



Research

Cite this article: Ewbank D. 2016 Measuring selection in human populations using the growth rate per generation. *Phil. Trans. R. Soc. B* **371**: 20150148.
<http://dx.doi.org/10.1098/rstb.2015.0148>

Accepted: 21 January 2016

One contribution of 14 to a theme issue 'Understanding variation in human fertility: what can we learn from evolutionary demography?'

Subject Areas:
evolution

Keywords:
demography, humans, evolution, reproduction, selection

Author for correspondence:
Douglas Ewbank
e-mail: ewbank@pop.upenn.edu

Measuring selection in human populations using the growth rate per generation

Douglas Ewbank

Population Studies Center, University of Pennsylvania, Philadelphia, PA 19104, USA

Estimates of the speed of evolution between generations depend on the association between individual traits and a measure of fitness. The two most frequently used measures of fitness are the net reproduction rate and the 1-year growth factor implied by the fertility and mortality rates. Results based on the two lead to very different results. The reason is that the 1-year growth factor is not a measure of change between generations. Therefore, studies of changes between generations should use the amount of growth over the length of a generation. This is especially important for studies of human populations because of the long length of generation. In addition, estimates based on a single year's growth are overly sensitive to data on individuals who fail to reproduce. The effects of using a generational measure are demonstrated using data from Kenya and Ukraine. These results demonstrate that using a 1-year growth rate to measure fitness leads to estimates that understate the rate at which evolution changes the characteristics of a human population.

1. Introduction

The composition of the human population is continually affected by the subtle effects of evolution. The rate of evolution is the speed at which the characteristics of a population change. For example, a recent study of a contemporary American population found that if the environment of the 1950s and 1960s continued, evolution would lead to a 1.8% decline in total cholesterol levels over a period of five generations [1]. Similarly, the age at menopause would increase by 0.4 years over the same period. These estimated effects are small compared with the changes resulting from the rapidly changing environment in which we live. However, they demonstrate that even in the context of very low mortality, the effect of evolution in humans is comparable with that in many other species (table 1).

Demonstrating these effects involved two steps. The first step was to demonstrate that the number of children a woman had in her lifetime was statistically related to those characteristics. Second, it was necessary to show that those characteristics tended to be inherited by their children. Those two parts of the evolutionary process, selection and inheritance, are often combined in a model called the 'breeder's equation' to estimate the effects of evolution on the characteristics of the next generation [2,3]. The breeder's equation is a model of change between generations and only gives reasonable estimates of year-to-year changes if the population has had a growth rate near zero in recent decades. Evolutionary demographers refer to this as the problem of overlapping generations [4].

It is rarely possible to measure both selection and inheritance in the same population; therefore, the two parts are generally studied separately. This paper examines an issue in the measurement of selection between generations in human populations. Studies of selection examine the statistical relationship between characteristics of individuals or groups and a measure of evolutionary fitness. Several measures of fitness have been used to study humans. The theoretically preferred measure of fitness is the population growth rate implied by the fertility and mortality rates. That growth rate is generally expressed as growth per year.

Many studies of human populations have examined the selection part of the evolutionary process by examining change between generations [1,5–16]. I have

Table 1. Nomenclature.

variable	definition
LRS	lifetime reproductive success—for an individual, the number of births that survive to the age of reproduction
R_0	for a population, the mean number of births implied by a set of age-specific fertility rates and life table survival rates weighted by the genetic contribution of each parent to the offspring, one-half for humans
r	the long-term population growth rate implied by a set of age-specific fertility and mortality rates. For human populations it is generally expressed per annum
λ	the factor by which the population will grow during a specified period of time given a set of fertility and mortality rates. It is equal to $\exp(r)$. For the present discussion, we assume it is the growth over 1 year
T	the mean length of generation—the time required for a population to increase by a factor of R_0

not found any studies of humans that employ a model that accounts for overlapping generations. Almost all of these studies rely solely on measures of either the quantity of reproduction such as the average children ever born [6,11], or the number of children surviving to a given age [1,5,7,12,15]. Several studies include values of the long-term growth rates [8,9,13,16,17].

The generally recommended method for adjusting for timing differences is to use the 1-year growth factor implied by the fertility and mortality rates, λ . However, it is demonstrated in §6 that using λ for humans leads to smaller selection differentials and, as a result, much slower estimates of the speed of evolution.

2. Selection differentials and selection gradients

Evolution is driven by differences in fitness among individuals that are correlated with heritable characteristics. These differences are described by directional selection differentials which are the change in the mean value of a phenotypic character produced by selection within a generation [2,3]. Selection differentials are interesting by themselves or they can be used as part of a projection of the characteristics of future generations.

Given a vector of characteristics, z , the vector of selection differentials, s , is given by

$$s = \bar{z}^* - \bar{z},$$

where \bar{z}^* is the vector of characteristics after selection [17,18]. Each set of characteristics is associated with fitness values, $W(z)$. Evolution is driven by relative fitness defined as

$$w(z) = \frac{W(z)}{\bar{W}},$$

where \bar{W} is the mean fitness across all individuals or groups. For quantitative traits, the observed selection differentials are given by

$$s = \text{Cov}[w, z].$$

Because selection is measured as the differences between successive generations, the measure of fitness must refer to the difference in the sizes of consecutive generations.

Our focus here is on how the estimates of selection are apt to change when we switch from one fitness measure to another. To do that, we use the definition of the correlation coefficient as the covariance divided by the standard deviations in fitness and that trait, σ_w and σ_i . Then solving for s_i we find:

$$s_i = \text{Cov}(w, z_i) = r_{i,w} \sigma_w \sigma_i,$$

where $r_{i,w}$ is the correlation between fitness and trait i .

It is difficult to predict how the correlation will change if we switch to a different measure of fitness. However, if the two measures of fitness are highly correlated than $r_{i,w}$ is not likely to change radically. On the other hand, as we shall see, we can sometimes predict that σ_w will change substantially when we switch fitness measures.

The selection differentials give the observed association between each trait and selection. However, the selection differential for a given trait is determined not only by its direct effect on selection but also by the fact that it is correlated with other traits that are directly associated with selection. Therefore, to estimate the direct force of selection associated with a given trait, it is necessary to control for the correlations between traits.

The direct force of selection for a trait is given by its selection gradient, β_i . It is equal to the partial regression coefficient from a multiple linear regression of relative fitness on all of the traits. For present purposes, we can simply state that β_i can be expressed as the ratio σ_w / σ_i multiplied by a term based on the correlations among the traits, and the correlations between each trait and relative fitness [19]. As before, it is difficult to guess how switching to a different measure of fitness might affect the correlations between relative fitness and each trait. However, as with the fitness differential, the fitness gradient is directly proportional to the standard deviation in the fitness measure.

It is worth noting that this use of the coefficients from a multiple regression is quite different from the usual application. Generally, the β_i are used to estimate the mean level of the left-hand variable associated with given levels of the right-hand variables. In that case, the ratio of σ_w to σ_i serves to translate from the units used to measure i to those used to measure w . However, when we use the β_i as selection differentials, we combine them with information on inheritance patterns to estimate new values for the right-hand variables. Now changing the scale of the measure of fitness need not change the scale of the new trait values.

3. Measuring fitness

The most commonly used measure of overall fitness for a population is the net reproduction rate, R_0 . It is the average number of offspring per person weighted by the genetic contribution of each parent to the offspring—one-half for humans [4,20]. It is defined in terms of $m(a)$ which is the

fertility rate at each age weighted by one-half and $p(a)$, the probability of surviving from birth to age a giving:

$$R_0 = \int_{\alpha}^{\omega} m(a)p(a) da,$$

where α and ω define the ages over which fertility occurs. Because $m(a)$ is a rate, it is expressed as births per unit of time. Rates for humans are almost always expressed per annum. The $p(a)$ are probabilities and do not depend on the units used for the $m(a)$. In practice, the integral is replaced by a summation using $m(a)$ values for age groups. For humans, R_0 is usually estimated from $m(a)$ values for 1-year or 5-year age groups. If the $m(a)$ are expressed per annum then the age-specific fertility rates are multiplied by five to derive rates per quinquennium for summations based on 5-year intervals.

One problem with R_0 as a measure of fitness is that it does not take into account the pace of reproduction. R_0 is a growth rate per generation. Therefore, two populations with the same value of R_0 grow at different annual rates if they have different lengths of generation. For example, consider two populations with an R_0 of 2.5 but one has a mean length of generation of 26 and the other has a mean of 28. The first is growing at 3.52% per year (i.e. $\ln(2.5)/26$), whereas the other is growing at only 3.27% per year. The same amount of growth per generation is spread over a longer period of time in the second population.

That problem is solved by using the long-term growth rate implied by the $m(a)$ and $p(a)$. If those rates stay constant long enough, the population growth rate will stabilize at r , the intrinsic rate of increase [21]. In practice, it is preferable to use $\exp(r)$ which is λ . Because it is a rate, r is expressed per unit of time. For studies of humans, r is expressed per annum and λ is then the result of 1-year's growth.

The relationship between these measures and R_0 is given by

$$R_0 = e^{rT} = \lambda^T,$$

where T is the mean length of generation measured in the same units as r and λ .

The intrinsic rate of increase and λ can also be calculated from the population projection matrix, \mathbf{A} [4]. The projection matrix is an $n \times n$ matrix, where the data are presented in terms of n age groups of length Δt years. The first row of \mathbf{A} contains the fertility rates, $F(a)$, which are one-half the number of births during the Δt years that survive to the start of the next projection period. The subdiagonal contains the proportion surviving from the start of the age interval until the start of the next interval, $P(a)$. The matrix projects the population Δt years forward.

If \mathbf{A} remains constant over time, the population eventually increases by a factor of $\lambda^{\Delta t}$ every Δt years, where λ is change per year. $\lambda^{\Delta t}$ is equal to the dominant eigenvalue of \mathbf{A} . The annualized growth factor, λ , is then the dominant eigenvalue raised to the power $1/\Delta t$ and the annualized intrinsic growth rate is $\ln(\lambda^{\Delta t})/\Delta t$ [22]. Therefore, with a projection matrix based on 5-year intervals, the annual growth factor is the fifth-root of the dominant eigenvalue.

There is nothing in the model of population growth that dictates what size age intervals we should use. Previous studies of selection in humans have used single years [9,14] or 5-year intervals [8] and the associated values of λ and λ^5 . However, I demonstrate below that estimates of selection differentials based on relative values of $\lambda^{\Delta t}$ are quite sensitive

to the choice of Δt . For example, using 1-year age intervals and λ will lead to quite different estimates of the selection differentials than using 5-year intervals and λ^5 . The two will always give the same relative ranking of subpopulations (subject to differences in the precision of the two estimates). However, the relative magnitudes of selection differentials for different traits based on λ will differ substantially from estimates based on λ^5 (see §6).

It is crucial that the growth factor used to estimate selection differentials is scaled properly. There is no reason why our choice of a fitness measure should depend on the age groups used in the available data or on the size of the projection matrix we are willing to manipulate. The time-scale for measuring fitness need not be the Δt used to form the projection matrix. For example, Lahdenperä *et al.* [23] used λ^{18} . The appropriate time-scale for fitness measures for studying selection is dictated by the definition of selection differentials as the change between generations. This is the only theoretical basis for selecting the appropriate time-scale for the measure of fitness used to estimate selection differentials. Therefore, the theoretically appropriate fitness measure for studying selection is λ^T .

In summary, the problem with comparing values of R_0 for different groups is a problem of differing time-scales. For group i , $R_{0,i}$ measures growth over T_i years. When we use R_0 to compare groups i and j , we are comparing a T_i -year growth factor to a T_j -year growth factor. Using λ for both groups solves this timing problem, but λ is not a generational measure. Therefore, the estimation of selection differentials should be based on growth factors for T years, that is λ_i^T , where T is the length of generation measured for the whole population. In this way, we are using the theoretically preferred measure, a growth rate, and scaling it appropriately for measuring selection.

4. What difference does using a generational measure make?

In most cases, rescaling a measure of fitness does not have any effect on estimates of selection. In particular, multiplying the fitness measure by a constant would not change the fitness values relative to the mean. However, rescaling λ involves raising it to a power which changes all of the relative values.

The first implication of raising the relative fitness values to the power T is that it increases the standard deviation in fitness, σ_w . We can see this in a simple example. In human populations, the values of λ for subpopulations are not apt to differ by more than a few per cent per year. Consider two subgroups with values of λ_i of 1.01 and 1.03 and a mean of 1.02. Their values relative to the mean are then approximately 0.99 and 1.01. Setting T at 29 and switching to λ_i^T leads to fitness values of 1.335 and 2.357. The mean fitness is now 1.846 (which is slightly different from 1.02^{29} or 1.776). This leads to relative fitness values of 0.723 and 1.277. As shown above (§2), increasing the deviations of the fitness of groups from mean fitness will tend to increase the selection gradients. This then translates to a more rapid estimated speed of evolution.

Although λ_i^T will always lead to more variation in fitness compared with what we find using λ , the same is not true of a comparison with R_0 . We can see this by noting that for group i

$$\lambda_i^T = R_{0,i}^{T/T_i}.$$

In human populations, T almost always falls within a few years of age 29 [24] and, therefore, the range of likely values of T/T_i is clustered around 1.0. As a result, the variance in λ_i^T and the covariances between it and other quantitative traits will be similar to those of R_0 .

There is a second important result from this. If groups with high values of R_0 also tend to have high values of T , then using λ_i^T will reduce relative fitness for those groups compared with using R_0 and increase it for groups with lower R_0 and T . In that case, using λ_i^T would reduce the variance in fitness. On the other hand, if groups with high R_0 tend to have lower values of T (perhaps because they start childbearing earlier), using λ_i^T will have the opposite effect: increasing the variance. Therefore, using λ_i^T can lead to larger or smaller variance in fitness than using R_0 . This is different from the effect of using λ which always leads to much lower variance.

5. Data and methods

I have used data from two surveys to demonstrate how changing the measure of relative fitness can change estimated selection differentials. These surveys were designed following the pattern set by the Demographic and Health Surveys (DHSs). The DHS program includes over 300 surveys in more than 90 countries covering important issues in population, health and nutrition. They are representative surveys that are generally national in scope. The central focus of most surveys is data collected from women of reproductive age. The data from the DHSs are freely available to researchers from the website DHSprogram.com.

The data used here come from the women's questionnaires and the pregnancy histories from the 2003 Kenyan Demographic and Health Survey (KDHS) [25] and the 2007 Ukraine Demographic and Health Survey (UDHS) [26]. These two countries were selected to represent extreme levels of fertility. The fertility rates in Kenya in the year preceding the survey implied 4.9 births per woman which implies rapid population growth. The comparable value in Ukraine was only 1.2, well below replacement. In Kenya, about 10.5% of children died before age 5. In Ukraine, only about 7% died before age 5 [27]. When mortality rates are this low, selection differentials are driven by differences in fertility. Both surveys are national in scope and are cluster surveys. All analyses use the appropriate sample weights to adjust for any oversampling of areas. The Kenya survey had a sample size of 7688 women aged 40–49. The sample size in Ukraine was 8007.

The limitation of pregnancy histories is that they only include women who are alive at the time of the survey. In order to get fertility information for all ages, it is necessary to limit the analysis to women who have completed (or nearly completed) reproduction. Therefore, we do not have life histories for women who died before reaching the end of the childbearing years. Estimates of λ for populations based solely on the maternity histories will be biased upward.

To incorporate mortality during the childbearing years, the completed maternity histories were used as a collection of potential histories. Each history was randomly assigned six ages at death over age 15 based on life tables for the two countries around 1990 [28]. The six replications for each history were then used as separate observations in the analyses by excluding births after the assigned age at death. The statistical tests then weighted each observation by one-sixth times the

original sample weight. With this design, the proportion childless includes women reaching age 40 without having a child that survived to age 5 as well as women whose simulated age at death preceded the birth of their first child.

This procedure incorporates the assumption that the risk of death during the childbearing years is not correlated with the phenotypes being studied. Therefore, the estimated selection differentials only incorporate the effects of fertility differentials. In addition, the populations of Kenya and Ukraine cannot be assumed to be a single breeding population. For these reasons, the estimated selection differentials may not be relevant to the real population of these countries.

Lande & Arnold [18] have shown how the selection gradients can be estimated using standard regression programs and data for individuals. McGraw & Caswell [29] proposed that this approach can be based on an estimate of λ for individuals. Their approach is to create a population projection matrix for individual i , $A^{(i)}$, with the age-specific number of births in the top row and ones in the subdiagonal for each year the individual survives. The dominant eigenvalue for this matrix is λ_i if the matrix is in single years of age.

The lifetime reproductive success (LRS) for each woman was calculated as the number of a woman's children who survived to 5 years of age. Because LRS is not adjusted for the genetic contribution to offspring, mean LRS is about twice as large as R_0 for the population. Then λ for individuals was estimated as $\ln(\text{LRS}_i/2)/T_i$. Following Coale [24], T_i was estimated as the average of the mean age of the fertility schedule and the mean age of mothers in a population with the growth rate implied by λ . This approximation is quite precise for humans and greatly simplified the calculations. The mean LRS of women aged 40–49 in Kenya is 5.20, three times as high as the mean in Ukraine (1.68).

Each survey provides data on a phenotype that is useful for estimating selection gradients. The KDHS included measurements of height (mean \pm 1 s.d.: 159.4 \pm 6.38 cm) and weight (60.19 \pm 13.49 kg). The UDHS provides systolic and diastolic blood pressure measurements (means 131.6 \pm 18.13 and 86.6 \pm 13.46). Therefore, the data provide a realistic example of the performance of different fitness measures using data for individual life histories.

All analyses were completed in STATA using appropriate weights.

6. Results

Table 2 shows fitness measures for three groups of Kenyan women defined by BMI (weight/height²). The mean LRS declines from 5.56 for the leanest to 4.32 for the heaviest. T_i declines from 27.02 to 25.66, a difference of 1.37 years. Therefore, we expect that adjusting for differences in T would reduce the relative advantage of the lean and moderate groups compared with the heaviest. In this case, using λ_i rather than LRS reduces the relative advantage of the leanest relative to the heaviest from 1.29 to only 1.03. By contrast, using λ_i^T leads to relative fitness values that are similar to the ratios using LRS. The use of λ_i also leads to a reduction in the variance of relative fitness across all individuals. The s.d. using λ_i (0.29) is about half as large as the values using LRS and λ_i^T (0.54 and 0.55).

Table 3 presents a similar comparison using systolic blood pressure data from the Ukraine. Again we see that using λ_i

Table 2. Estimated absolute and relative values of LRS, λ and λ_i^T , and mean length of generation (T) by body mass index (BMI), Kenya [25].

	LRS	λ	λ_i^T	T
means				
BMI < 20	5.55	0.977	2.69	27.02
BMI 20–27.5	5.11	0.949	2.55	26.28
BMI 27.5+	4.32	0.947	2.22	25.66
relative fitness values				difference in T
BMI < 20	1.29	1.03	1.21	1.37
BMI 20–27.5	1.18	1.00	1.15	0.62
BMI 27.5+	ref.	ref.	ref.	ref.

Table 3. Estimated absolute and relative values of LRS, λ and λ_i^T , and mean length of generation (T) by systolic blood pressure (SBP), Ukraine [26].

	LRS	λ	λ_i^T	T
means				
SBP < 130	1.62	0.937	0.805	24.98
SBP < 140	1.69	0.927	0.838	25.12
SBP < 160	1.78	0.951	0.881	25.26
SBP < 180	1.88	0.913	0.930	24.47
SBP > 179	1.88	0.936	0.925	25.36
relative fitness values				difference in T
SBP < 130	0.86	1.00	0.87	–0.38
SBP < 140	0.90	0.99	0.91	–0.24
SBP < 160	0.95	1.02	0.95	–0.10
SBP < 180	1.00	0.97	1.01	–0.89
SBP > 179	ref.	ref.	ref.	ref.

leads to much smaller differences among the groups. The relative values using LRS range from 0.86 to 1.0, a range of 0.14. Using λ_i , the range shrinks to only 0.04. Also the standard deviation in relative fitness among all individuals is reduced by about 50% from 0.53 using LRS to 0.24 using λ_i .

Table 4 presents selection gradients for Kenya and Ukraine from regressions of relative fitness using LRS, λ_i , and λ_i^T on age at first birth and age at last birth. All of the estimates suggest that lower age at first birth and higher age at last birth are associated with higher fitness. However, using λ_i leads to selection gradients that are only 5.7% as large as those estimated using LRS in Kenya and only 3.4% as large in Ukraine.

In Kenya, the three fitness measures also lead to very different estimates of the relative importance of ages at first and last birth. The ratio of the two gradients is determined solely by the correlations and not affected by the change in the variation in fitness. LRS suggests that the ages at first and last birth have roughly the same effect on fitness. Using λ_i leads to a coefficient on age at first birth that is 1.32 times as large as the coefficient on age at last birth. Using λ_i^T leads to an even larger ratio: 1.63. In Ukraine, λ_i and λ_i^T lead to about the same effect on the relative importance of the two ages.

Studies of ages at first and last birth exclude childless individuals for whom these variables are not defined.

Table 4. Estimated selection gradients for ages at first and last birth based on different measures of fitness: LRS, λ and λ_i^T [25,26].

	selection gradient for age at:		
	1st birth	last birth	ratio of gradients
Kenya			
LRS	–0.053	0.052	–1.01
λ	–0.003	0.002	–1.32
λ_i^T	–0.070	0.043	–1.63
Ukraine			
LRS	–0.085	0.076	–1.12
λ	–0.003	0.003	–0.97
λ_i^T	–0.079	0.076	–1.03

When we examine other characteristics, differences in the prevalence of childlessness can become a very important determinant of selection differentials. Estimates based on λ_i and λ_i^T differ in the extent to which they are determined by the proportion childless.

Figure 1 plots values of LRS and λ_i for individual Kenyan women. The value of λ_i for nulliparous individuals (zero) is an extreme outlier. Having a single birth that survives increases λ_i by about 10 times as much as going from a single child to 12 children. The levelling off of the curve is surprising given that evolution is driven by differences in the number of descendants produced by individuals with different characteristics. It is not plausible that fitness should almost cease responding to differences in reproduction after only a few births. Because of the similar magnitudes of R_0 and λ_i^T , a graph of LRS and λ_i^T would show a nearly linear relationship between reproduction and fitness. The graph for Ukraine would look almost exactly like figure 1 except for the lack of data at the highest levels of LRS. At each value of LRS, the two graphs differ imperceptibly because of small differences in the mean length of generation.

Studies by Jones & Bird [14] and by Käär & Jokela [8] present similar graphs. Their graphs show similar effects, although they appear somewhat different from figure 1. Jones and Bird have examined this relationship using historical data from Utah. Their graph excludes childless women and the vertical axis begins above 0.95. This stretches out the curve for LRS values above unity which visually de-emphasizes the flattening. Their discussion emphasizes both the decline in the marginal increase in λ and the enormous increase in fitness associated with having a first birth.

Käär and Jokela [8] examine the relationship between LRS and λ using historical data for the Sami populations of northern Finland. They begin their curve at three surviving children (an LRS of 1.5 by their definition). Their curve appears less flat than figure 1 or Jones and Bird's curve. The reason is that they calculate λ using a projection matrix based on 5-year age groups. Therefore, they are actually plotting a 5-year growth factor. For example, a value of three surviving children would not be sufficient to give an annual increase of 6% for a human population. Although their x -axis is labelled from 1.0 to 1.3, that is equivalent to 1-year growth factors ranging from 1.0 to 1.054 which is similar to the range shown in figure 1 and the figure given by Jones and Bird. If the data from these two studies were plotted on figure 1, they would look very similar to the

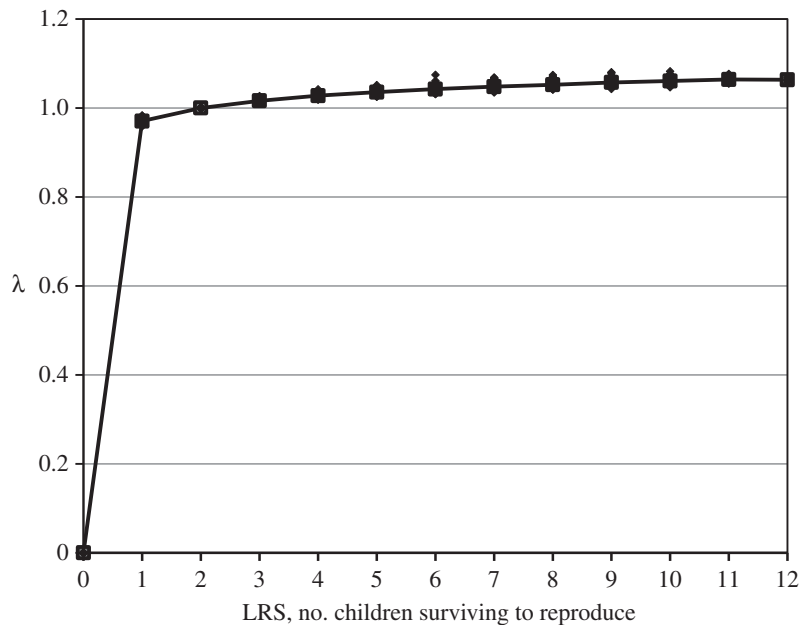


Figure 1: Relationship between lifetime reproductive success (LRS) and estimates of the 1-year intrinsic growth rate, λ , for Kenyan women [25].

data from Kenya. Any differences would be due to differences in the values of T in these populations.

With such a large difference between the value of λ for an LRS of zero and the values for all other parities, the regression estimates can be dominated by differences in the proportion childless across groups. With a relatively high proportion of childless individuals, a regression on λ approaches a regression on a binary measure of childlessness. The sensitivity of selection differentials based on λ to childlessness is apparent in data presented by Korpelainen [13]. She presents mean LRS and mean λ for groups of women in Finland defined by year of birth (1870–1899 and 1900–1929) and three ages at death (less than 50, 50–79 and 80+). The variance among the six estimates of relative fitness using LRS is 0.023, whereas the variance among the values based on λ_i is only 0.008. The differences in mean λ_i are almost completely determined by the proportion childless. The proportion of women without surviving children varies from 4.3 to 20.7% across the six groups and the number of births per women with children varies from 1.41 to 2.05. The correlation of mean LRS with the proportion childless was -0.87 , whereas the correlation between mean λ and childlessness was -0.98 .

Table 5 demonstrates this effect in regressions for Kenya using the estimated selection gradients associated with height and weight. Just as in table 2, using λ_i leads to a huge reduction in the selection gradients. This is true for both height and weight. In addition, dropping the women who failed to reproduce has a much larger effect when using λ_i . The selection gradients based on LRS and λ_i^T change very little when we limit the sample. The estimates for height drop by 12% and 15% and the estimates for weight both change by 5%. However, when λ_i is used as the measure of fitness, dropping the childless women reduces the estimates by 78% and 57%.

In addition, the relative importance of the two measures changes substantially using λ but remains virtually unchanged using LRS and λ_i^T . Using LRS, the ratio of the gradient for height to the gradient for weight changes very little when we drop the childless women (-1.33 versus -1.22). The same is true using λ_i^T . However, the ratio of the two

Table 5. The effect on estimated selection gradients for height (β_{ht}) and weight (β_{wt}) of limiting the sample to woman with children that survived to age 5 using three measures of fitness: LRS, λ and λ_i^T [25].

	selection gradient for:		
	height	weight	β_{ht}/β_{wt} ratio
LRS	0.0098	-0.0074	-1.33
λ	0.0017	-0.0007	-2.47
λ_i^T	0.0084	-0.0061	-1.37
limited to LRS > 0			
LRS	0.0086	-0.0070	-1.22
λ	0.0003	-0.0003	-1.09
λ_i^T	0.0071	-0.0058	-1.24

gradients drops dramatically when using λ (-2.47 to -1.09). This demonstrates the extent to which the extreme value of λ_i for the nulliparous can cause the selection differentials to be very sensitive to the characteristics of childless individuals. Similar effects are seen in analyses of blood pressure in Ukraine (results not shown).

This sensitivity to observations with LRS of zero is especially troublesome given a problem highlighted by McGraw & Caswell [29]. They discuss the use of a population projection matrix for person i , $\mathbf{A}^{(i)}$, as an estimate of that person's propensity to reproduce. In that model, each individual has risks of giving birth and dying at each age. However, we do not observe their underlying risks. As McGraw and Caswell note, the observed life history is only one of the possible lives which that individual might have lived. They note that it is the propensity to reproduce that is relevant for the study of evolution, not the observed reproduction. In most cases, the observed LRS for an individual can be assumed to be an unbiased estimate of their propensity to reproduce and certainly the mean values for groups provide unbiased estimates. However, McGraw and Caswell note that that is not true for most individuals that

did not reproduce. Most individuals with LRS of zero had a non-zero propensity to reproduce and, therefore, their observed LRS provides a biased estimate. As a statistical fix for this problem, they proposed estimating the propensity to reproduce for these individuals using regressions of survival and fertility on observed characteristics. It is possible that something of this sort could reduce this problem. In the meantime, estimates of selection differentials based on λ_i are much more heavily affected by these biased estimates than are differentials estimated from LRS or λ_i^T .

7. Summary

The mathematical models of population growth and evolution provide remarkable insights into the complex interactions of genetics, environment and evolution. The strength of these models is the extent to which they generalize basic biological phenomena that are observed in whales, trees, bacteria and viruses. One reason for this is that they are scalable with respect to time. However, it is necessary to specify a unit of time when they are applied to a specific species. In general, the unit of time appropriate for studying population dynamics will not be an appropriate unit of time for studying selection and evolution.

Studies of human populations almost always use annual fertility, mortality and growth rates, and single-year or 5-year age groups. However, selection differentials are based on the changes in phenotype distributions between generations. The appropriate measure of the length of a generation is the mean length of generation, T . Therefore, it is necessary to rescale fitness measures derived from the population growth models before estimating selection differentials. In practice, this involves raising the annual growth factor for each individual or group, λ_i , to the power T measured for the whole population.

In many cases, rescaling a measure has little effect on substantive findings. However, using λ to study selection leads to radically different results than using R_0 or LRS. It is not intuitively clear that adjusting a quantitative measure for slight differences in timing should have such a radical effect. It is

especially not obvious that an adjustment for timing should always greatly reduce estimates of selection—even when quantity and timing are negatively correlated.

The effect of using λ_i^T rather than λ for studying selection is especially large for studies in humans. The first reason is simply that T for humans is very large. Raising estimates of λ for groups or individuals to the 25th or 30th power has a large effect on the variation in fitness and the selection gradients. Secondly, selection gradients based on λ for individuals are very sensitive to observations for childless individuals because of the extreme value of λ .

In a population, λ^T is equal to R_0 . Therefore, when we adjust for timing differentials by raising the estimate of λ_i for each individual or group to the value of T for the whole population, we derive a measure that is scaled in the same units as R_0 . As a result, λ_i^T can be considered a timing adjusted R_0 .

Evolution is basically a game of numbers: what kinds of individuals produce the largest number of descendants. Differences in the timing of reproduction and the mean length of generation make a difference but this is secondary to the basic numbers. The solution to the timing issue is to use the long-term growth factor implied by the quantity of reproduction (lifetime reproduction) and the mean length of generation. However, using the implied 1-year growth factor for humans leads to relative fitness values that are radically different from the basic quantities of reproduction. As a result, estimates of the speed of evolution using a 1-year growth rate are radically different from those based on the unadjusted quantities. Using a growth factor that is consistent with the length of generation leads to fitness values that are much more similar to the numbers that dominate evolution. This approach maintains the primacy of the amount of reproduction while providing a modest adjustment for timing.

Competing interests. We have no competing interests.

Funding. We received no funding for this study.

Acknowledgement. This manuscript benefitted greatly from numerous suggestions and probing questions from the reviewers.

References

- Byars SG, Ewbank D, Govindaraju DR, Stearns SC. 2010 Colloquium papers: natural selection in a contemporary human population. *Proc. Natl Acad. Sci. USA* **107**, 1787–1792. (doi:10.1073/pnas.0906199106)
- Lush JL. 1943 *Animal breeding plans*. Ames, IA: Iowa State College Press.
- De Jong G. 1994 The fitness of fitness concepts and the description of natural selection. *Q. Rev. Biol.* **69**, 3–29. (doi:10.1086/418431)
- Caswell H. 2001 *Matrix population models: construction, analysis, and interpretation*, 2nd edn. Sunderland, MA: Sinaur Associates.
- Low BS. 1991 Reproductive life in 19th-century Sweden: an evolutionary perspective on demographic phenomenon. *Ethol. Sociobiol.* **12**, 411–448. (doi:10.1016/0162-3095(91)90024-K)
- Røskoft E, Wara A, Viken Å. 1992 Reproductive success in relation to resource-access and parental age in a small Norwegian farming parish during the period 1700–1900. *Ethol. Sociobiol.* **13**, 443–461. (doi:10.1016/0162-3095(92)90012-5)
- Käär P, Jokela J, Helle T, Kohola I. 1996 Direct and correlative phenotypic selection on life-history traits in three pre-industrial human populations. *Proc. R. Soc. Lond. B* **263**, 1475–1480. (doi:10.1098/rspb.1996.0215)
- Käär P, Jokela J. 1998 Natural selection on age-specific fertilities in human females: comparison of individual-level fitness measures. *Proc. R. Soc. Lond. B* **265**, 2415–2420. (doi:10.1098/rspb.1998.0592)
- Korpelainen H. 2000 Fitness, reproduction and longevity among European aristocratic and rural Finnish families in the 1700s and 1800s. *Proc. R. Soc. Lond. B* **267**, 1765–1770. (doi:10.1098/rspb.2000.1208)
- Clarke AL, Low BS. 2001 Testing evolutionary hypotheses with demographic data. *Popul. Dev. Rev.* **27**, 633–660. (doi:10.1111/j.1728-4457.2001.00633.x)
- Low BS, Simon CP, Anderson KG. 2002 An evolutionary ecological perspective on demographic transitions: modeling multiple currencies. *Am. J. Hum. Biol.* **14**, 149–167. (doi:10.1002/ajhb.10043)
- Strassmann BI, Gillespie B. 2002 Life-history theory, fertility and reproductive success in humans. *Proc. R. Soc. Lond. B* **269**, 553–562. (doi:10.1098/rspb.2001.1912)
- Korpelainen H. 2003 Human life histories and the demographic transition: a case study from Finland, 1870–1949. *Am. J. Phys. Anthropol.* **120**, 384–390. (doi:10.1002/ajpa.10191)
- Jones JH, Bird RB. 2004 The marginal valuation of fertility. *Evol Hum Behav.* **35**, 65–71. (doi:10.1016/j.evolhumbehav.2013.10.002)
- Alvergne A, Jokela M, Lummaa V. 2010 Personality and reproductive success in a high-fertility human population. *Proc. Natl Acad. Sci. USA* **107**, 11 745–11 750. (doi:10.1073/pnas.1001752107)

16. Gillespie DOS, Russell AF, Lummaa V. 2013 The effect of maternal age and reproductive history on offspring survival and lifetime reproduction in preindustrial humans. *Evolution* **67**, 1964–1974. (doi:10.1111/evo.12078)
17. Falconer DS. 1981 *Introduction to quantitative genetics*, 2nd edn. London, UK: Longman.
18. Lande R, Arnold SJ. 1983 The measurement of selection on correlated characters. *Evolution* **37**, 1210–1226. (doi:10.2307/2408842)
19. Johnston J. 1972 *Econometric methods*, 2nd edn. New York, NY: McGraw-Hill.
20. Stearns SC. 1992 *The evolution of life histories*. Oxford, UK: Oxford University Press.
21. Cohen JE. 1979 Ergotic theorems in demography. *Bull. New Ser. Am. Math. Soc.* **1**, 275–295. (doi:10.1090/S0273-0979-1979-14594-4)
22. Keyfitz N, Caswell H. 2005 *Applied mathematical demography*, 3rd edn. New York, NY: Springer.
23. Lahdenperä M, Lummaa V, Helle S, Tremblay M, Russell AF. 2004 Fitness benefits of prolonged post-reproductive lifespan in women. *Nature* **428**, 178–181. (doi:10.1038/nature02367)
24. Coale AJ. 1957 A new method for calculating Lotka's—the intrinsic rate of growth in a stable population. *Popul. Stud. (Camb.)* **11**, 92–94.
25. Kenya Central Bureau of Statistics, Kenya Ministry of Health, ORC Macro 2004 Kenya demographic and health survey 2003. Calverton, MD: CBS, MOH, and ORC Macro.
26. Ukraine Center for Social Reforms, Macro International 2008 Ukraine demographic and health survey 2007: key findings. Calverton, MD: UCSR and Macro International.
27. UNICEF. 2012 *Ukraine: maternal, newborn & child survival*. New York, NY: UNICEF.
28. United Nations DoEaSAPD. 2013 World population prospects: the 2012 revision 2013. DVD edition. (POP/DB/WPP/Rev.2012/MORT/F17-3)
29. McGraw JB, Caswell H. 1996 Estimation of individual fitness from life-history data. *Am. Nat.* **147**, 47–64. (doi:10.1086/285839)