

Supplemental methods

R code example

Cox proportional hazard analysis in R (R Core Team 2023) with the survival package (Therneau 2022) was applied.

```
The code is set up as
result <- coxph(Surv(time, status) ~ Sex + BirthLocation + Birth
year, data = data)
when using the birth year as a continuous variable, and
res.cox <- coxph(Surv(time, status) ~ Sex + BirthLocation + Birth
Decade, data = data)
when using birth decades as cohorts. 'time' is the age of the
animal (either at death or at the set cutoff), 'status' is whether
the animal is dead or alive at the cutoff. Note that the number of
factors used may vary between models.
```

The Siler mortality model

The Siler mortality model consists of three simple competing hazards which are mathematically artificial but have a biological interpretation: a hazard decreasing with age, a constant base hazard and a hazard which increases with the individual's age (Siler 1979). This corresponds to the general pattern of a neonate/juvenile mortality risk that declines with birth (i.e. being highest directly at birth), a basic mortality risk all animals face and an increasing likelihood of dying as animals become older (senescence) (Figure 3A).

The hazard (= risk) at time t (written as $h(t)$) in the model is expressed by five non-negative parameters,

$$h(t) = a_1 \exp(-a_2 t) + b_1 \exp(b_2 t) + c$$

where $a_1 \exp(-a_2 t)$ expresses the decreasing hazard with age, $b_1 \exp(b_2 t)$ is the increasing hazard and c denotes the age-independent base hazard.

It is a routine task to obtain different functions of the age t describing the mortality of a population, as the cumulative hazard contingent conditioned on having survived to age t_0 , by integration of the hazard from t_0 to t , formally,

$$H(t_0, t) = \int_{t_0}^t h(x) dx = \tilde{H}(t) - \tilde{H}(t_0)$$

where

$$\tilde{H}(t) = \frac{e^{-a_2 t}}{a_2 b_2} (a_2 b_2 c t e^{a_2 t} + a_2 b_1 e^{a_2 t + b_2 t} - a_1 b_2)$$

This equation yields the typical U-shaped graph of a Siler mortality curve of a high neonate and declining juvenile mortality, a turning point at the lowest (basal) mortality and an increase in mortality towards old age (Fig. 3A)

Given the cumulative hazard contingent, the survival contingent conditioned on having survived to age t_0 is given as

$$S(t_0, t) = \exp(-H(t_0, t)) = \frac{S(0, t)}{S(0, t_0)}$$

For brevity, we will abbreviate $S(t) = S(0, t)$. Hence, $S(t')$ is the proportion of all born individuals which survive at least until age t' and $S(t_0, t')$ is the proportion of all individuals that reached age t_0 and survived at least until age t' . Simple calculus reveals

$$S(t) = \exp\left(-ct - \frac{a_1}{a_2}(1 - e^{-a_2 t}) + \frac{b_1}{b_2}(1 - e^{b_2 t})\right)$$

This equation yields the typical descending survivorship graph with an initial drop for neonate/juvenile mortality, a possible plateau and subsequent drop (type I survivorship curve), or just a

less steep descent during post-juvenile period (type II curve) (Fig. 3B).

Having defined the survival contingent, we can express the life expectancy. The life expectancy conditioned on having reached age t_0 , $E(t_0, t)$, is the expected additional years lived before dying given that an individual reached at least age t_0 . Analogously as before, we let $E(t) = E(0, t)$. Formally,

$$E(t_0, t) = t + \int_t^{\infty} \frac{S(x) dx}{S(t_0)}$$

and

$$E(t) = t + \int_t^{\infty} S(x) dx$$

Given a higher neonate/juvenile mortality than the basal mortality, the average life expectancy at birth is necessarily lower than the life expectancy at the age of basal mortality. From the age of basal mortality onwards, life expectancy decreases again with age (Fig. 3C).

In the current contribution, we also use the lifespan inequality or its reciprocal – the lifespan equality. Given the life expectancy and the survival contingent, the lifespan inequality and the lifespan equality are defined as

$$L_{ineq}(t) = -\frac{1}{E(t)} \int_t^{\infty} S(x) \cdot \ln(S(x)) dx$$

and, respectively,

$$L_{Eq}(t) = -\ln(L_{ineq}(t))$$

The lifespan inequality can be seen as a measure of how evenly the attained age is distributed among individuals of a population across time; statistically speaking, it is a normalised entropy of the survival contingent.

As for life expectancy, equality is expressed in relation to a certain age. Typically, a population with a higher neonate/juvenile mortality will have a lower equality at birth (because some animals die as neonates/juveniles, whereas some survive to medium age) compared to a population with a lower neonate/juvenile mortality. By contrast, when the age of basal mortality is reached, a population in which all individuals die soon afterwards will have a higher equality at that age, whereas a population in which some individuals achieve very long lives will, at this age, have a lower equality (Fig. 3D).

Fitting the model

While the Siler mortality model itself is easy to describe mathematically, it is far from obvious how to apply the model to specific animal data. Challenges that arise are, possibly, unknown dates of birth, currently alive animals, and small sample sizes in certain age categories. All measures (H, S, E, L) described above define the same model and can be analytically translated into each other. Hence, it is the user's choice which measure is best estimated from the empirical data and yields the most stable fit to the data. We decided to fit the survival contingent. Given a population V , the empirical survival contingent

$$\bar{S}(t)$$

is defined as the proportion of observed individuals whose age exceeded t , formally,

$$\bar{S}(t) = \frac{|x \in V: age(x) > t|}{|V|}$$

This quantity was calculated for $t=1,2,\dots$ where t represents the time in days from an individual's date of birth. Then, the Python

programming language and the scipy library were used to fit $S(t)$ by $S(t)$ under the condition that all parameters are positive.

As the data consisted only of animals born in zoos, all birth dates were known. To overcome the challenge of still living individuals, the death age of each still living individual was sampled from a probability distribution – in other words, it was extrapolated. More precisely, the time that an alive individual x of current age t will survive was described by a truncated geometric random variable with parameter

$$\frac{1}{m-t}$$

m is defined by the conditional empirical mean of the death age of individuals born after 1980, conditioned on reaching at least age t . The estimated death age was set to

$$\min(t + \text{Geom}\left(\frac{1}{m-t}\right), m_{\text{age}})$$

with the maximum lifespan of giraffe $m_{\text{age}}(\text{Giraffe})=39$ (the oldest animal in the dataset). Pseudo-randomness was implemented by numpy's random library. To account for random effects, both in the sampling of the death ages as well as in the fitting of the survival contingent, and to guarantee stable model parameters, the model was fitted 1,000 times per analysed population. In each fit, death ages were sampled independently and a subset of the individuals was drawn with replacement as per standard bootstrap sampling. Reported means of model parameters are with respect to all sources of randomness involved.

Calculations were performed using the scipy library of the Python programming language. The code can be downloaded at https://github.com/Klimroth/siler_mortality_methods, or obtained from the authors (hahnklim@mathematik.uni-frankfurt.de, mclauss@vetclinics.uzh.ch).

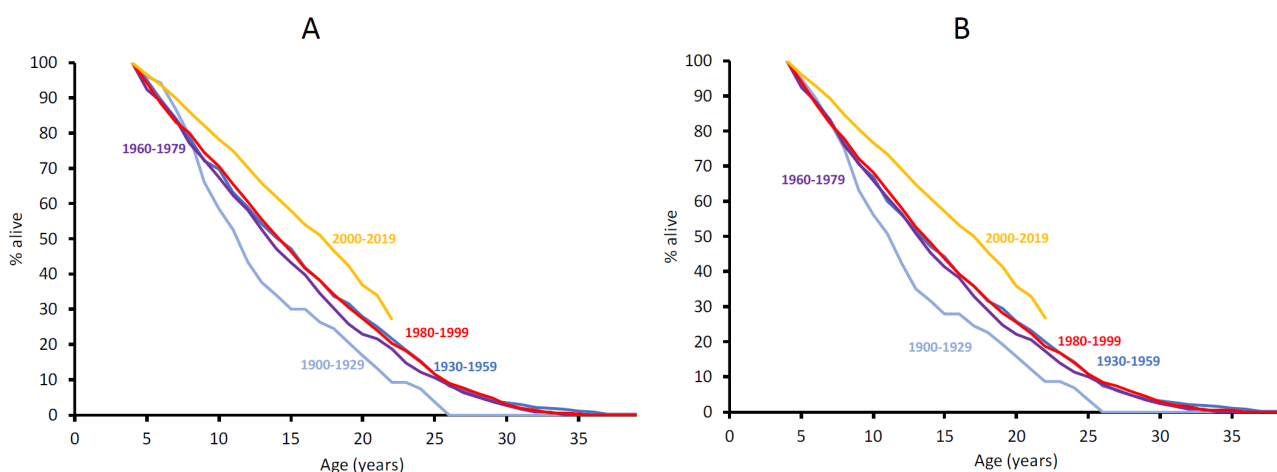


Figure S1 Survivorship graphs for zoo- and wild born giraffe *Giraffa camelopardalis* starting from the age of 4 years (A) excluding and (B) including animals lost to follow up, by individual birth cohorts. For statistics, see Table 1.