ON SOME COMMON ERRORS
IN ANALYSIS OF AGE-STRUCTURED POPULATIONS

Richard B. Harris
Montana Cooperative Wildlife Research Unit
University of Montana
Missoula, Montana 59812

Lee H. Metzgar
Division of Biological Sciences
University of Montana
Missoula, Montana 59812

Abstract: Analyzing age-structured population data can be useful in deducing factors affecting population trend, and projecting future population levels. However, in a review of recent literature, we noted 3 errors in analyses, all of which may lead to erroneous inferences: 1) Equating $S_x$ derived from a series of ages at death with $l_x$, survival rate, in life-table analyses; 2) pooling birth rates across age-categories of mothers without accounting for mortality of mothers within the pooled age category; and 3) substituting $m_x$ for $F_x$ when calculating the finite growth rate $\lambda$ using a Leslie matrix. We discuss implications of these errors and suggest ways to avoid making them.

Key Words: Population analysis, life table, Leslie matrix

ACTA THERIOLOGICA SINICA 13:217-222

In recent years, many scientists have expanded our knowledge of the dynamics of age-structured populations. In particular, new research has focused on mammalian species with rather complex life-histories. These efforts are of great help in better managing, conserving, and rationally utilizing these species. As wildlife biologists and educators in the United States, we are impressed by the interest, concern, and hard work currently being shown in China toward mammalian species.

However, we have discovered three problems in the use of life tables and Leslie matrices to estimate population trajectory when data consist of a series of ages at death (often determined from skulls or carcasses). In some cases, conclusions have been based on analytical techniques that are inappropriate, or applied incorrectly, given the data available. These errors have also been made by biologists working in the West, and some recent literature has appeared to point them out.
to us (e.g., Caughley and Birch 1971, Caughley 1977, Taylor and Carley 1988). However, these clarifying articles were all written in English, and may not be known to biologists working in China.

In this paper, we describe where some recent analyses have been inappropriately applied, and suggest alternatives. The errors we discuss fall into 3 categories:

1. The use of the standing age structure (i.e., the number of skulls for animals of each age) to generate the $l_x$ series (survivorship), and incorporating these survivorship rates into a life-table;
2. Pooling birth rates across age-categories of mothers without accounting for mortality of mothers within the pooled age category; and
3. The correct calculation of values for the $F_x$ series (the birth parameters) when solving for the finite rate of growth, $\lambda$, using a Leslie matrix.

### Generation of the $l_x$ (survivorship) series from ages at death for life table analyses

As has been recognized in other papers (Caughley 1966, Caughley 1974, Caughley 1976, Caughley and Birch 1971, Downing 1980:253, Murray and Garding 1984, Wilson and Bossert 1981), the probability of surviving to age $x$ (usually referred to as $l_x$) relates to the population age structure (usually referred to as $S_x$) by the following equation:

$$ l_x = S_x e^{rx} \quad (1) $$

In this equation, $e^r$ represents the ratio by which the population changes in one unit of time, (for example, one year), so that $e^r = N_{t+1}/N_t$, where $N_t$ = population size at time $t$. By convention, $S_x$ represents the number in age category $x$ as a proportion of the number of newborns, that is:

$$ S_x = s_x/s_0 \quad (2) $$

where $s_x$ animals actually occupy age category $x$ during the data collection period.

In some papers, including those by Wei et al. (1989) on giant pandas ($Ailuropoda melanoleuca$), Sheng et al. (1990) on black muntjack ($Muntiacus crinifrons$), and Yang et al. (1990) on musk deer ($Moschus berezovksii$), the two series, $l_x$ and $S_x$ have been equated, and the analysis
continued to estimate the intrinsic rate of growth \( (r) \). However, doing so requires both a stable age structure (that is, \( l_x \) as well as the \( m_x \) schedule [age-specific fecundity rates] constant through a sufficiently long period, usually 1 or 2 generation lengths), and a constant population size throughout the data collection period (that is, \( e^r = \lambda = 1 \), or equivalently, \( r = 0 \)).

Most often, data are lacking to support the requirement that the age structure be stable, or that \( r = 0 \). More importantly, if the assumption that \( r = 0 \) is known to be true, the remaining analysis (that is, estimating \( r \)) is unnecessary. Continuing with such an analysis is not meaningful because "the calculation simply retrieves an assumption and disguises it as a conclusion" (Caughley and Birch 1971). This type of tautological error has frequently been made by scientists working in North America, Europe and Australia. Caughley and Birch (1971) cite a number of examples; prominent among them are Lowe (1969), and Watson (1970).

If the objective of life-table analysis is to determine the population's rate of increase over some time period in the past, or its potential rate of increase at the present, one must in general use a method of determining age-specific mortality that does not rely on the age-structure. This is much more difficult because it usually requires following individually marked animals.

A hypothetical example can serve to illustrate these points. Consider a study of a small ungulate species in which skulls are collected from the field. The skulls are an unbiased sample of the animals dying in year \( t \), except for the youngest age class which is known to be under-represented because the skulls are small and fragile, thus easily destroyed. From zoo studies, it is known that adult females produce, on average, 0.5 females per year beginning in their second year of life, and continue to do so until they are 5-year olds, at which time their reproductive rate declines slightly, to about 0.452 female young/year. Females in older age classes do not produce young. A total of 505 skulls from age 1+ animals dying during year \( t \) are collected. These hypothetical data appear in Table 1. Because of the known bias in the number of 0-year old skulls collected, this entry in the table is completed by applying the known \( m_x \) values to the remainder of the age-structure. This scenario roughly mimics a typical study using skulls to develop an age-structure (e.g., Wei et al. 1989). If this age-structure is taken directly to represent the \( l_x \) schedule, the life table appears as in Table 1, and the resulting \( r = 0 \) (or, \( \lambda = 1.0 \)). That is, the population
Harris and Metzgar  4/7/92: Errors in population analysis

appears to be stable.

However, these 505 skulls (column 3 of Table 1) could just as easily result from a very
different scenario. Imagine that a larger population in existence during year $t-1$ had recently come
under increased mortality pressure, such that yearly adult survival rates had fallen during year $t-1$ to
0.5 (Table 2). Even if the $m_x$ values had not changed, i.e., were identical to those in Table 1, this
population would be declining rapidly.

In actuality, the larger population existing during year $t-1$ in Table 2 should have produced
354, rather than 200 young. However, this is precisely where data of this type are usually weakest;
we generally must estimate the number in the 0-year class, rather than depend directly on the
number found. Had we done so following the above procedures, we would again have completed
the table by entering 200 in the 0-year age-class.

Thus, if the biologist had no knowledge of population sizes in years $t-1$ and $t$, there would
be no way to determine which of these 2 scenarios was correct. The data in each case, in the form
of the $S_x$ series determined from skulls and the $m_x$ series from zoo studies, are identical. Yet, the
actual population dynamics are quite different. If one analyzed the data in Table 2 according to the
assumptions of Table 1, one would conclude that the population was stable. Conversely, correct
inference could be made if a sample of individually identified animals were followed from year $t-1$
to year $t$; this would have produced true estimates of yearly survival rates (0.9 in Table 1, 0.5 in
Table 2), and thus a correct estimation of $r$.

This hypothetical example may be extreme, but it serves to illustrate that, in general, a series
of ages at death ($S_x$) cannot be substituted for survival rates ($l_x$). Additional data will always be
necessary to correctly estimate the population's rate of growth.

Pooling age classes.

Quite frequently, sample sizes are insufficient to allow constructing a life-table with each
year acting as an age-class, even when individuals have been categorized as such. In such
situations, individuals alive at the start of each n-year age interval may be multiplied by the number
of young that would be produced by a female that lives through the entire age interval. For example, in the paper by Xia and Hu (1989), female giant pandas entering the 9-11 age group were grouped together, and the reproductive rate for this group (1.0 female young over the 3 years, in this example) calculated as if all females survived through that age category. However, when calculated this way, the results will suggest an artificially high rate of population increase.

The problem arises because not all of the females that enter any three-year age category will live to their second and third birthdays within that age category. Continuing with the example from Xia and Hu (1989), while 0.321 newborns will live to the start of the 9-11 age category (0.36 X 0.95 X 0.94 = 0.321), only some smaller fraction of them will be alive on their tenth and eleventh birthdays.

We suggest two solutions to this dilemma. First and most simple, the matrix calculations may be done with each year treated separately. The difficulty is, as already stated, that sample sizes will be quite small, and thus estimates of yearly survival and reproductive rates unstable.

Alternatively, one may wish to discount the \( m_x \) values for the second, third, etc. years of each n-year age interval, so that the combined \( m_x \) values more accurately represent the number of female births produced by a female alive at the start of the interval. Again using the Xia and Hu (1989) example, newborn females produce 0.321 female young as they pass through the 9-11 age category. However, not all females entering the category live through it. One simple way to discount is to assume constant yearly mortality rates during the interval (in this case, \([(1-0.86)/3]\)), and construct a life-table such that these females will produce not 0.321 female young, but rather 0.306 female young \([(0.321 \times 0.33) + (0.306 \times 0.33) + (0.292 \times 0.33) = 0.306]\). The 0.306 and 0.292 figures come from a linear application of the survival rate to the 0.321 that began the interval. The disadvantage of this approach lies in its combination of natality and survivorship in a manner that some readers may find confusing.

**Birth parameters for the Leslie matrix projection.**

Birth rate data coming either from field or zoo studies are usually in the form of female
young/mother at the time of birth, which in the life-table method is usually referred to as $m_x$. Some recent papers, including Xia and Hu (1989), and Yang et al. (1990), have constructed Leslie matrices using $m_x$ for the birth parameter. However, for use in the Leslie-matrix method of estimating $\lambda$, the appropriate statistic to use in adding new members to the population is not $m_x$, but is $F_x$. The difference between $m_x$ and $F_x$ is subtle but important. The $F_x$ parameter was defined by Leslie (1948) as:

"...the number of daughters born in the interval $t$ to $t+1$ per female alive aged $x$ to $x+1$ at time $t$, who will be alive in the age group 0 to 1 at time $t+1$...".

Put another way, Leslie's $F_x$, as used in the matrix formulation, is the number of females alive that were born to mothers during the previous time interval. As such, $F_x$ incorporates both the raw birth rate and the probability of the newborns surviving the initial period until the next census. Other sources to consult on this issue include Emlen (1973:251), Caughley (1977:111), and Taylor and Carley (1988). This confusion in definition and nomenclature between life-table and Leslie matrix methods has plagued researchers in the U.S. and Canada lately. If one substitutes $m_x$ for $F_x$ in the Leslie matrix calculations, the resulting population projection will be biased in a positive manner except in those rare instances in which mortality during the first time period is zero.

**Discussion**

We have outlined 3 errors that commonly occur when using age-structure data to project population trends. These issues are quite technical in nature, and may appear to have only minor consequences. However, biases introduced by incorporating these errors can be significant when considered across the entire age structure. If one assumes the $S_x$ data truly represent $l_x$ data when they do not, all analyses will yield population growth rates very near to zero, regardless of the population's true growth rate. When lumping age-classes in analyses without discounting throughout the entire age-structure, biases in the resulting $r$ can be large. For example, when we made corrections (as suggested above) in each of the three-year age categories of Xia and Hu (1989), we produced a matrix whose projected population (once a stable age distribution has been
achieved), grew at only about 3% per generation (one generation equals the mean age of giving birth), instead of the nearly 15% per generation estimated by them. Finally, using \( m_x \) in place of \( F_x \) in Leslie matrix projections ignores first-year mortality, which is often high, and a major determinant in overall population dynamics.

Researchers using age-structure data should take care to avoid inappropriate application of these methods. We acknowledge the great difficulty of obtaining detailed population data, particularly in remote locations. Some solutions we recommend here require data that are even more difficult to obtain. Regrettably, we wildlife biologists are rarely able to document population parameters as well as we'd like. However, it is usually better to admit our limitations than to continue with analyses where important assumptions are violated, or where necessary approximations may carry substantial potential for bias. It would be unfortunate if well-intentioned, but faulty use of analytical techniques in population studies led to incorrect conclusions.

ACKNOWLEDGEMENTS

Improvements to the manuscript were suggested by R. Taber, D. Pletscher, B. O'Gara and B. Burke. Translation help provided by Li Fengru.

LITERATURE CITED

Leslie, P.H. 1948. Some further notes on the use of matrices in population mathematics.


Harris and Metzgar  4/7/92: Errors in population analysis

<table>
<thead>
<tr>
<th>Age</th>
<th>Population year t-1</th>
<th>Skulls year t</th>
<th>(1-q_x) year t-1</th>
<th>l_x</th>
<th>m_x</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000</td>
<td>0</td>
<td>200</td>
<td>(200)</td>
<td>0.50</td>
<td>1.000</td>
</tr>
<tr>
<td>0.500</td>
<td>1</td>
<td>100</td>
<td>100</td>
<td>0.90</td>
<td>0.500</td>
</tr>
<tr>
<td>0.500</td>
<td>2</td>
<td>90</td>
<td>90</td>
<td>0.90</td>
<td>0.450</td>
</tr>
<tr>
<td>0.500</td>
<td>3</td>
<td>81</td>
<td>81</td>
<td>0.90</td>
<td>0.405</td>
</tr>
<tr>
<td>0.500</td>
<td>4</td>
<td>72</td>
<td>72</td>
<td>0.90</td>
<td>0.365</td>
</tr>
<tr>
<td>0.500</td>
<td>5</td>
<td>63</td>
<td>63</td>
<td>0.90</td>
<td>0.328</td>
</tr>
<tr>
<td>0.452</td>
<td>6</td>
<td>54</td>
<td>54</td>
<td>0.90</td>
<td>0.295</td>
</tr>
<tr>
<td>0.000</td>
<td>7</td>
<td>45</td>
<td>45</td>
<td>0.90</td>
<td>0.266</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>505</td>
<td>(705)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

0-year old produced in year t+1: 200
r = 0.00, e' = λ = 1.00

Table 1. A hypothetical population, yielding a collection of 505 skulls, and calculations assuming that yearly survival rates (and hence the l_x schedule) are taken directly from the standing age-structure, as represented by skulls. Skulls are not available for the 0-year class; rather this entry is completed by applying age-specific natality rates to the age-structure, as indicated in column 3. This population is stable; r = 0.
<table>
<thead>
<tr>
<th>Age</th>
<th>Population year t-1</th>
<th>Skulls year t</th>
<th>(1-q_x) year t-1</th>
<th>l_x</th>
<th>m_x</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>200</td>
<td>(200)</td>
<td>0.50</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>180</td>
<td>100</td>
<td>0.50</td>
<td>0.500</td>
<td></td>
</tr>
<tr>
<td>0.500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>162</td>
<td>90</td>
<td>0.50</td>
<td>0.250</td>
<td></td>
</tr>
<tr>
<td>0.500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>144</td>
<td>81</td>
<td>0.50</td>
<td>0.125</td>
<td></td>
</tr>
<tr>
<td>0.500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>126</td>
<td>72</td>
<td>0.50</td>
<td>0.063</td>
<td></td>
</tr>
<tr>
<td>0.500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>108</td>
<td>63</td>
<td>0.50</td>
<td>0.031</td>
<td></td>
</tr>
<tr>
<td>0.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>90</td>
<td>54</td>
<td>0.50</td>
<td>0.016</td>
<td></td>
</tr>
<tr>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>81</td>
<td>45</td>
<td>0.50</td>
<td>0.008</td>
<td></td>
</tr>
<tr>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total 505  
(1091)  (705)

0-year old produced in year t+1: 200  
\( r = -0.24, \lambda = 0.78 \)

Table 2. A hypothetical population, yielding the same collection of 505 skulls as in Table 1. Were skulls from 0-year olds available, there would be 354, rather than 200. They are not, hence, the 200 is calculated as in Table 1. Note, however, that unlike in Table 1, this population is declining rapidly; the population during year t-1 was much larger, but a recent decrease in survival rates has produced the skulls observed during year t.