Brown bear population trends from demographic and monitoring-based estimators

Jedediah F. Brodie¹,³ and Michael L. Gibeau²

¹Division of Biological Sciences, University of Montana, Missoula, MT 59812, USA
²Parks Canada, Banff National Park, Box 213, Lake Louise AB T0L 1E0, Canada

Abstract: A primary goal of monitoring wildlife populations is the estimation of population growth rate, \( \lambda \). Two common methods by which biologists estimate \( \lambda \) are demographic studies of marked individuals, which tend to be expensive and labor-intensive, and estimators derived from time series of population indices. We compare grizzly bear (Ursus arctos) population growth rates in the Banff ecosystem (Alberta, Canada) from a published demographic study to estimates from concurrent monitoring of an index of population size, the number of females with cubs-of-the-year \( F_{\text{cub}} \). We estimated population trends by transforming the index into 2 population estimators (bias-corrected Chao and summation), and used each to estimate \( \lambda \). The 95\% confidence intervals of \( \lambda \) from the 2 monitoring-based estimators overlapped the point estimate of the demographic study. Precision of the bias-corrected Chao estimator was very low (95\% CI of \( \lambda = 0.572–1.679 \)); its application to the time-series used here is essentially fruitless. Precision of the summation estimator (95\% CI of \( \lambda = 0.847–1.137 \)) and the demographic study (0.99–1.09) were higher, but the CI of the former at least could be artificially narrow. Because all estimates were close to 1.00, the long-term fate of this population may depend critically on subtle changes in growth rate and on environmental stochasticity. Given that long-term demographic studies are not feasible in this system, population monitoring may be a worthwhile way to assess population dynamics. However, given the low power of many monitoring techniques to detect trends and the low precision of the \( F_{\text{cub}} \) estimators in particular, long time-series and explicit measures to remove sampling variance should be employed to increase trend estimate precision.

Key words: Banff National Park, carnivores, demography, estimators, \( F_{\text{cub}} \), females with cubs, indices, Kananaskis country, monitoring, population growth rate, Ursidae, Ursus arctos


Research has consistently indicated that 3 dominant factors drive extinction risk: population size, average population growth rates, and temporal variation in population growth rates (Fagan et al. 2001, Inchausti and Halley 2003, Reed and Hobbs 2004). The minimum criterion for a population to persist is a geometric mean annual population growth rate \( \geq 1.0 \), meaning that births exceed or numerically balance deaths. However, this criterion alone is not sufficient for population persistence; small or highly variable populations can go extinct despite relatively high population growth rates (Dennis et al. 1991, Mangel and Tier 1994). Therefore, population size and variance in growth rate are both key parameters for understanding population status.

The true abundance and growth rate of vertebrate populations can seldom, if ever, be ascertained by direct enumeration. Instead, wildlife biologists have several options. Individuals in the population can be marked and their fates followed over time, allowing demographic rates (e.g., age-specific survivorship or fecundity) to be calculated (Leslie 1945, Wakkinen and Kasworm 2004, Kovach et al. 2006). Dominant eigenvalues of the resulting transition matrices then provide a measure of \( \lambda \) (Caswell 2000). If large sample sizes are available over relatively long periods, such studies can estimate trends precisely and accurately. However, because they are labor intensive and expensive, demographic studies of marked individuals are seldom conducted over
enough years to estimate variance in $\lambda$ over time (Fieberg and Ellner 2000). In addition, small sample sizes obtained in many vertebrate population studies can seriously affect performance of demographic estimators (McKelvey and Pearson 2001).

Population size, $\lambda$, and variance in $\lambda$ can also be estimated from long-term monitoring of population size estimates. These repeated estimates can be derived from, for example, mark–recapture studies (Pradel 1996, Kendall et al. 1997) or index-based density estimates (Knight et al. 1995, Keating et al. 2002). Indices are measurable quantities that are assumed to be proportional to actual population size. If the functional relationship between the index and true population size is known, the index can be turned into an estimator of the actual population size. An advantage of index-derived estimators is that they are typically easier to measure than actual population sizes and, therefore, make monitoring over large scales in space and time feasible. A disadvantage is that the relationship between index values and true population size is seldom known. Use of population parameters derived from indices can be especially problematic if underlying assumptions are not met (Thompson 2003).

Problems resulting from extrapolation from small samples are particularly acute in studies of large, free-ranging carnivores. These species usually occur at very low densities, are difficult to locate and count, and yet are often of extremely high conservation or management concern (Soule and Terborgh 1999). In principle, demographic studies of marked (e.g., radiocollared) individuals may provide the best estimates of population growth rates, yet these types of studies face several limitations, described above. Sampling techniques using genetic tagging can provide population estimates non-invasively (Bellemain et al. 2005, Solberg et al. 2006), but remain expensive and labor-intensive and are also, therefore, usually short in duration.

Demographic and monitoring-based estimators each have advantages and disadvantages, and we make no claim here that either method is necessarily preferable. Although many studies have assessed the efficacies of index-based estimators (e.g., Don 1984, Hallett et al. 1991, Calvert and Robertson 2002), few have compared index-based estimators with demographic estimators to assess similarity in calculated growth rates. This is surprising because both of these methods for monitoring populations are common. It would be useful for biologists using one monitoring method to have a sense of how their results would differ if using an alternative method. Here, we compare grizzly bear ($Ursus arctos$) population growth rates calculated from monitoring-based estimators with that from a previously published demographic study in Banff National Park and Kananaskis Country, Alberta, Canada (hereafter, “Banff–Kananaskis”). We generated estimates of $\lambda$ using yearly counts of unduplicated female bears with cubs-of-the-year (hereafter $F_{\text{cub}}$) after conversion by 2 common formulae to estimators of population size (Knight et al. 1995, Keating et al. 2002). We compared growth rates generated from these analyses with estimates derived from a concurrent demographic study by Herrero (2005; also see Garshelis et al. 2005a) to ascertain: (1) how well population growth rates generated from the index-based estimators matched those of the demographic estimator, and (2) which of the $F_{\text{cub}}$ estimators generated growth rate estimates closest to that of the demographic estimator.

**Materials and methods**

**Study area**

The Bow River watershed of southwestern Alberta constituted the core of the study area. This area is 11,400 km$^2$ of mountainous terrain 50–180 km west of Calgary and was the focus of an intensive grizzly bear research program during 1994–2004 (Herrero 2005). The area includes roughly 50% of Banff National Park (BNP) and all adjacent Alberta Provincial land known as Kananaskis Country. Neither jurisdiction permitted grizzly bear hunting, although bears were exposed to hunting outside the Bow River Watershed. Differing agencies oversee preservation, industrial tourism, recreation, forestry, oil and gas extraction, mining, and stock grazing. Native councils, towns and municipalities, commercial developers, and residential owners all manage lands.

**Field methods**

Reproductive status of female grizzly bears was determined as part of a larger ongoing research effort (Herrero 2005) that maintained 10–15 radiocollared females out of a low density population of approximately 100 bears. Observations of $F_{\text{cub}}$ were by research staff as well as sighting records from provincial and federal agencies. Both agencies employ a bear monitoring system where the public is encouraged to report grizzly bear sightings. We
did not include in the $F_{\text{cub}}$ calculations sightings where the females were located for demographic monitoring; the $F_{\text{cub}}$ sighting records only included bears (collared or not) seen without specifically looking for them or by radiotracking by biologists or the general public. Agency and research staff followed up on all reports of family groups for data verification.

We followed procedures and criteria similar to Knight et al. (1995; also see Interagency Conservation Strategy Team 2003) to determine whether sequential sightings belonged to the same family group or different groups. The combination of range size, physical barriers, and population density made distinguishing individual females with litters relatively straightforward. The mean standard diameter for annual ranges of radiocollared females with cubs of the year was 13 km ($n = 27$, SD = 5.2 km, Gibeau unpublished data). Given the similarity in range size to Knight et al. (1995), we simply adopted their criteria of 30 km separation in judging whether 2 females with litters of the same size were distinct. Based on evidence from radiotelemetry (Herrero 2005) and genetics (Proctor 2005), we considered the Trans Canada Highway, which bisects the area, a barrier to female bears with cubs. In 6 of the 12 year monitoring period, there were overlapping litters of the same size. In all 6 cases this involved only 2 family groups.

We made decisions whether family groups were distinct following the rule set outlined by Knight et al. (1995:246), “Once a female with a specific number of cubs was sighted in an area, no other female with the same number of cubs in that same area was regarded as distinct unless 2 family groups were seen by the same observer on the same day, or by 2 observers at different locations but similar times, or 1 or both of the females were radio-marked. Because of possible cub mortality, no female with fewer cubs was considered distinct in that area unless she was seen on the same day as the first female or unless both were radio-marked.” Cubs were classified from their size and, if known, the reproductive status of the female from the previous year. The maximum number of cubs observed was considered the litter size, although cubs lost very early in the season would not have been recorded.

**Population estimators**

Raw counts of $F_{\text{cub}}$ alone are not good metrics of population size in a given year because a different fraction of bears breed in each year (Eberhardt and Knight 1996) and because individual $F_{\text{cub}}$ vary in their detectability (Mattson 1997). A simple method of using $F_{\text{cub}}$ to establish a lower bound for population size is to sum the observations across the mean interbirth interval; this is a commonly used estimator in the continental US as part of the recovery plan under the Endangered Species Act (16 US Code 1531–1544; Knight et al. 1995), where $F_{\text{cub}}$ are summed over 3 years, the mean interbirth interval (in Yellowstone National Park, USA; see Knight and Eberhardt 1985, Eberhardt and Knight 1996). However, Keating et al. (2002) pointed out that this method biased the trend estimation by using minimum counts rather than actual population estimates, and furthermore that the method does not permit calculation of valid confidence limits. They supported other estimator functions based on recording the number of sightings of each female over the course of the year (analogous to building a “capture history” in a mark–recapture study). These asymptotic estimates of population size are less affected by variation in detectability (Boyce et al. 2001). Sighting history can then be used with various models to estimate the total number of females in the population. Keating et al. (2002) used Monte Carlo simulations to test a number of non-parametric $F_{\text{cub}}$ estimator models and determined that the Chao (Chao 1984, 1989) and second-order sample coverage (Chao and Lee 1992) estimators were the best in terms of robustness to variation in number of unique females, overall sample size, and coefficient of variation. Further simulations suggested that the bias-corrected Chao estimator (Chao 1989, referred to as “Chao2” in Cherry et al. 2007) should be used for management applications because it is less likely to be biased high than the sample-coverage estimator (Cherry et al. 2007). It should be noted that the bias of the Chao estimator increases as detectability becomes more heterogeneous and sample size decreases (K. Keating, US Geological Survey, personal communication, Bozeman, Montana, USA, 2006), but that the magnitude of this bias is much lower than the potential bias incurred by assuming a female population size equal to the number of raw $F_{\text{cub}}$ observed.

We tested the $F_{\text{cub}}$ summation estimator (Knight et al. 1995) and the bias-corrected Chao estimator (Chao 1989, Keating et al. 2002) using data collected in Banff–Kananaskis from 1993–2004, concurrent with the demographic study of Garshelis et al.
For each estimator, we converted raw estimates of $F_{\text{cub}}$ into indices of female population size. The $F_{\text{cub}}$ summation estimator (Knight et al. 1995) was:

$$\hat{N}_t = \sum_{i=1}^{t} \hat{m}_t$$ (1)

where $\hat{m}_t$ is the estimated number of unique $F_{\text{cub}}$ in year $t$. This is essentially Eberhardt and Knight’s (1996) method, as exemplified by Morris and Doak (2002), except that we used a 4 instead of a 3-year summation because Banff–Kananaskis grizzlies have 4-year mean interbirth intervals (Herrero 2005).

The second estimator we tested was the bias-corrected Chao estimator (Chao 1989, Keating et al. 2002):

$$\hat{N}_{\text{Chao}}(t) = \hat{m}_t + \frac{(f_1^2 - f_1)}{2(f_2 + 1)}$$ (2)

where $f_1$ and $f_2$ are the numbers of unique $F_{\text{cub}}$ seen once and twice, respectively, in year $t$.

For both estimators, annual log growth rates were calculated as:

$$\ln(\hat{l}_t) = \ln\left(\frac{\hat{N}_{\text{est}}(t + 1)}{\hat{N}_{\text{est}}(t)}\right)$$ (3)

where $\hat{N}_{\text{est}}$ was either $\hat{N}_{\text{sum}}$ or $\hat{N}_{\text{chao}}$. The exponent of the mean of these rates provides an estimate of $\lambda$ (Dennis et al. 1991, Morris and Doak 2002). We regressed the ln $\hat{l}$ array against an array of ones (the square root of 1-year time intervals between censuses) with the intercept forced through zero (Dennis et al. 1991); the slope of this regression was $\hat{\mu}$. The 95% confidence limits for $\lambda$ were then:

$$\left(e^{(\hat{\mu} - t \times SE(\hat{\mu})}, e^{(\hat{\mu} + t \times SE(\hat{\mu})}\right)$$ (4)

where $SE(\hat{\mu})$ is the standard error of the regression slope and $t$ is the critical value of the 2-tailed Student’s $t$ distribution with a significance level $\alpha = 0.05$ and degrees of freedom equal to the number of transitions in the time series minus 1 (Morris and Doak 2002). We note that the confidence interval for the Chao estimator is based on variance in annual counts, whereas the variance for the summation estimator is reduced by the autocorrelation inherent in the method.

Finally, we used simulations in MATLAB (version 7.0.4; The Math Works Inc., Natick, Massachusetts, USA) to estimate the sensitivity of the indices to variation in the number of unique females observed (e.g., caused by observation error such as missing observations that would underestimate population size in a year as well as misidentifications that could overestimate population size). For each index-based estimator (Chao and summation), we created 10,000 simulated time-series with 12 years each (1993–2004). For each year we randomly selected a population size from a Poisson distribution with a mean equal to the observed $m_t$ for that year. We randomly drew the ratios of $f_1$-simulated and $f_2$-simulated to $m$-simulated in each year from the range of ratios observed in the data. We then estimated $\lambda$ for each simulated time-series using equation 3. The degree to which random variation in the annual population estimates affected measures of population growth rate was assessed as the difference between the mean simulated $\lambda$ estimates and the observed $\lambda$.

Results

Eleven years of field monitoring recorded year-to-year fluctuations in records of $F_{\text{cub}}$ (Fig. 1, Table 1). Because human activity was relatively common in both jurisdictions, sightings from all regions within the study area ($n = 407$) were easily obtained. Annual female abundance estimates from the Chao estimator were very close to actual counts of $F_{\text{cub}}$. 
When applied to the interval over which the demographic study took place (1993–2004), the point estimate from the Chao estimator was 0.980, that of the summation estimator was 0.981. The \( \hat{N} \) of the matrix model was 1.04. The confidence interval for the Chao estimator was 0.572–1.679; that of the summation estimator was 0.847–1.137 (Fig. 2), closer to the precision of the matrix model (95\% CI = 0.99–1.09).

Sensitivity analysis showed that random fluctuation in bear sightings caused the trend estimates of both monitoring-based estimators to deviate little from zero. Mean differences between simulated and observed \( \hat{l} \) for each estimator. Boxes are 1 standard deviation around the mean, error bars are 95\% percentiles.

Discussion

The bias-corrected Chao estimator inflates raw counts of known \( F_{\text{cub}} \) seen only once, but reduces this inflation by females seen twice. In this study most \( F_{\text{cub}} \) were seen more than twice per year (Table 1), implying that relatively few escaped detection; thus, the Chao-estimated populations are largely the same as actual annual \( F_{\text{cub}} \) counts. This may be a relatively common scenario for relatively small populations in well-studied areas. We also note that \( \hat{n}/m \) ratios (Table 1) differed widely across years. This variability is due to differences in visibility of individual bears, not differences in search effort. In certain years highly habituated animals or those whose home ranges made them especially visible (e.g., from park roads) were observed almost daily due to their proximity to people.
We cannot determine which of the 3 methods (2 monitoring-based estimators and the demographic model) best approximates the true population growth rate. As presented here, the results of the Chao estimator-based analysis are unusable due to their extremely low precision. The wide confidence intervals are due to high inter-annual variance in population estimates which, again, may be fairly common for small populations. For example, a random fluctuation in 2 bear sightings per year has much more effect on abundance estimates in a small population (40% change from \( n = 5 \)) than in a larger one (4% change from \( n = 50 \)). Importantly, the performance of any estimator varies with population size (Keating et al. 2002). Thus, precision should increase positively with population size, introducing a potential source of bias in the resulting estimates of \( \lambda \). No study has yet explored the implications of such bias for management (Keating, personal communication, 2006). Moreover, trend analysis using this estimator cannot account for variation in the proportion of females that breed each year. The \( \hat{\lambda} \) point estimate from the demographic study indicates 4% annual growth. This implies a 40% increase in abundance over the duration of the study, but no such increase was noticed. Exclusion of stochasticity and density dependence from their model, as well as imprecision in age-of-senescence estimates, could have biased their \( \hat{\lambda} \) upwards (Garshelis et al. 2005b).

Though the summation estimator appears to give more precise trend estimates than the Chao estimator, those of the former may be artificially narrow. Inter-annual variance in population estimates is necessarily reduced in a running sum, and to our knowledge no methods account for this in the calculation of confidence limits. Furthermore, given that the female bear population in Banff–Kananaskis is relatively small and that the summation method relies on summing \( F_{cub} \) across the average interbirth interval, this method risks random bias due to over-counting or under-counting bears whose interbirth interval differs from four years (D. Garshelis, University of Minnesota, Grand Rapids, Minnesota, USA, personal communication, 2005).

Sensitivity analysis shows that the bias of both monitoring-based estimators was relatively unaffected by random variation in annual counts. Nevertheless, the variance in the differences between simulated and observed \( \hat{\lambda} \) was relatively high, implying that the trend predictions of both estimators will be importantly affected by, for example, observation error. While the variance of the summation estimator was lower than that of the Chao estimator, this may again be an artifact for the reasons discussed above.

Only through continued \( F_{cub} \)-based monitoring will we learn the minimum time-series length necessary for reasonably precise trend estimates; though if trends are not monotonic during this time, we will likely have very little power to detect thresholds or inflection points. Indeed, trend estimation from monitoring data is often bedeviled by low power (see Doak 1995).

For \( F_{cub} \)-based monitoring to be useful, even over longer time-series, we strongly recommend methods to remove sampling variance and other forms of observation error before estimating confidence intervals. Observation error refers to inaccuracies in population size and trend estimation, part of which (sampling variation) comes from the extrapolation of subsets of the population up to the entire population (Morris and Doak 2002). Some facets of observation error can be reduced by careful attention to detail and accuracy during data collection; there are also techniques to reduce sampling variation during the annual censuses (Morris and Doak 2002). Also, recently developed statistical models allow researchers to separate sampling variance from environmental stochasticity using relatively short (15–20 year) time-series (de Valpine and Hastings 2002, Lindley 2003, Holmes 2004). For example Lindley (2003) shows that a time-series of \( F_{cub} \) counts in Yellowstone can be converted to state space (where the population process and observation process are modeled separately) and, through the application of a Kalman filter (Harvey 1989), likelihood functions can be generated to partition total variance into process and sampling components. It may be possible to apply these methods after only a few more years of \( F_{cub} \)-based monitoring in the Banff–Kananaskis system.

Demographic and monitoring-based estimators remain 2 of the most common methods by which biologists monitor populations, yet few studies compare results generated from the 2 methods. Thus, it is difficult for biologists using one monitoring method to assess how their results would change using an alternative method. In Banff National Park and Kananaskis Country, biologists have conducted a decade-long demographic study of radiocollared grizzly bears that was terminated due to funding and political pressure. It was not feasible to continue the
demographic study over the long term, but it may be possible that an intensive $F_{\text{sub}}$-based monitoring program would eventually be able to confidently estimate population trends and the magnitude of process variation.

Acknowledgments

A decade of research was funded and guided by the many partners contributing to the Eastern Slopes Grizzly Bear Project. E. Crone gave advice on the analysis presented here; the manuscript also benefitted greatly from comments by R. Harris, K. Keating, A. Ordiz, P. McLoughlin, C. Schwartz, D. Garshelis, M. Kauffman, J. Maron, and G. White. These analyses were partly funded by Parks Canada.

Literature cited


Received: 8 May 2006
Accepted: 17 March 2007
Associate Editor: P. McLoughlin