
		$\pi = 0.5$	$\rho = 0.05$	.051	.088	.771 (.216)	.070	.153	.551 (.25)
			$\rho = 0.5$	.047	.063	.912 (.213)	.058	.096	.729 (.217)
$\overline{g} = 10$	$\beta = 0.1$	$\pi = 0.3$	$\rho = 0.05$	.005	.025	1.000 (.599)	.432	.186	.479 (1.567)
			$\rho = 0.5$	.005	.018	1.000 (.581)	.184	.602	.642 (.903)
		$\pi = 0.5$	$\rho = 0.05$	.003	.018	1.000 (.367)	.276	.852	.476 (.743)
			$\rho = 0.5$	.003	.011	1.000 (.355)	.088	.287	.640 (.444)
	$\beta = 0.3$	$\pi = 0.3$	$\rho = 0.05$	.005	.025	1.000 (.295)	.060	.251	.556 (.336)
			$\rho = 0.5$	.005	.018	1.000 (.287)	.039	.149	.739 (.308)
		$\pi = 0.5$	$\rho = 0.05$	.003	.018	1.000 (.183)	.039	.157	.536 (.200)
			$\rho = 0.5$	.004	.011	1.000 (.178)	.027	.085	.753 (.188)

Table 4. Empirical values of the relative bias (RB), relative mean error (RME), interval coverage (C) and relative mean interval length (RMIL) for the Minta-Mangel estimator $\widetilde{N}_{MM}$.

Simulation parameters				RB	RME	C (RMIL)	RB	RME	C (RMIL)
Mark distribution					EVEN			AGGREGATED	
$\overline{g} = 2$	$\beta = 0.1$	$\pi = 0.3$	$\rho = 0.5$	.127	.308	.459 (.298)	.214	.605	.309(.297)
			$\rho = 0.5$	.095	.194	.453 (.200)	.123	.309	.321(.196)
		$\pi = 0.5$	$\rho = 0.05$	.079	.201	.478 (.219)	.119	.354	.323(.216)
			$\rho = 0.5$	.061	.121	.471(.132)	.075	.182	.324(.128)
	$\beta = 0.3$	$\pi = 0.3$	$\rho = 0.05$	.079	.135	.705(.264)	.098	.221	.520(.268)
			$\rho = 0.5$	.076	.104	.627(.183)	.086	.154	.485(.328)
		$\pi = 0.5$	$\rho = 0.05$	.055	.092	.744(.195)	.068	.154	.526(.199)
			$\rho = 0.5$	.052	.067	.655(.122)	.057	.097	.515(.121)
$\overline{g} = 10$	$\beta = 0.1$	$\pi = 0.3$	$\rho = 0.05$	.007	.043	1.000 (.287)	.424	1.187	.165(.228)
			$\rho = 0.5$	.009	.034	1.000 (.186)	.174	.602	.176(.147)
		$\pi = 0.5$	$\rho = 0.05$	.008	.027	1.000 (.214)	.275	.858	.223(.178)
			$\rho = 0.5$	.010	.022	1.000 (.134)	.085	.291	.199(.104)
	$\beta = 0.3$	$\pi = 0.3$	$\rho = 0.05$	.007	.037	1000(.255)	.049	.254	.462(.253)
			$\rho = 0.5$	.009	.031	1.000(.178)	.030	.151	.493 (.174)
		$\pi = 0.5$	$\rho = 0.05$	.008	.027	1.000(.192)	.032	.167	.502(.194)
			$\rho = 0.5$	.010	.022	.415(.214)	.022	.090	.546(.121)

Table 5. Empirical values of the relative bias (RB), relative mean error (RME), interval coverage (C) and relative mean interval length (RMIL) for the Minta-Mangel pseudo maximum likelihood estimator $\hat{N}_{MM}$.

Simulation parameters				RB	RME	C (RMIL)	RB	RME	C (RMIL)
Mark distribution					EVEN			AGGREGATED	
$\bar{g} = 2$	$\beta = 0.1$	$\pi = 0.3$	$\rho = 0.05$	.128	.309	.942 (.979)	.218	.608	.779(1.042)
			$\rho = 0.5$	.093	.193	.935 (.631)	.124	.310	.750(.622)
		$\pi = 0.5$	$\rho = 0.05$	.078	.201	.944 (.693)	.121	.354	.769(.699)
			$\rho = 0.5$	.058	.119	.927(.398)	.073	.181	.737(.380)
	$\beta = 0.3$	$\pi = 0.3$	$\rho = 0.05$	.079	.135	.972(.480)	.104	.217	.852(.493)
			$\rho = 0.5$	.074	.103	.939(.329)	.086	.152	.779 (.328)
		$\pi = 0.5$	$\rho = 0.05$	.073	.093	.983(.355)	.104	.149	.875 (.365)
			$\rho = 0.5$	.065	.073	.905(.213)	.083	.100	.731(.213)
$\bar{g} = 10$	$\beta = 0.1$	$\pi = 0.3$	$\rho = 0.05$	.008	.027	1.000 (.825)	.433	1.189	.420(.933)
			$\rho = 0.5$	.009	.020	1.000 (.199)	.178	.603	.417(.483)
		$\pi = 0.5$	$\rho = 0.05$	.008	.023	1.000 (.614)	.293	.858	.549(.632)
			$\rho = 0.5$	.009	.017	.999 (.370)	.095	.286	.481(.295)
	$\beta = 0.3$	$\pi = 0.3$	$\rho = 0.05$	.022	.044	.992(.426)	.098	.239	.833(.456)
			$\rho = 0.5$	.024	.042	.978(.289)	.065	.141	.784 (.291)
		$\pi = 0.5$	$\rho = 0.05$	.108	.127	.766(.346)	.147	.185	.0639(.357)
			$\rho = 0.5$	.122	.136	.415(.214)	.136	.149	.326(.211)

Table 6. Empirical values of the relative bias (RB), relative mean error (RME), interval coverage (C) and relative mean interval length (RMIL) for the Bowden estimator $\hat{N}_B$.

Simulation parameters				RB	RME	C (RMIL)	RB	RME	C (RMIL)
Mark distribution					EVEN			AGGREGATED	
$\bar{g} = 2$	$\beta = 0.1$	$\pi = 0.3$	$\rho = 0.05$	.068	.256	.939 (.904)	.152	.530	.775 (.951)
			$\rho = 0.5$	.067	.173	.946 (.626)	.098	.283	.769 (.613)
		$\pi = 0.5$	$\rho = 0.05$	.046	.175	.945 (.679)	.086	.319	.779 (.678)
			$\rho = 0.5$	.047	.112	.943 (.409)	.062	.172	.760 (.390)
	$\beta = 0.3$	$\pi = 0.3$	$\rho = 0.05$	.065	.124	.976 (.472)	.084	.209	.822 (.481)
			$\rho = 0.5$	.068	.098	.952 (.329)	.078	.148	.782 (.328)
		$\pi = 0.5$	$\rho = 0.05$	.046	.085	.984 (.347)	.059	.147	.820 (.352)
			$\rho = 0.5$	.047	.063	.943 (.212)	.051	.093	.771 (.209)
$\bar{g} = 10$	$\beta = 0.1$	$\pi = 0.3$	$\rho = 0.05$	.038	.049	1.000 (.773)	.369	1.074	.409 (.852)
			$\rho = 0.5$	.013	.022	1.000 (.558)	.154	.560	.424 (.471)
		$\pi = 0.5$	$\rho = 0.05$	.019	.028	1.000 (.605)	.246	.801	.500 (.617)
			$\rho = 0.5$	.001	.010	1.000 (.380)	.076	.280	.447 (.298)
	$\beta = 0.3$	$\pi = 0.3$	$\rho = 0.05$	.002	.024	1.000 (.413)	.040	.243	.685 (.428)
			$\rho = 0.5$	.004	.017	1.000 (.285)	.025	.147	.684 (.281)
		$\pi = 0.5$	$\rho = 0.05$	.001	.018	1.000 (.312)	.026	.160	.714 (.319)
			$\rho = 0.5$	.006	.012	1.000 (.191)	.019	.086	.724 (.188)

4. Discussion with a case study

The analysis of the tables motivates the use of the $\hat{N}_B$ as the method giving rise to the best accuracy results together with the best coverage of the confidence intervals. As theoretically asserted in section 2.3, under an even distribution of marks among groups, the bias level achieved by $\hat{N}_B$ turns out to be very satisfactory, being always less than 7%. Moreover, the interval coverage of the method is very near to or greater than 95% in all the situations. On the other hand, an uneven distribution of marks among groups heavily deteriorates the performance of BE, which however remains the best when compared with the performance of the other estimators.

The presence of large groups exacerbated the accuracy losses even if the main reason for failure remains by far the uneven distribution of marks. The presence of dependence among occasions also has a (less marked) detrimental effect on the accuracy of the estimators. Under the worst conditions, such as the uneven distribution of marks, $\beta = 0.1$, $\pi = 0.3$ and a high level of dependence among occasions $(\rho = 0.05)$, the bias of BE increases from 15% with a relative mean error of 53% when $\bar{g} = 2$ to 37% with a relative mean error of 100% when $\bar{g} = 10$. The aggregated distribution of marks heavily impacts also on the coverage of the confidence intervals constructed around BE, which even if no worse than the coverage of the other methods, falls from 77% to 41% under the above-mentioned less-favourable conditions.

In accordance with these results, there does not seem to be any reason to adopt the most complex NOREMARK procedures, *i.e.* JHE or MME. Indeed, as to JHE, the Hudson method gives rise to confidence intervals with coverage always lower than those provided by BE. Moreover, as to MME, the confidence intervals constructed around $\tilde{N}_{MM}$ by using the quantiles of the waiting time distribution turn out to be completely unreliable with a coverage which turns out to be much smaller than the nominal level even when the marks are evenly distributed among groups (these results confirm the theoretical consideration made in section 2.2). Finally, even if the use of $\hat{N}_{MM}$ with the related confidence intervals greatly improves the coverage over the Minta-Mangel procedure, and even if $\hat{N}_{MM}$ is in some cases comparable with $\hat{N}_B$, its use involves greater computational complexities without offering substantial gains in precision.

Accordingly, providing that the marks are evenly distributed among groups, the Bowden criterion constitutes a valid procedure for estimating population size from mark-resighting data, accomplishing robustness and computational simplicity. On the other hand, if the marks are unevenly distributed, no mark-resighting procedure seems to be reliable.

These findings may be of some help to interpret the less-than-obvious results obtained in a case study in which NOREMARK methods were adopted to estimate the size of a population of chamois *Rupicapra rupicapra* (Linneaus, 1758) females on a study area (12°.00' E – 46°.30' N) located in the Paneveggio - Pale di San Martino Natural Park (Trentino, eastern Italian Alps). The area size was of about 7 km^2, with altitudes ranging from 1600 to 2200 m a.s.l. and vegetation mainly consisting of alpine grassland, coniferous (larch and spruce) forest and shrubs (knee pine, alder, juniper, and rhododendron trees). Chamois detection was almost complete in alpine grassland and rocky slopes (36,4% of total surface), heterogeneous but generally poor in larch forest (34,5%) and very low or null in other forest and shrubs habitats.

Animal radio-tracking was conducted since the end of the 90's, funded by "Paneveggio-Pale di San Martino" Natural Park. As a consequence, from $M = 20$ to $M = 14$ radio-tagged females were available from 2000 to 2003, which were yearly used to estimate the female population size by means of mark-resight surveys. Marks were quite evenly distributed between groups, in the sense that the greatest group, referred to by the researchers as group 3, and the second one, referred to as group 1, contained the greatest number of marked animals (from 11 to 7 in group 3, from 4 to 4 in group 1) while few or no marked animals were contained in the smaller groups, referred to as groups 2, 4 and 5 (from 3 to 1 in group 2, 1 in group 4 and from 0 to 1 in group 5).

In the summer of each year, in the month of July or August, counts were performed over 4 or 5 consecutive days. The study area was partitioned into 4 spatial units which were searched simultaneously. The surveys were conducted in all weather conditions and observations were made from purposively-selected fixed points and transects in the early morning (06.00-10.00 AM) by at least 2 observers for each unit. Observers were equipped with binoculars (7x42 or 10x40) and spotting scopes (20-60x75).

Table 7 reports the values of $\hat{N}_{JH}$, $\tilde{N}_{MM}$, $\hat{N}_{MM}$ and $\hat{N}_B$ as well as the resulting 0.95 confidence intervals of the population size. While a drastic and highly significant decrease of female abundance (from about 200 to 80 individuals) is deduced using JHE and MME from 2000 to 2001, with the lower ends of the 2000 confidence intervals much greater than the upper ends of the 2001 intervals (more than 80 individuals in both cases), a similar but less significant reduction is deduced using BE, with the lower end of the 2000 confidence interval very close to the upper end of the 2001 interval. Results akin to those obtained using BE are also obtained by means of the pseudo-maximum likelihood estimates. The presence of such clashing results achieved using NOREMARK could be justified only on the basis of the unreliability of JHE and MME, which has been theoretically and empirically emphasized throughout this work. On the other hand, owing to the even distribution of marks performed among the five groups partitioning the population, the results based on BE should be the most reliable. Accordingly, even if a clear reduction of the female population from 2000 to 2001 should be deduced (as can also be asserted from the 2000-01 high mortality rate caused by dramatic winter climatic conditions) such a reduction might not be as substantial as that which arises from two of the three estimation procedures adopted in NOREMARK.

Table 7. Mark-resighting size estimates and 0.95 confidence intervals (in brackets) for the female community of chamois on Paneveggio - Pale di San Martino Natural Park in the years 2000-03.

year	$\hat{N}_{JH}$	$\tilde{N}_{MM}$	$\hat{N}_{MM}$	$\hat{N}_{B}$
2000	202 (156-279)	205 (176-240)	206 (104.4-307.6)	193.4 (102.6-284.2)
2001	85 (76-102)	83 (76-92)	83 (65.9-100.1)	82.6 (65.3-100.0)
2002	103 (87-129)	95(86-105)	95 (72.5-117.5)	93.75 (71.3-116.2)
2003	95 (79-123)	96 (85-108)	96 (67.2-124.8)	93.8 (65.5-122.2)

Acknowledgements

The authors thank Professor Sandro Lovari who stimulated this work, Professor Luca Pratelli for his helpful suggestions in the theoretical aspects of the work, and the park wardens of Paneveggio - Pale di San Martino Natural Park as well as colleagues and students who contributed to capturing animals and collecting data. The field study on chamois was funded by “Paneveggio-Pale di San Martino” Natural Park.

References

Arnason A.N., Schwarz C.J., Gerrard J.M. (1991) Estimating closed populations size and number of marked data from sighting data. *Journal of Wildlife Management*, **55**, 716-730.

Bartmann R.M., White G.C., Carpenter L.H., Garrott, R.A. (1987) Aerial mark-recapture estimates of confined mule deer in pinyon-juniper woodland. *Journal of Wildlife Management*, **51**, 41-46.

Bowden D.C. (1993) *A simple technique for estimating population size*. Department of Statistics, Colorado State University, Fort Collins.

Bowden D.C., Kufeld R.C. (1995) Generalized mark-sight population size estimation applied to Colorado moose. *Journal of Wildlife Management*, **59**, 840-851.

Burnham K.P., Overton W.S. (1978) Estimation of the size of a closed population when captures very among animals. *Biometrika*, **65**, 625-633.

Furlow R.C., Haderlie M., Van den Berge R. (1981) Estimating a bighorn sheep population by mark-recapture. *Desert Bighorn Council Transactions*, **25**, 31-33.

Gong G., Samaniego F.J. (1981) Pseudo maximum likelihood estimation: theory and applications. *The Annals of Statistics*, **8**, 861-869.

Hein E.W., Andelt W.F. (1995) Estimating coyote density from mark-resight surveys. *Journal of Wildlife Management*, **59**, 164-169.

Hudson D.J. (1971) Interval estimation from the likelihood function. *Journal of the Royal Statistical Society*, **B 33**, 256-262.

Le Cren E.D. (1965) A note on the history of mark-recapture population estimates. *Journal of Animal Ecology*, **34**, 453-454.

Leslie D.M. Jr., Douglas C.L. (1979) Desert bighorn sheep of the River Mountains, Nevada. *Wildlife Monograph* **66**, 1-56.

Leslie D.M. Jr., Douglas C.L. (1986) Modeling demographics of bighorn sheep: current abilities and missing links. *North American Wildlife and Natural Resources Conference Transactions*, **51**, 62-73.

Miller S.D., Becker E.F., Ballard W.H.(1987) Black and brown bear density estimates using modified capture-recapture techniques in Alaska. *International Conference on Bear Reasearch and Management*, **7**, 23-35.

Minta S., Mangel M. (1989) A simple population estimate based on simulation for capture-recapture and capture-resight data. *Ecology,* **70**, 1738-1751.

Neal A.K., White G.C., Gill R.B., Reed D.F., Olterman J.H. (1993) Evaluation of mark-resight model assumptions for estimating mountain sheep numbers. *Journal of Wildlife Management*, **57**, 436-450.

Otis D.L., Burnham K.P., White C.W., Anderson D.R. (1978) *Statistical inference from capture data on closed animal populations*. Wildlife Monographs n.62, The Wildlife Society, USA.

Ries E. H., Hiby L.R., Reijnders P.J.H. (1998) Maximum likelihood population size estimation of harbour seals in the Dutch Wadden Sea based on capture-recapture experiment. *Journal of Applied Ecology*, **35**, 332-339.

Rice W. R., Harder J.D. (1977) Application of multiple aerial sampling to a mark-recapture census of white-tailed deer. *Journal of Wildlife Management*, **41**, 197-206.

Seber G.A. (1982) *The estimation of animal abundance and related parameters*. Arnold Edward, London.

Thompson S.K. (1992) *Sampling*. Wiley, New York.

White G.C. (1993) Evaluation of radio tagging marking and sighting estimators of population size using Monte Carlo simulations. *Marked Individuals in the Study of Bird Populations* (eds J.D. Lebreton and P.M. North), pp. 91-103. Birkhäuser Verlag, Basel.

White G.C. (1996) NOREMARK: Population estimation from mark-resighting surveys. *Wildlife Society Bullettin*, **24**, 50-52.

White G.C., Garrot R.A. (1990) *Analysis of wildlife radio-tracking data*. Academic Press, San Diego (CA).

White G.C., Shenk, T.M. (2001) Population estimation with radio-marked animals. *Designs and analysis of wildlife radiotelemetry studies*, (eds. J.J. Millspaugh and J.M. Marzluff), Academic Press, San Diego (CA).