ECOLOGY LETTERS

Ecology Letters, (2016) doi: 10.1111/ele.12619

LETTER

The importance of the timescale of the fitness metric for estimates of selection on phenotypic traits during a period of demographic change

Katherine Scranton,[1]* Virpi Lummaa[2] and Stephen C. Stearns[1]

[1]*Department of Ecology and Evolutionary Biology, Yale University, 165 Prospect Street CT 06520-8102, New Haven, CT, USA*

[2]*Department of Biology, University of Turku, FI-20014 Turku, Finland*

Correspondence: E-mail: katherine.scranton@ucla.edu

Abstract

Although fitness is central to the evolutionary process, metrics vary by timescale. Different timescales may give rise to different estimates of selection, especially during demographic transitions caused by rapid environmental and socioeconomic change. In this study, we used a dataset of a human population in Finland from 1775 to 1950 to compare two fitness metrics and their estimates of selection pressures, before and during a demographic transition. Both metrics, lifetime reproductive success and an annual metric of individual performance, declined while selection on the ages at first and last reproduction remained nearly constant, favouring individuals with wider reproductive windows. The ability to partition the annual metric into contributions from reproduction and survival revealed the short-term effects of a famine and the reversal of selection pressure via the survival component of annual fitness. Although the metrics generally agreed, the annual metric detected the effects of environmental variation and demographic change occurring within a generation.

Keywords

Demographic change, fitness, fitness timescale, life-history traits, phenotypic selection.

Ecology Letters (2016)

INTRODUCTION

The concept of fitness is at the core of our understanding of evolutionary change. In general, fitness measures an individual's genetic contribution to a future population. However, there is no single method for defining and calculating fitness; approaches are very different when considering a gene or allele vs. a phenotype. For phenotypic traits, fitness definitions and metrics differ at the individual vs. population level (McGraw & Caswell 1996; Link *et al*. 2002) and by timescale (Mylius & Diekmann 1995; Coulson *et al*. 2006). Metrics may also be affected by intraspecific variation in generation time (Kruuk *et al*. 1999; Brommer *et al*. 2004). The timescale over which fitness is measured is especially important for populations in variable environments with changing or variable demographic rates (Mylius & Diekmann 1995; Coulson *et al*. 2006). Long-lived species with overlapping generations can be affected by variation within a generation and datasets that capture multiple generations are rare (Benton & Grant 2000). Metrics are rarely compared across timescales with respect to the predictions of selection gradients on a trait. Here, we evaluate two metrics of individual fitness measured over different timescales for their ability to detect trends in selection on phenotypic traits during a demographic transition in a human population.

Fitness is commonly measured empirically with a generation-based metric that can be clearly defined, estimated with little error, and cleanly separates the fitness of individuals in one generation from any differences in their offspring in response to selection (Arnold & Wade 1984). However, estimates of fitness using a generational metric such as lifetime reproductive success can be sensitive to variation in generation time (Brommer *et al*. 2004) and to environmental variation within a single generation (Benton & Grant 2000; Coulson *et al*. 2003). Annual metrics have been developed as an alternative to generational metrics to better represent the variation present within a generation in populations with overlapping generations, where the contributions of different year-cohorts to the population in any subsequent year will vary (McGraw & Caswell 1996; Benton & Grant 2000; Lande *et al*. 2003; Brommer *et al*. 2004; Coulson *et al*. 2006; Courtiol *et al*. 2013). A few studies have compared the two types of metrics (Korpelainen 2003; Coulson *et al*. 2006), but none have tested the predictions of the strength of selection over periods of dramatic and rapid change such as the human demographic transition. This gap in our ability to estimate selection on phenotypic traits, regardless of fitness metric represents a serious limitation to future studies of selection over short time periods or time periods with rapid demographic change.

Modern human societies have recently experienced demographic transitions marked by shifts in environmental pressures due to advances in medicine, hygiene, and food availability and corresponding declines in childhood mortality and fertility rates (Korpelainen 2003; Moorad 2013). Though declines in demographic rates have been hypothesised to weaken selection, shifting environments can alter the strength of selection in complex ways. The underlying environmental changes can directly affect the contribution to fitness of phenotypic traits. Changes in demographic rates also alter the opportunity for selection: the variance of relative fitness and the variance in the fitness contributions from mortality or fertility (Korpelainen 2003; Courtiol *et al*. 2013; Moorad 2013).

Selection on phenotypic traits is ongoing in contemporary human populations. Many human populations have high variation in reproductive success (Brown *et al.* 2009) and key life-history traits such as age at first and last reproduction and interbirth interval have high heritability (Pettay *et al.* 2005; Stearns *et al.* 2010; Milot *et al.* 2011; Vink *et al.* 2012; Bolund *et al.* 2015). Studies of human populations that have experienced a demographic transition have found continued selection on age at first and last reproduction and age-specific fertility and survival (Byars *et al.* 2010; Milot *et al.* 2011; Moorad 2013). However, there is also evidence that selection on some traits may change direction during a demographic transition (Courtiol *et al.* 2013). Consequently, it is likely that human life-history traits are under selection during a demographic transition, but the relative ability of different fitness metrics to estimate the strength of that selection is unknown.

In this study, we evaluate two metrics of individual fitness that operate on different timescales for their ability to detect secular trends in selection on two key life-history traits during the demographic transition in a human population. The first goal is to determine whether changes in fertility and survival are associated with secular trends in the mean and variance of each fitness metric and whether our conclusions depend on the timescale at which fitness is measured. The two fitness metrics we compare are lifetime reproductive success (LRS) and an annual metric (W) related to individual performance. Our second goal is to use estimates of the selection gradients on individuals' ages at first and last reproduction (AFR, ALR) to determine whether selective pressures change across the demographic transition. To do so, we contrast the two fitness metrics, comparing overall strength of selection and the trends in selection through time to see whether our conclusions depend on the method of calculating fitness.

MATERIALS AND METHODS

Dataset

We used demographic data from human populations throughout Finland to characterise the demographic transition. Demographic transitions in contemporary human populations are characterised by shifts from high fertility rates and high mortality to low fertility and low mortality, due to economic and industrial growth and the cultural shifts that follow (Korpelainen 2003). The demographic transition in Finland began between 1870 and 1880 (Korpelainen 2003; Bolund *et al.* 2015). We used the subset of data on individuals alive between 1775 and 1950 to capture the trends in selection before and during the demographic transition. We used data from inland parishes (Ikaalinen, Jaakkima, Rautu, Pulkkila and Tyrvää) and coastal parishes (Kustavi, Rymättylä, Hiittinen). Lutheran churches were required by law to track births, deaths and marriages in each parish, and we used these records to create pedigrees and life histories for a total of 82,961 individuals from the 15th through 21st century (Bolund *et al* 2015). We restricted our analysis to the 39,773 individuals who lived between 1775 and 1950 with complete life histories. We only used individuals who were observed for the complete reproductive period (to age 50) to avoid incomplete reproductive histories of those people who emigrated or dispersed to untracked parishes in Finland. We calculated annual birth rates and survival rates for adults and children (under 16 years) for the 175 year time period.

Calculating fitness metrics

We used two metrics of fitness in our analyses: lifetime reproductive success LRS and annual fitness W. We measured LRS as the total number of offspring per individual. We calculated relative LRS by standardising within a birth cohort (subtracting the cohort mean and dividing by the cohort standard deviation) in order to address the changes in cohort mean and variance over time.

To calculate the annual metric W, we followed Courtiol *et al.* (2013) and partitioned the population by year, so that each row is a 'person-year' corresponding to each person alive in that year. We estimated an individual's annual total fitness ($W_{\mathrm{y,total}}$) as its genetic contribution to the population in the next year: the sum of the contribution from survival (1 for surviving from year y to $y + 1$, 0 otherwise) and from reproduction (0.5 for each offspring produced in year y). In this way, there were several discrete scores possible: 0 for dying without reproducing, 0.5 for producing one offspring and dying, 1 for surviving without reproducing, 1.5 for surviving and producing one offspring, 2 for surviving and producing two offspring, and so on for each additional birth. If no individuals in the population had multiple births, there would be only four categories, but births of twins or triplets allowed higher annual fitness values. We parsed the metric into the contribution to each individual's fitness from reproduction ($W_{\mathrm{fertility}}$) and survival (W_{survival}). We also analysed childhood fitness using the individual total fitness from children under the age of 16, excluding adults (W_{child}).

We calculated the relative total fitness of an individual i within a given year y as $w_{total,i,y} = \frac{W_{total,i,y}}{W_{total,y}}$. We similarly calculated the relative fitness contributions from reproduction $w_{\mathrm{fertility}}$ and childhood survival w_{child}. Relative annual fitness is the more appropriate metric for studies of selection as a trait confers an advantage to an individual only as compared to the rest of the population. The variance in relative annual fitness w_{total} represents the 'opportunity for selection' I, which summarises the total capacity for adaptation: the total selection on all possible traits (Moorad 2013). We calculated the variance of relative total fitness and each of its components and estimated the errors by bootstrapping annual populations.

Estimating selection coefficients

We estimated the strength of selection on AFR and ALR in each year using relative cohort LRS, relative annual fitness w_{total} and the components $w_{\mathrm{fertility}}$, w_{survival}. Analysing age at first and last reproduction necessarily restricts our data further to those individuals who reproduced; only those individuals have a recorded AFR or ALR trait. This leaves 18,362 individuals in the total population and 17,685 individuals in annual birth cohorts. We used the annual estimates of the selection gradients to determine whether selective pressures changed across the demographic transition.

We performed basic linear regressions using the relative LRS data of each year's birth cohort to estimate selection coefficients for age at first reproduction (AFR) and age at last reproduction (ALR). We used a linear model (lm in R) while controlling for three additional fixed effects: gender, lineage parish as an 8-level factor and socioeconomic status, as estimated from an individual's occupation (Pettay *et al.* 2007). Individuals were grouped into one of two categories: landed and landless, to create an indicator of socioeconomic status based on property that is known to associate with key life-history traits and fitness in the study population (Courtiol *et al.* 2012; Liu *et al.* 2012). 'Landed' individuals were those who either owned or rented property (or were married to such an individual) and 'landless' were those without property, who were likely servants or itinerant poor. We fit this full model as well as all of the limited models with any of the five fixed effects, without including any interactions. This created 32 possible models, including the null model of equal fitness. We fit all models and ranked them by AIC. We selected the model with the lowest AIC value, unless the full model's AIC is within Δ2 of the minimum value, in which case we took the full model. If the model with the lowest AIC value did not have AFR or ALR as a factor, we considered the strength of selection on that trait to be zero. We followed the same procedure for each year's cohort to create a dataset of independent annual regression coefficients.

We then estimated a linear model for the response of regression coefficients on AFR and ALR over time. We used the estimated selection coefficient (or zero) regardless of p-value, but substituting zeros for coefficients with $P > 0.05$ did not qualitatively change our results. This yielded slopes β_{AFR} and β_{ALR} that represent the change in the strength of selection over time.

We similarly performed regressions in each year using relative annual fitness (w_{total}) to determine the strength of selection on AFR and ALR. However, the relative fitness values within a year can take one of several discrete, ordered values, necessitating an ordinal regression without the proportional odds assumption using a cumulative link model with a probit link. We performed analyses in R with the package *ordinal* (Christensen 2015). In order to extract information specifically on the contributions to fitness from fertility and survival, we performed binomial regressions using the binary data of individual reproduction $w_{\text{fertility}}$ and survival w_{survival}. We could not isolate childhood survival in this analysis because this subset of data included only individuals who had a trait value for AFR and ALR, excluding those individuals that did not survive childhood. As in the LRS analysis, we fit all 32 models including the full model with five fixed effects and all nested limited models (excluding interactions). We performed model selection by comparing AIC values and extracted the regression coefficients on AFR and ALR in the same way as above.

To determine if there were any trends in selection over time, we would have liked to perform a linear regression of selection coefficients on time, but the annual populations are not independent. There was in fact a moderate to high degree of autocorrelation (lag 1) in the time series' of selection coefficients (on ALR as measured by relative annual fitness: 0.646, by the contribution from reproduction: 0.585, and the contribution from survival: 0.444, and on AFR as measured by relative annual fitness: 0.712, by the contribution from reproduction: 0.636). The selection coefficients on AFR as measured by the contribution from survival showed almost no autocorrelation (0.0455) but we proceeded with the analyses, as we know the data points are not independent.

To account for the non-independence, we estimated the trend in selection coefficients in two ways: with a corrected Mann-Kendall test and with Spearman's correlation coefficient. We used two approaches because the analyses account for autocorrelation in fundamentally different ways and have slightly different null hypotheses. First, we performed a corrected Mann-Kendall test for a trend in a time series that accounts for autocorrelation (Hamed & Rao 1998), using the R package *fume*. This test relaxes the assumption that the differences between pairs of values in a time series are independent of all other pairs, making it appropriate for any autocorrelation structure in a dataset (Hamed & Rao 1998). It tests against the null hypotheses that the data are randomly ordered in time, specifically the population-level associations between trait and fitness are unordered in time.

We also estimated trends by calculating Spearman's nonparametric correlation coefficient ρ between each time series of selection coefficients and time. In order to account for autocorrelation, we created a null distribution of correlation coefficients that explicitly recreated the source of autocorrelation. Autocorrelation in this dataset comes from the fact that populations in subsequent years are comprised of many of the same individuals (with a few births and deaths). We specified a null hypothesis that individuals are randomly ordered in time, or the individual-level associations between trait and fitness are unordered in time. We created a null distribution by simulating datasets by bootstrap resampling on individuals, then shuffling birthdates without replacement, realigning each individual in the 175 year window. This breaks the correlations of individuals' traits and fitness with time, but leaves the sequence of each individual's annual fitness values and the subsequent autocorrelation intact. We created 1000 replicate bootstrapped datasets in this way, ran the annual regressions to estimate the selection coefficients, and recalculated Spearman's correlation coefficient between selection gradients and time. We compared the correlation coefficient ρ calculated from the original dataset to this set of 1000 ρ_0 values simulated under the null hypothesis.

RESULTS

The shifts in the reproductive and survival rates around 1870 correspond to the beginning of the demographic transition (Fig. 1a). The annual per capita birth rate decreased by more than half from 0.094 in the 1870′s to 0.037 in 1950 (Fig. 1a dashed line). This would correspond to a decline from *c.* nine births each year for every 100 people in the population to about four births per 100 people. Childhood mortality also decreased from 4.9% in the 1870′s to 0.7% by 1950 (Fig. 1a solid line). Compounded, these annual rates would correspond to a cumulative childhood survival rate (survival from birth until the age of 15) of 47% before the transition and 90% after the transition. The famine of 1866–1868 (Hayward *et al.* 2013) had drastic effects on the population, causing childhood

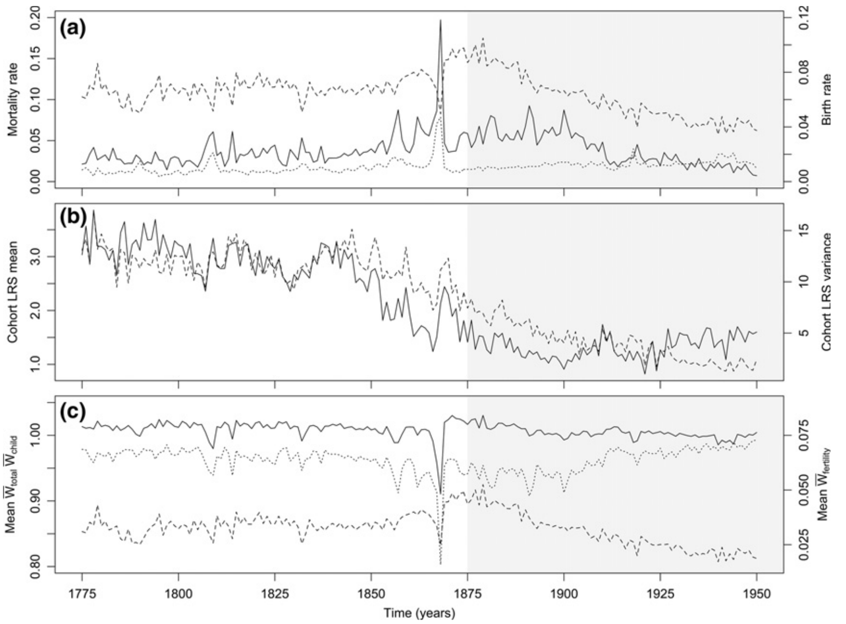

Figure 1 Demographic rates and fitness metrics for the sampled population in Finland 1775–1950. (a) Annual rates of birth (dashed line), adult mortality (dotted line) and childhood mortality (solid line), (b) the mean (solid line) and variance (dashed line) of LRS of the cohort of individuals born in each year, (c) mean annual fitness ($\overline{W}_{total}$, solid line) and the contribution from reproduction ($\overline{W}_{fertility}$, dashed line) and from childhood survival ($\overline{W}_{child}$, dotted line). All rates and metrics were estimated from the population of 39 773 individuals with fully tracked life histories. The shaded grey area shows the time period since the start of the demographic transition.

mortality to increase four-fold (from 5.9% in 1864 to 19.7% in 1868, Fig. 1a) and per capita birth rate to decrease by about a third (from 0.079 in 1864 to 0.050 in 1868, Fig. 1a).

Shifting demographic rates were accompanied by decreases in fitness, as measured by both LRS and annual fitness W. The mean LRS of annual birth cohorts started to decline just before 1850 (Fig. 1b solid line), corresponding to those individuals who were reproductively active at the beginning of the demographic transition. Variance in LRS within each birth cohort also declined with a similar pattern to mean cohort LRS (Fig. 1b dashed line). Mean annual fitness of the population ($\overline{W}_{total}$) was fairly constant before the demographic transition and declined slightly from 1.023 in the 1870′s to 1.000 by 1950, with a sharp drop during the famine (Fig. 1a dotted line). This slight secular decline in annual fitness ($\overline{W}_{total}$) represents a change from a growing population to a population in stasis.

The components of the annual metric revealed compensatory changes in the contributions to annual fitness from fertility and childhood survival. Decreases in $W_{\mathrm{fertility}}$ were almost balanced by increases in W_{child}, reflecting the demographic transition in the population. The mean contribution to fitness from reproduction ($\overline{W}_{fertility}$) declined from around 0.047 in the 1870s to 0.019 in 1950, mirroring the decline in per capita birth rate (Fig. 1c dashed line). The mean contribution to fitness from childhood survival ($\overline{W}_{child}$) increased from 0.95 in the 1870′s to 0.99 in 1950 (Fig. 1c dotted line), showing the effects of decreased childhood mortality. The famine was too brief to leave a signature in the generational metric LRS, but was reflected in the annual metric.

The variance of relative annual fitness or opportunity for selection declined during the demographic transition for relative annual fitness w_{total} (Fig. 2a). The decline in the variance of relative annual fitness was dominated by a similar decline in the contribution from childhood survival w_{child} (Fig. 2b). The variance in the contribution to relative fitness from fertility $w_{\mathrm{fertility}}$ increased during the demographic transition (Fig. 2c), but that trend reflects the calculation of $w_{\mathrm{fertility}}$. The variance of relative fitness was calculated as the variance of absolute fitness divided by the mean of absolute fitness, which each declined at different rates. This slight difference caused an observed increase in relative annual fitness, consistent with similar analyses of this annual metric (Courtiol *et al.* 2013).

Both fitness metrics revealed negative selection on AFR and positive selection on ALR, favouring individuals with wider reproductive windows (Fig. 3). Mean AFR and ALR were 27.2 years and 36.9 years, respectively, although both were highly variable within cohorts. Selection as measured by relative cohort LRS increased slightly through time for both traits with significant, but very small effect sizes (Fig. 3a, $\beta_{AFR} = -3 \times 10^{-4}$, $P = 1.9 \times 10^{-18}$, $\beta_{ALR} = 2.3 \times 10^{-4}$, $P = 9.8 \times 10^{-13}$).

We used two methods to assess the trends through time in selection as measured by annual fitness and its components. The two methods were consistent, always estimating the same direction of trend (increasing or decreasing). The strength of the relationship was also consistent between analyses, but the significance level differed slightly. Selection on AFR measured by relative annual total fitness weakened through time (Fig. 3b). The trend was significant for selection on AFR through the Mann-Kendall test ($Z_{c,AFR} = 3.651$, $P = 2.61 \times 10^{-4}$), but not as measured by correlation coefficients ($\rho_{AFR} = 0.68$, $P = 0.15$). Selection on ALR remained

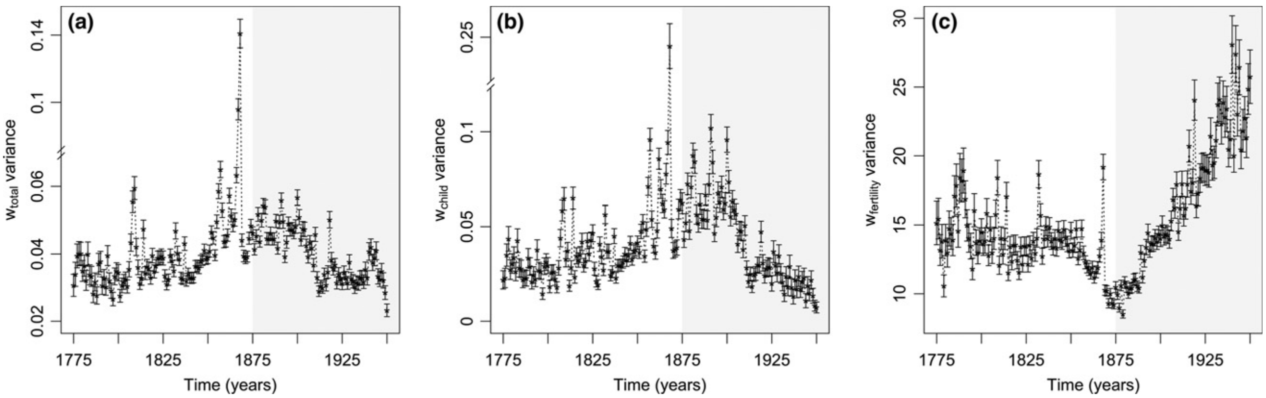

Figure 2 Variances of (a) relative annual fitness w_{total} and its components (b) w_{child} and c) $w_{\text{fertility}}$. The grey regions denote the time period since the start of the demographic transition. Error bars are ± 1 standard deviation of the estimated variances in fitness for 1000 replicate bootstrapped samples of the annual populations.

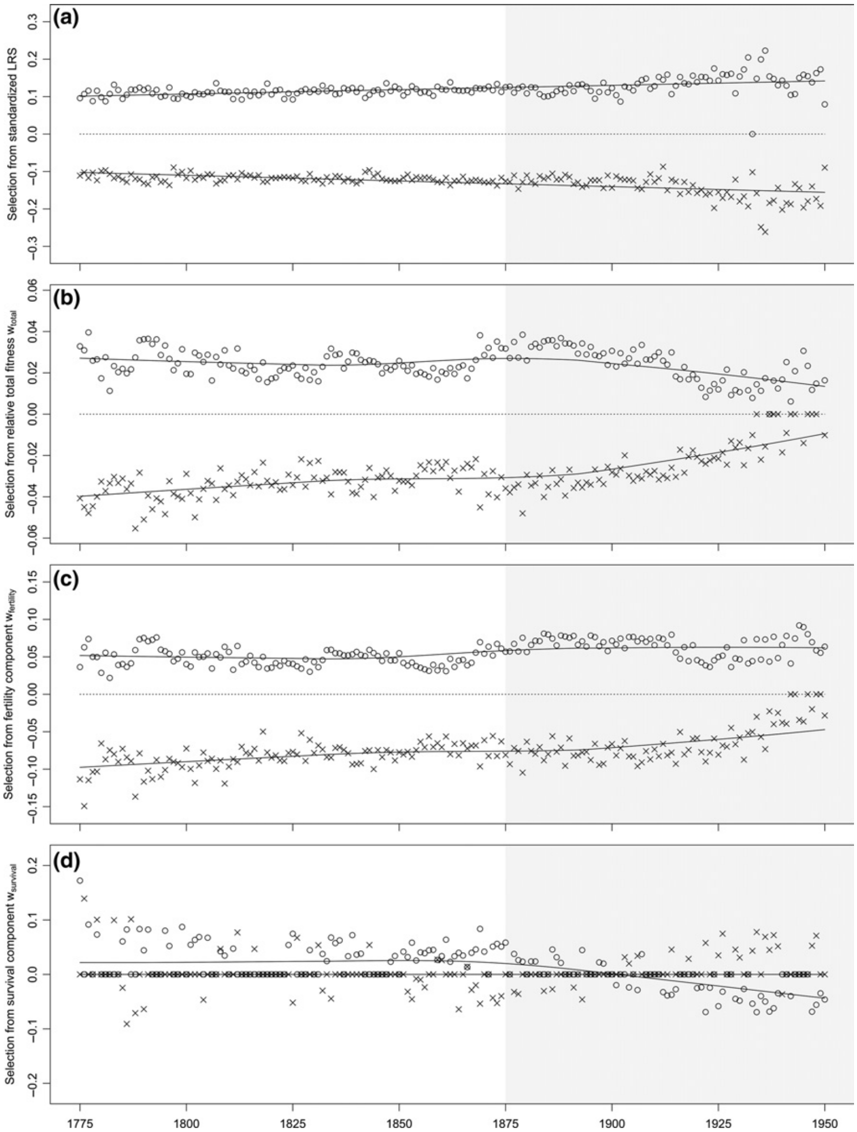

Figure 3 Annual selection coefficients on AFR (x) and ALR (o) as measured by each fitness metric. We estimated selection coefficients using (a) cohort LRS, (b) relative total fitness w_{total}, (c) the contribution to relative fitness from reproduction $w_{\text{fertility}}$, (d) the contribution to relative fitness from survival w_{survival}. For each fitness metric, the relationship with time is given by a line fitted by a regression (a) or by a LOWESS smoother (b,c and d). The shaded grey area shows the time period since the start of the demographic transition.

consistently positive through time ($\rho_{ALR} = -0.24$, $P = 0.061$, $Z_{c,ALR} = -1.165$, $P = 0.244$).

Selection as measured by the contribution to fitness from reproduction showed similar trends as selection as measured by relative annual total fitness (Fig. 3c). Selection weakened through time for AFR, significantly according to the Mann-Kendall test ($Z_{c,AFR} = 3.177$, $P = 1.49 \times 10^{-3}$) and non-significantly for Spearman's correlation coefficient ($\rho_{AFR} = 0.58$, $P = 0.099$). Selection on ALR as measured through reproduction remained positive throughout the study period, but did show a slight trend towards weaker positive selection ($\rho_{ALR} = 0.37$, $P = 0.26$, $Z_{c,ALR} = 2.05$, $P = 0.041$).

However, selection as measured by the contribution to fitness from survival was markedly different than selection as measured by any other metric (Fig. 3d). In many years, we estimated zero selection on AFR (118 of 175 years) and ALR (80 of 175 years). Overall, we found a highly significant reversal of selection on ALR through survival differences, moving from positive selection before the demographic transition to negative selection by 1950 ($\rho_{ALR} = -0.52$, $P < 0.001$, $Z_{c,ALR} = -3.609$, $P = 3.07 \times 10^{-4}$). Selection on AFR also reversed, moving from negative to positive, but with a very weak trend, non-significant in the Mann-Kendall test ($\rho_{AFR} = 0.15$, $P = 0.001$, $Z_{c,AFR} = 0.855$, $P = 0.393$). By the end of the study period, the population experienced positive selection on AFR and negative selection on ALR, meaning individuals with narrower reproductive windows were more likely to contribute to the population in the next year by surviving.

DISCUSSION

We have shown the advantages of using an annual metric of fitness to elucidate the effects of environmental variation and demographic change that occur within a generation. The effects of a short, severe famine were reflected in fitness as measured by the annual metric but not by the generational metric LRS. The components of the annual metric, the contribution from reproduction and childhood survival, revealed that compensatory changes in fitness occurred during the demographic transition. Both metrics consistently estimated negative selection on AFR and positive selection on ALR, favouring wide reproductive windows. The annual metric and LRS produced qualitatively similar patterns of selection through the demographic transition. However, the ability to break down the annual metric into survival and reproductive components revealed a reversal of selection as measured by adult survival.

The main advantage of the annual metric is its ability to capture temporal variation or secular change in selection within a generation. The ability to estimate a time series of selection pressures allows us to quantify the trends in selection and their association with environmental processes (Kingsolver *et al.* 2001; Coulson *et al.* 2003; Courtiol *et al.* 2013). Our results are consistent with other evidence for directional selection on life-history traits (Kingsolver *et al.* 2012). The direction of selection on the traits in our study is also constant, with very weak trends in intensity through time. This is of interest in the light that also the genetic structure and heritability of those traits were relatively stable across the demographic transition (Bolund *et al* 2015). We do see some temporal variation in strength of selection, particularly in selection as estimated by the survival component of annual fitness and at the end of the study period in selection as estimated by LRS. As shown in previous reviews and meta-analyses (Kingsolver *et al.* 2012; Morrissey & Hadfield 2012), sampling error may contribute to the variation in selection as estimated by LRS: our sample size decreases after about 1925 due to the difficulty of tracking individuals as many emigrated from Finland.

The demographic transition in recent human populations offers the most dramatic and rapid shift in mortality and fertility observed in any species (Bolund *et al.* 2015), and thus an excellent opportunity to test how these two approaches to measuring fitness perform under changing environmental conditions. Both metrics estimated the ongoing selection on human life-history traits during the demographic transition, an interpretation that has been widely debated (Smith *et al.* 2001). The congruence of estimates produced by different metrics has not been tested before in such rapidly changing conditions as this human dataset offers. Although we use this human demographic transition to illustrate the similarities and differences in the two metrics, our results would apply to the many ecosystems in which species experience considerable fluctuations in selection pressures over time (Bell 2010).

The ability to estimate multiple snapshots of selection over short time periods also allows us to identify dramatic changes in fitness that may be of short duration. Although we used the famine of 1866–1868 (Hayward *et al.* 2013) as an example of dramatic, short-lived environmental conditions, the metrics would perform similarly in any swift, dramatic change, such as a pulsed resource input (Holt 2008). Using an annual metric will also allow future studies to use shorter datasets to assess the ongoing selection for populations of long-lived species; long-lived species with overlapping generations require very long datasets to produce reliable estimates of fitness and selective pressures using LRS (Grafen 1988). In fact, the metric could be modified to estimate fitness as the contribution to the population over any time interval (a day or an hour instead of a year) for a species with a shorter generation time.

Another advantage of using annual fitness is the ability to decompose it into contributions from reproduction and survival. The ability to combine the effects of selection through reproduction and survival allows a more nuanced analysis of selection, which can operate through either (or both) life-history processes (Lande 1982). Recent extensions of stage structured models to explicitly allow trait evolution have allowed exploration of the differential contributions of survival and reproduction to fitness, selection intensity, phenotype mean and variance (Coulson *et al.* 2010; Barfield *et al.* 2011; Steiner *et al.* 2014). Theoretical studies of eco-evolutionary dynamics have shown that selection on a phenotype can act through both survival and reproduction, in opposite directions in some cases (Rees & Ellner 2016). Especially in temporally variable environments, the ability to partition fitness into demographic components can be essential (Horvitz *et al.* 2010).

Both fitness metrics produced similar qualitative trends of selection, overall, estimating negative selection on AFR and positive selection on ALR. However, although our dataset spanned an extraordinary length of time, only a few traits were recorded. Other unmeasured traits might be correlated with AFR and ALR, creating tradeoffs of which we are unaware (Stearns 1989). It is not surprising that there were strong similarities in the trends of selection estimated by both metrics since they measure the same fundamental processes of reproduction and survival in the same population.

However, there were slight differences in the secular trends of the strength of selection. Selection as measured by relative LRS did significantly strengthen through time, but with very small effect sizes and with increased residual variance. In contrast, selection as measured by relative annual fitness showed signs of weakening selection that varied in the significance compared to a null hypothesis. One reason for the differences in the analyses is that LRS fully describes reproduction, but only implicitly captures survival: individuals with more offspring must live long enough to reproduce repeatedly. Variation in longevity and in the timing of reproduction caused the discrepancies between the fitness estimates of the two metrics.

There was also a lag time between the response from the annual metric and from LRS because they represent fundamentally different populations. In any year, the annual metric summarises the survival and reproductive output of all individuals alive in that year, while LRS summarises future reproductive output of individuals born in that year. The same individuals contribute to the metrics in different years: to LRS when they are born and to the annual metric throughout their lifetime. The analyses of the annual metric that account for autocorrelation differed in the significance level of the trends, reflecting the fundamental differences in the two tests: the Mann-Kendall test applies a general model of autocorrelation structure, whereas the null distribution of ρ values recreates the underlying source of autocorrelation in the data. The differences in significance level could also be due to the null hypotheses of random ordering of trait-fitness associations at the population vs. individual level.

Resolving the annual fitness metric into its contributions from survival and reproduction showed that selection as measured by the contribution to fitness from reproduction produced patterns of selection coefficients similar to those produced by LRS. However, selection as measured by the contribution to fitness from adult survival was markedly different, reversing direction on both AFR and ALR. We do note that estimated selection coefficients via the survival component were variable throughout the study period. In many years, we estimated zero selection via annual survival on AFR (118 of 175 years) and ALR (80 of 175 years), indicating that the reproductive window was not always linked to annual survival. In spite of periods of negligible selection via survival, the positive correlations between annual survival and early age at first reproduction and late age at last reproduction reversed during the demographic transition to become negative. The comparisons between the estimates of selection made via reproduction and survival highlight the fact that selection may be acting through both life-history processes simultaneously (Lande 1982) with variation in the relative importance of each component (Coulson *et al.* 2003).

Our results agree with previous studies that have compared fitness metrics on different timescales. An analysis of another dataset of the human population in Finland found similar patterns of declining LRS and steady individual fitness (measured as λ) during the demographic transition, but did not compare estimates of selection on any trait (Korpelainen 2003). Using a 'de-lifing' approach to estimate annual fitness and selection on birth weight in Soay sheep, Coulson *et al.* (2006) found that an estimate of selection using LRS fell close to the mean of selection coefficients estimated with their annual metric. Our results mirror the qualitative similarities between metrics and the ability of the annual metric to pick up variation smoothed by LRS. Our results also extend the analyses to comparing trends of selection over periods of demographic change. The similarities between previous studies and aspects of our results lend confidence to the use of the annual metric in future studies.

In conclusion, we found that both metrics revealed ongoing, relatively constant selection before and during the demographic transition, favouring individuals who begin reproducing earlier and those who cease reproduction later in life, providing more evidence for the importance of selection in contemporary human populations. The annual metric of fitness yielded the same qualitative conclusions about the direction of selection operating on a human population across the demographic transition as lifetime reproductive success, lending confidence to the use of such an annual metric. Further, decomposing the annual metric into components contributed by survival and reproduction revealed details about the ways selection acts through different life-history processes, especially important for the unparalleled rapid change in vital rates during demographic transitions. The annual metric was also better able to detect the effects of the short dramatic change in environmental conditions during the famine. Our results encourage the use and application of the annual metric in other ecosystems with any species, especially when radically shifting environmental conditions cause swift changes in demographic rates.

ACKNOWLEDGEMENTS

This work was supported in part by the facilities and staff of the Yale University Faculty of Arts and Sciences High Performance Computing Center and by Academy of Finland and European Research Council grants to VL.

AUTHOR CONTRIBUTIONS

SS developed the motivation and main questions with substantial input from KS and VL, VL contributed the dataset, KS performed all of the analyses and wrote the manuscript, with substantial revisions from SS and VL.

REFERENCES

Arnold, S.J. & Wade, M.J. (1984). On the measurement of natural and sexual selection: theory. *Evolution*, 38, 709–719.

Barfield, M., Holt, R.D. & Gomulkiewicz, R. (2011). Evolution in stage-structured populations. *Am. Nat.*, 177, 397–409.

Bell, G. (2010). Fluctuating selection: the perpetual renewal of adaptation in variable environments. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 365, 87–97.

Benton, T.G. & Grant, A. (2000). Evolutionary fitness in ecology: comparing measures of fitness in stochastic, density-dependent environments. *Evol. Ecol. Res*., 2, 769–789.

Bolund, E., Hayward, A., Pettay, J.E. & Lummaa, V. (2015). Effects of the demographic transition on the genetic variances and covariances of human life-history traits. *Evolution*, 69, 747–755.

Brommer, J., Gustafsson, L., Pieti¨ainen, H. & Meril¨a, J. (2004). Single Generation Estimates of Individual Fitness as Proxies for Long Term Genetic Contribution. *Am. Nat*. 163:505–517.

Brown, G.R., Laland, K.N. & Mulder, M.B. (2009). Bateman's principles and human sex roles. *Trends Ecol. Evol*., 24, 297–304.

Byars, S.G., Ewbank, D., Govindaraju, D.R. & Stearns, S.C. (2010). Natural selection in a contemporary human population. *Proc. Natl Acad. Sci. USA*, 107, 1787–1792.

Christensen, R.H.B., 2015. ordinal—regression models for ordinal data. R package version 2015.6-28. Available at: http://www.cran.r-project.org/package=ordinal/. Last accessed 25 June 2015.

Coulson, T., Kruuk, L.E.B., Tavecchia, G., Pemberton, J.M. & Clutton-Brock, T.H. (2003). Estimating selection on neonatal traits in red deer using elasticity path analysis. *Evolution*, 57, 2879–2892.

Coulson, T., Benton, T.G., Lundberg, P., Dall, S.R.X., Kendall, B.E. & Gaillard, J.M. (2006). Estimating individual contributions to population growth: evolutionary fitness in ecological time. *Proc. R. Soc. B Biol. Sci*., 273, 547–555.

Coulson, T., Tuljapurkar, S. & Childs, D.Z. (2010). Using evolutionary demography to link life history theory, quantitative genetics and population ecology. *J. Anim. Ecol*., 79, 1226–1240.

Courtiol, A., Pettay, J.E., Jokela, M., Rotkirch, A. & Lummaa, V. (2012). Natural and sexual selection in a monogamous historical human population. *Proc. Natl Acad. Sci*., 109, 8044–8049.

Courtiol, A., Rickard, I., Lummaa, V., Prentice, A., Fulford, A.J.C. & Stearns, S. (2013). The demographic transition influences variance in fitness and selection on height and BMI in rural gambia. *Curr. Biol*., 23, 884–889.

Grafen, A., 1988. On the uses of data on lifetime reproductive success. In: *Reproductive Success: Studies of Individual Variation in Contrasting Breeding Systems* (ed Clutton-Brock, T.H.). University of Chicago Press, Chicago, IL, USA, pp. 454–471.

Hamed, K.H. & Rao, A.R. (1998). A modified Mann-Kendall trend test for autocorrelated data. *J. Hydrol*., 204, 182–196.

Hayward, A.D., Rickard, I.J. & Lummaa, V. (2013). Influence of early-life nutrition on mortality and reproductive success during a subsequent famine in a preindustrial population. *Proc. Natl Acad. Sci. USA*, 110, 13886–13891.

Holt, R.D. (2008). Theoretical Perspectives on Resource Pulses. *Ecology*, 89, 671–681.

Horvitz, C.C., Coulson, T., Tuljapurkar, S. & Schemske, D.W. (2010). A new way to integrate selection when both demography and selection gradients vary over time. *Int. J. Plant Sci*., 171, 945–959.

Kingsolver, J.G., Hoekstra, H.E., Hoekstra, J.M., Berrigan, D., Vignieri, S.N., Hill, C.E. *et al*. (2001). The strength of phenotypic selection in natural populations. *Am. Nat*., 157, 245–261.

Kingsolver, J.G., Diamond, S.E., Siepielski, A.M. & Carlson, S.M. (2012). Synthetic analyses of phenotypic selection in natural populations: lessons, limitations and future directions. *Evol. Ecol*., 26, 1101–1118.

Korpelainen, H. (2003). Human life histories and the demographic transition: a case study from Finland, 1870-1949. *Am. J. Phys. Anthropol*., 120, 384–390.

Kruuk, L.E.B., Clutton-Brock, T.H., Rose, K.E. & Guinness, F.E. (1999). Early determinants of lifetime reproductive success differ between the sexes in red deer. *Proc. R. Soc. B Biol. Sci*., 266, 1655–1661.

Lande, R. (1982). A quantitative genetic theory of life history evolution. *Ecology*, 63, 607–615.

Lande, R., Engen, S. & Sæther, B.-E., 2003. *Stochastic Population Dynamics in Ecology and Conservation*. Oxford University Press, Oxford, UK.

Link, W.A., Cooch, E.G. & Cam, E. (2002). Model-based estimation of individual fitness. *J. App. Stat*., 29, 207–224.

Liu, J., Rotkirch, A. & Lummaa, V. (2012). Maternal risk of breeding failure remained low throughout the demographic transitions in fertility and age at first reproduction in Finland. *PLoS ONE*, 7, e34898.

McGraw, J.B. & Caswell, H. (1996). Estimation of individual fitness from life-history data. *Am. Nat*., 147, 47–64.

Milot, E., Mayer, F.M., Nussey, D.H., Boisvert, M., Pelletier, F. & Reale, D. (2011). Evidence for evolution in response to natural selection in a contemporary human population. *Proc. Natl Acad. Sci. USA*, 108, 17040–17045.

Moorad, J.A. (2013). A demographic transition altered the strength of selection for fitness and age-specific survival and fertility in a 19th century American population. *Evolution*, 67, 1622–1634.

Morrissey, M.B. & Hadfield, J.D. (2012). Directional selection in temporally replicated studies is remarkably consistent. *Evolution*, 66, 435–442.

Mylius, S.D. & Diekmann, O. (1995). On evolutionarily stable life histories, optimization and the need to be specific about density dependence. *Oikos*, 74, 218–224.

Pettay, J.E., Kruuk, L.E.B., Jokela, J. & Lummaa, V. (2005). Heritability and genetic constraints of life-history trait evolution in preindustrial humans. *Proc. Natl Acad. Sci. USA*, 102, 2838–2843.

Pettay, J.E., Helle, S., Jokela, J. & Lummaa, V. (2007). Natural selection on female life-history traits in relation to socio-economic class in pre-industrial human populations. *PLoS ONE*, 2, e606.

Rees, M. & Ellner, S.P. (2016). Evolving integral projection models: evolutionary demography meets eco-evolutionary dynamics. *Methods Ecol. Evol*., 7, 157–170.

Smith, E.A., Mulder, M.B. & Hill, K. (2001). Controversies in the evolutionary social sciences: a guide for the perplexed. *Trends Ecol. Evol*., 16, 128–135.

Stearns, S. (1989). Trade-offs in life-history evolution. *Funct. Ecol*., 3, 259–268.

Stearns, S.C., Byars, S.G., Govindaraju, D.R. & Ewbank, D. (2010). Measuring selection in contemporary human populations. *Nat. Rev. Genet*., 11, 611–622.

Steiner, U.K., Tuljapurkar, S. & Coulson, T. (2014). Generation time, net reproductive rate, and growth in stage-age-structured populations. *Am. Nat*., 183, 771–783.

Vink, J.M., Bartels, M., van Beijsterveldt, T.C.E.M., van Dongen, J., van Beek, J.H.D.A., Distel, M.A. *et al*. (2012). Sex differences in genetic architecture of complex phenotypes? *PLoS ONE*, 7, e47371.

Editor, Tim Coulson
Manuscript received 10 March 2016
First decision made 12 April 2016
Manuscript accepted 21 April 2016