

1 **Supplementary Methods, Tables and Figures for:**

2

3 **Reduced costs of reproduction in females mediate a shift from a male-biased to a**
4 **female-biased lifespan in humans**

5

6 Elisabeth Bolund, Virpi Lummaa, Ken R. Smith, Heidi Hanson, Alexei A. Maklakov

7

8

9 **Supplementary Methods**

10

11 *Data*

12 The Utah Population Database (UPDB) is one of the most comprehensive computerised
13 genealogies in the world. Genealogy records are gathered and provided by the Genealogical
14 Society of Utah and span the pioneer settlement era in Utah in the mid-1800s as well as their
15 Utah descendants until the present day¹. These records include information on fertility and
16 mortality for parents and their children. Before 1890 polygamy was practiced by a minority
17 of the population². Non-paternity rates in Utah are lower than the US average³. The Utah
18 population during the period studied here suffered less mortality than most comparable
19 populations, especially at early ages⁴.

20

21 For this study, we focussed on the time period surrounding the demographic transition and
22 included individuals from the UPDB that were born between 1820 and 1920 (no individual
23 born before 1920 was still alive at the end of data coverage for this study in 2012). We
24 required full information for a number of variables that were entered as fixed and random
25 effects in the models, namely reproductive history, whether the individual was involved in a

26 polygamous marriage (4.7% of females and 1.9% of males), whether the individual was born
27 in Utah (54% of females and 52% of males), birth order in three levels denoting whether an
28 individual was the firstborn son, the firstborn daughter, or laterborn of either sex (to control
29 for effects of inherited wealth), and the identity of the birth mother (maternal identity, to
30 account for non-independence of individuals born in the same family). All individuals with
31 full information for these fixed and random effects also had known birth and death dates,
32 hence censoring was not employed in survival models. Individuals were divided into four
33 birth cohorts of 25 years each, commencing 1820. This resulted in a total sample size of
34 75,667 reproducing females and 64,933 reproducing males.

35

36 Because the dataset is built from descendant genealogies, it likely underrepresents individuals
37 that are known to not reproduce. In the present study, 1.7% of married individuals were
38 known to be nulliparous, likely representing a non-random subset of the population because
39 involuntary childlessness may be more common in frail individuals. Because of the difficulty
40 to achieve a representative subset of nulliparous individuals and our interest in the cost of
41 increased parity and the effect of changes in parity on lifespan, we therefore focussed on only
42 reproducing individuals in the main analyses. For comparison, we include Kaplan-Meier
43 survivor curves for presumably nulliparous individuals of the two sexes over the studied time
44 period in Supplementary Fig. 1. These should be interpreted with knowledge of the lowered
45 accuracy of the data in these individuals, because individuals that were registered as non-
46 reproducers may be so because of a lack of information regarding their reproduction.

47

48 *Sexual dimorphism in lifespan*

49 To look at the differences in survival patterns between the two sexes and how this changed
50 over time, we used the *survival* package in R⁵ to plot Kaplan-Meier survival curves for the two

51 sexes in four cohorts of 25 years each, starting at the year 1820. Because the interest was in
52 adult survival and we focussed on only reproducing individuals, individuals that died before
53 reproductive age are not included in the survival analyses. Hence, the first deaths occur at age
54 15. We tested if the sexes differed significantly with log rank and Wilcoxon tests in each
55 cohort (Figure 1). We then proceeded with parametric accelerated failure time (AFT) survival
56 models using the *survreg* function in the survival package. AFT models assume that the
57 effects of covariates are multiplicative with respect to survival time. Parametric models
58 require that a distribution is specified for the survival times. We compared the fit of
59 exponential, Weibull, log-normal and log-logistic models using two approaches: graphically,
60 by comparing the predicted estimates from the fitted model with the observed Kaplan-Meier
61 estimate of the survivor function, and analytically, by comparing model fits with different
62 distributions specified, using χ^2 -tests for nested models (exponential, Weibull and log-
63 logistic) and using Akaike's information criterion (AIC) for non-nested models. This showed
64 that a Weibull distribution provided the best fit. The Gompertz model was not intended to
65 study mortality patterns beyond the age of about 80 years and thus was not considered⁶.
66 Models included as fixed effects polygamy status, birth in or outside of Utah, birth order and
67 birth cohort. Sex was added as a stratified fixed effect to allow different baseline survival
68 shapes for the two sexes. Observations were clustered by maternal identity to account for
69 clustering within mothers. The interaction between sex and birth cohort indicates whether the
70 two sexes differ in each cohort. To ensure that these results are not dependant on the arbitrary
71 division of the data into 25 year birth cohorts, we repeated the analyses, first with the data
72 divided into 5 cohorts of 20 years each, and second, with birth year treated as a continuous
73 variable. Results were robust to these alternative model specifications (not shown). Mortality
74 during migration to Utah is likely to have had sex-specific effects on survival. The analyses
75 detailed above includes all individuals with known birth and death information, regardless of

76 whether they were born or died in Utah or elsewhere. Because all individuals in the first
77 cohort were born outside Utah, most of these individuals suffered the hardships of migration,
78 affecting the sexes differently and resulting in an overall more robust cohort surviving to old
79 ages. Subsequent cohorts are composed of four categories: individuals that 1) were born
80 elsewhere and migrated to Utah, 2) were born and died in Utah, 3) were born in Utah and
81 died elsewhere and, 4) were born and died outside of Utah. Cohort 2 thus offers the
82 opportunity to compare migrants into Utah with individuals that spent their entire lives in
83 Utah during a time period when migration still imposed hardships (Supplementary Fig. 1a).
84 To test for a healthy migrant effect, we used a parametric survival model that adjusted for
85 polygamy status and birth order as fixed effects, and clustered the observations by maternal
86 identity. Because the conditions during migration changed over time, we validated the use of
87 cohort 2 for this comparison by comparing migrants in cohort 2 with migrants in cohort 1
88 (Supplementary Fig. 1b) and looked at the difference in survival between migrants in these
89 two cohorts in a parametric survival model within each sex, with covariates as described
90 above.

91

92 *Reproduction-lifespan trade-off*

93 To test for a trade-off between reproduction and lifespan in the two sexes, we investigated the
94 relationship between number of children born and post-reproductive lifespan with two
95 approaches. First, we employed the R-package *MCMCglmm*, which uses an iterative
96 Bayesian approach⁷ and provides meaningful error estimates for derived variables by direct
97 sampling from the posterior distribution. Individuals were required to live until age 55 or
98 older, in order to focus on the effect of reproduction on late life mortality (100% of all
99 women and 96.8% of all men had finished reproduction by this age). This avoids
100 confounding deaths caused by childbearing (Figure 1) with post-reproductive mortality.

101 Hence, we subtracted 55 years from the total lifespan to obtain the response variable post-
102 reproductive lifespan. Our selection criteria restricted the total sample size to 118,911
103 individuals. Models controlled for a number of fixed effects as described above (polygamy
104 status, whether the individual was born in Utah, birth order and birth cohort). Maternal
105 identity was included as a random effect to account for non-independence of observations
106 within families. Number of children born was standardised within sexes prior to the analysis
107 (mean of zero; SD of one) and we used post-reproductive lifespan as a Gaussian response
108 variable. The initial model included the interaction between sex and both the linear and the
109 quadratic (non-linear) term for number of children born. The 95% Bayesian credibility
110 intervals for the interaction term indicate whether the linear or quadratic slope of the
111 relationship between number of children and post-reproductive lifespan differed significantly
112 between the two sexes. The quadratic term was not significant and subsequently excluded
113 from the model. Because the maximum number of children was 64 in males and 21 in
114 females, we investigated whether males with extreme values influenced the result by
115 removing the 282 males with 22 or more children, generally polygamist men. After this
116 equalisation of potential reproductive effort in the two sexes, we re-estimated the above
117 model. This did not influence the results so we present results from the full dataset. To obtain
118 sex-specific estimates for the relationship between reproductive effort and lifespan, we re-
119 estimated the model for each sex separately. In the sex-specific models, the quadratic effect
120 was significant in both sexes and was thus retained in both models. Because males that
121 continued to reproduce after age 55 might inflate the estimates, the male-specific model was
122 re-estimated excluding the 3.2% of males that continued to reproduce after age 55. This did
123 not influence the results so we present results from the full dataset. All models were run with
124 a prior with $V = 1$ and a degree of belief parameter (ν) of 0.002. Convergence of runs was
125 assessed by visual inspection of output plots.

126

127 Alternatively, post-reproductive lifespan can be analysed in a survival model framework. To
128 increase comparability with the models on sexual dimorphism in lifespan, and to obtain sex-
129 specific estimates of acceleration factors for individuals with different levels of reproductive
130 output, we used parametric survival models using the function *survreg* in the R-package
131 survival. The subset used for analyses was restricted to individuals that survived until age 55
132 and their lifespan was used as the response variable. A Weibull distribution provided the best
133 fit to the data in both sexes. Models included fixed effects as above, observations were
134 clustered by maternal identity and birth cohort was added as a stratified fixed effect to allow
135 different baseline survival shapes for each birth cohort. To obtain acceleration factors for
136 different levels of reproductive output, we grouped the number of children born into six
137 different levels representing low to high reproductive investment: one child, two to four
138 children, five to eight children, nine to 14 children, 15 to 21 children and 22 to 64 children
139 (males only).

140

141 Infant mortality may influence reproductive decisions and thus affect costs of reproduction,
142 potentially in a sex-specific manner. The main cause of infant mortality in historical Utah
143 was infectious diseases⁴ and it is thus possible that infant mortality could affect the health and
144 survival of mothers and fathers differently. While infant mortality in historical Utah was
145 comparatively low, in the current study, 33% of reproducing women with full known
146 reproductive history from birth cohorts 1 and 2 (that reproduced when fertility rates were
147 high, Figure 1) lost one or more infants during the first year of life, while this dropped to 23%
148 for women from cohorts 3 and 4 (that reproduced when fertility rates were lower). To account
149 for the effects of infant mortality, we repeated the analyses of costs of reproduction with
150 infant loss included as a two-level covariate (the individual had one or more children that

151 died during the first year of life versus all children survived their first year of life). This did
152 not influence the results (not shown) and we thus present models excluding this variable,
153 because the predictions regarding how infant mortality is expected to influence total fertility
154 and sex-specific survival patterns are not straightforward.

155

156 **References**

157

158 1 Albright, L. A. C. Utah family-based analysis: Past, present and future. *Hum Hered*
159 65, 209-220, doi:10.1159/000112368 (2008).

160 2 Bean, L. L. & Mineau, G. P. The polygyny-fertility hypothesis - a re-evaluation. *Pop*
161 *Stud-J Demog* 40, 67-81, doi:10.1080/0032472031000141846 (1986).

162 3 Bean LL, Mineau GP & DL, A. Fertility change on the American frontier: Adaptation
163 and innovation. (University of California Press, 1990).

164 4 Bean, L. L. et al. Infant deaths in Utah, 1850-1939. *Utah historical quarterly*. 70, 16
165 (2002).

166 5 A package for survival analysis in S v. R package version 2.37-7 (2014).

167 6 Olshansky, S. J. & Carnes, B. A. Ever since Gompertz. *Demography* 34, 1-15,
168 doi:10.2307/2061656 (1997).

169 7 Hadfield, J. MCMC methods for multi-response generalized linear mixed models: The
170 MCMCglmm R package. *J Stat Software* 33, 1-22, doi:10.18637/jss.v033.i02 (2010).

171

172 **Supplementary tables and figures**

173

174 **Table S1.** Averages (\pm SD) and sample sizes of number of children and lifespan in the two
175 sexes in four birth cohorts in historical Utah.

176

Birth cohort	Average children \pm SD		Average lifespan \pm SD		N	
	females	males	females	males	females	males
1820 to 1844	8.5 \pm 3.3	11.1 \pm 5.7	68.3 \pm 17.8	70.1 \pm 14.3	3298	3793
1845 to 1869	7.8 \pm 3.3	8.3 \pm 3.8	69.5 \pm 17.5	70.2 \pm 14.6	17331	16403
1870 to 1894	5.8 \pm 3.1	5.8 \pm 3.0	72.4 \pm 18.1	70.8 \pm 15.6	37156	33919
1895 to 1919	4.2 \pm 2.5	4.2 \pm 2.4	76.7 \pm 17.5	72.6 \pm 15.7	17882	10818

177

178 **Table S2. Proportion of reproducing women who died the same year in which they gave**
179 **birth to their last child,** in four birth cohorts in historical Utah. Sample sizes as reported in
180 Table S1.

181

Birth cohort	1820 to 1844	1845 to 1869	1870 to 1894	1895 to 1919
Proportion deceased	0.0633	0.053	0.037	0.019

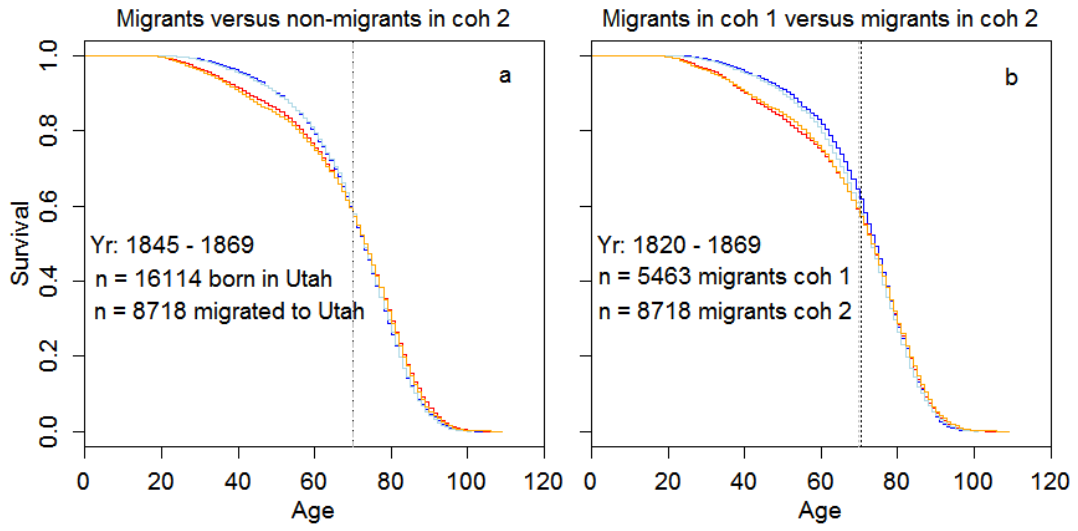
182

183

184 **Figure S1.**

185 Comparison of sex-specific survival patterns between (a) individuals that migrated into Utah
186 and individuals that spent their entire lives in Utah in birth cohort (coh) 2 and (b) individuals
187 that migrated into Utah in birth cohort 1 compared to migrants in birth cohort 2. Birth cohort
188 2 is composed of four groups of individuals with respect to migratory status: individuals that
189 were born and died in Utah, individuals that were born elsewhere and died in Utah,
190 individuals that were born in Utah but died elsewhere and individuals that spent their entire
191 lives outside Utah. In (a), the first two of these groups are compared, separately in the two
192 sexes. The early migrants to Utah faced harsher conditions during migration than later
193 migrants. Therefore, figure (b) compares the effects of migration on survival in the early
194 migrants into Utah of cohort 1 with the later migrants of cohort 2, separately in the two sexes.
195 Note that migrants into Utah in cohort 2 thus appear in both figures, males in light blue and
196 females in orange. In (a), these migrants are compared to non-migrants within the same
197 cohort (males in blue and females in red), whereas in (b), the migrants in cohort 2 are
198 compared to migrants in cohort 1 (males in blue and females in red). Shown are the birth
199 years (Yr) contained in birth cohort 1 (a) or 1 and 2 (b) and the sample size in each group
200 (N). Dotted vertical lines indicate the average lifespan of (a) non-migrants (black) and
201 migrants (grey) in both sexes pooled and of (b) migrants in cohort 1 (black) and cohort 2
202 (grey). All p-values were > 0.2 from GW and LR tests of the difference between individuals
203 within each sex that were born in Utah and individuals that migrated into Utah (a), and
204 likewise (b), all p-values comparing migrants in cohort 1 to migrants in cohort 2 were > 0.25
205 in females. Migrant males in cohort 1 tended to live longer than migrant males in cohort 2,
206 GW $\chi^2 = 9.3$, $p = 0.002$, LR $\chi^2 = 3.7$, $p = 0.053$.

207



209 **Figure S2.**

210 Age-specific adult survival for females (red) and males (blue) that presumably did not
211 reproduce, divided into four 25-year birth cohorts that cover the demographic transition in
212 Utah. Shown in each figure are the birth years contained in that cohort (Yr), the sample size
213 (n), the average \pm SD number of children born to females (CHf), the χ^2 and p-values
214 indicating differences in the survivor function between the two sexes from the Peto and Peto
215 modification of the Gehan-Wilcoxon test (GW, this test weights differences in survivorship
216 that occur early more heavily than differences at later survival times) and from a log-rank test
217 (LR, this test weights differences at later survival times more heavily). Dotted vertical lines
218 indicate the average lifespan of each sex in each cohort.

