DISCUSSION PAPER SERIES

DP17835

THE INHERITANCE OF SOCIAL STATUS: ENGLAND, 1600-2022

Gregory Clark

ECONOMIC HISTORY AND LABOUR ECONOMICS

CEPR

ISSN 0265-8003

THE INHERITANCE OF SOCIAL STATUS: ENGLAND, 1600-2022

Gregory Clark

Discussion Paper DP17835
Published 23 January 2023
Submitted 17 January 2023

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

THE INHERITANCE OF SOCIAL STATUS: ENGLAND, 1600-2022

Abstract

A lineage of 422,215 English people 1600-2022 contains correlations in social outcomes among relatives as distant as 4th cousins. These correlations show striking patterns. First is the strong persistence of social status across family trees. Correlations decline by a factor of only 0.8 across each generation. Even fourth cousins, with a common ancestor only five generations earlier, show significant status correlations. The second remarkable feature is that the decline in correlation with genetic distance in the lineage is unchanged 1600-2022. Vast social changes in England between 1600 and 2022 would have been expected to increase social mobility. Yet people remain correlated in outcomes with their lineage relatives in exactly the same way as in pre-industrial England. The third surprising feature is that the correlations parallel those of a simple model of additive genetic determination of status, with a genetic correlation in marriage of 0.6.

JEL Classification: N/A

Keywords: Intergenerational social mobility

Gregory Clark - gclark@ucdavis.edu
University of California, Davis and CEPR

Acknowledgements
The FOE and marriages database employed in this paper was developed in collaboration with Neil Cummins, LSE. I thank the members and leadership of the Guild of One-Name Studies for allowing us access to their genealogies. I thank the members of the FreeReg organization for the transcripts of parish wedding records they have made publicly available. Zhiming Zhu of LSE provided excellent RA work in collecting the modern Essex marriage records. I thank James Lee, David Cesarini, Alex Young and Rosalind Ardern for helpful suggestions on the paper.

The Inheritance of Social Status: England, 1600-2022

Gregory Clark[1]
Department of Economics, University of California, Davis, CA 95616 and
Department of Economic History, London School of Economics, UK
Email: gclark@ucdavis.edu

Abstract

A lineage of 422,215 English people 1600-2022 contains correlations in social outcomes among relatives as distant as 4th cousins. These correlations show striking patterns. First is the strong persistence of social status across family trees. Correlations decline by a factor of only 0.8 across each generation. Even fourth cousins, with a common ancestor only five generations earlier, show significant status correlations. The second remarkable feature is that the decline in correlation with genetic distance in the lineage is unchanged 1600-2022. Vast social changes in England between 1600 and 2022 would have been expected to increase social mobility. Yet people remain correlated in outcomes with their lineage relatives in exactly the same way as in pre-industrial England. The third surprising feature is that the correlations parallel those of a simple model of additive genetic determination of status, with a genetic correlation in marriage of 0.6.

Introduction

Using a large genealogical database, which details the family connections of 422,215 people with rarer surnames in England for births 1600-2022, the paper

[1] The FOE and marriages database employed in this paper was developed in collaboration with Neil Cummins, LSE. I thank the members and leadership of the *Guild of One-Name Studies* for allowing us access to their genealogies. I thank the members of the FreeReg organization for the transcripts of parish wedding records they have made publicly available. Zhiming Zhu of LSE provided excellent RA work in collecting the modern Essex marriage records. I thank James Lee, David Cesarini, Alex Young and Rosalind Ardern for helpful suggestions on the paper.

examines patterns of inheritance of social status in both pre-industrial and contemporary England. Social status is measured by six outcomes: occupational status, higher education status, literacy, dwelling value, company directorships, and the index of multiple deprivation (IMD) for the residence location. Status correlations are calculated for all these outcomes for relatives up to fourth cousins. Table 1 shows that set of correlations.

These status correlations reveal four things. The first is that status persists strongly across even very distant relatives, across all measures of status. Even fourth cousins, who shared a common ancestor only five generations earlier, typically show statistically significant correlations in status. The second is that the decline in status correlations with each step outward in the lineage is a constant 0.8, for different measures of status, and for different epochs from 1600 to 2022. The vast social changes in England since the Industrial Revolution, including mass public schooling, have not increased, in any way, underlying rates of social mobility.

The third interesting feature of the correlations are that they conform closely to those predicted by Ronald Fisher in 1918, for familial correlations in the presence of strong assortment in mating.[1,2,3,4] In particular, the correlation in mating in the genetics underlying social outcomes would have to be 0.6 to generate the persistence rate of 0.8. There is ancillary evidence that the phenotypic assortment in marriage in England for underlying social status is around 0.8, and largely unchanged for marriages 1837-2022.[5]

Since this is observational data there is no proof here that additive genetic transmission causes social status. All we can determine is that whatever social processes are producing the observed outcomes have a form of transmission which mimics that of additive genetic effects, in the presence of the important social institution of strong assortative mating. Two recent whole genome studies for Britain, however, show correlation in marital partners of genetic predictors of educational attainment that are consistent with the 0.6 correlation.[6,7]

Table 1: Social Status Correlations by Familial Connection, England, 1600-2022

Outcome	Modstat	House value	IMD	CoDir	Occstat	Occstat	HighEd	HighEd	Literacy
Birth Period	1910-96	1910-1996	1910-1996	1910-1996	1780-1859	1860-1919	1780-1859	1860-1919	1725-1869
Pairs observed	98,119	98,117	98,711	219,552	104,854	241,941	103,506	219,507	53,047
Correlations									
Full Sibling	0.381	0.334	0.271	0.175	0.584	0.496	0.558	0.359	0.407
Child	0.396	0.352	0.319	0.141	0.585	0.513	0.500	0.305	0.436
Sibling-rem	0.273	0.252	0.181	0.061	0.502	0.369	0.398	0.213	0.339
Grandchild	0.320	0.268	0.246	0.104	0.434	0.347	0.426	0.273	0.252
Cousin	0.223	0.213	0.155	0.064	0.465	0.277	0.366	0.146	0.270
Cousin-rem	0.148	0.158	0.101	0.015	0.325	0.257	0.235	0.181	0.206
Cousin2	0.137	0.139	0.074	0.069	0.232	0.151	0.103	0.045	0.225
Cousin2-rem	0.098	0.093	0.057	0.025	0.189	0.140	0.098	0.076	0.182
Cousin3	0.099	0.095	0.051	0.049	0.114	0.138	0.186	0.099	0.189
Cousin3-rem	0.063	0.076	0.017	0.019	0.065	0.139	0.118	0.079	0.202
Cousin4	0.079	0.084	0.032	0.024	0.079	0.090	0.020	0.040	0.146
Unrelated	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: "-rem" indicates "once removed". Cousin2, 3 and 4 indicates second, third and fourth cousins. Modstat is a PCA index that combines House Value, IMD and the Company Director indicator (CoDir). "Occstat" is occupational status, "HighEd" an indicator for higher education. "Pairs Observed" is the total number of pairs of relatives used in estimating the parameters of equation (3).

Even if it is the case that in England 1600-2022 social status was mainly determined by genetic inheritance, this does not in itself imply that social interventions cannot change social outcomes. There has been much discussion in recent years of an alternative causal path through genetic nurture. However, the constancy of the patterns of status persistence across the interval 1600-2022 does suggest social interventions have surprisingly modest effects. Before 1870 there was little public provision of education, of health care, or of income support. Families largely depended on their own resources. Since 1920 there have been

increasing levels of public provision of education, health care, and basic needs. These services should have helped, in particular, poorer families.[8] Yet we see no corresponding increase in rates of social mobility.

Results

As Fisher demonstrated in 1918, the expected correlation in phenotypes of relatives for a trait transmitted through additive genetic determination, with marital assortment, is dependent on only three parameters: the hereditability of the phenotype, h^2, and the genetic correlation of parents in the relevant genetics, m, and the number of steps distant are the relatives in the family tree, n. The details are shown in table 2.

The key parameter determining long term persistence of correlations is m, the genetic correlation of spouses. With no assortment, the expected correlation of a trait, even with a heritability of 0.7, for fourth cousins, would be 0.001. Even for second cousins it would be only 0.02.

Using observed correlations in status across relatives, m and h^2 can be estimated from the set of observed correlations. For all but linear descendants of one parent the expected correlation on the Fisher formulae, ρ_n has the form

$$\ln(\rho_n) = \ln(h^2) + n\ln\left(\frac{1+m}{2}\right) \quad (1)$$

In the other cases the expected correlation is

$$\ln(\rho_n) = \ln(h^2) + n\ln\left(\frac{1+m}{2}\right) + \ln\left(\frac{1+r}{1+m}\right) \quad (2)$$

This means that we can estimate m and h^2 from the parameters of a linear regression

$$\ln(\rho_n) = a + \ln(b).n + c.d_{lin} \quad (3)$$

Table 2: Correlations between relatives with assortative mating

Relative to Child	Parents Match on Phenotype	Relative to Child	Parents Match on Phenotype
Average of parents	h^2	Ave. Grandparents	$h^2\left(\frac{1+m}{2}\right)$
Full Sibling	$h^2\left(\frac{1+m}{2}\right)$	Single parent	$h^2\left(\frac{1+r}{2}\right)$
Uncle/Aunt	$h^2\left(\frac{1+m}{2}\right)^2$	Single grandparent	$h^2\left(\frac{1+m}{2}\right)\left(\frac{1+r}{2}\right)$
Cousin	$h^2\left(\frac{1+m}{2}\right)^3$	Cousin removed	$h^2\left(\frac{1+m}{2}\right)^4$
Second Cousin	$h^2\left(\frac{1+m}{2}\right)^5$	Second Cousin rem.	$h^2\left(\frac{1+m}{2}\right)^6$
Third Cousin	$h^2\left(\frac{1+m}{2}\right)^7$	Third Cousin rem.	$h^2\left(\frac{1+m}{2}\right)^8$
Fourth Cousin	$h^2\left(\frac{1+m}{2}\right)^9$		

Note: m is the correlation of parents on the relevant genotype, r the correlation on the relevant phenotype. h^2 is the regression coefficient of the child phenotype on the average of the parents' phenotypes.

where d_{lin} is an indicator which is 1 for the cases where the phenotype marital correlation appears. $b = \frac{1+m}{2}$ is the persistence rate of the correlation as we move one step down the family tree, or one step across between full siblings.

It would normally be anticipated that $m = rh^2$. However, we shall see below that for social status outcomes there is good evidence that marital matching is on latent status, not captured well by observed phenotypes such as years of education. In this case the observed phenotype correlation of parents could be

less than m. Thus it will be hard to distinguish empirically the two cases where the phenotype correlation appears.

Once *m* is estimated we can graph the link between the implied fraction of shared genotype and the correlation of social outcomes.

Figure 1 shows the estimated values of $b = \left(\frac{1+m}{2}\right)$ for each of the outcome measures from equation (3), the confidence intervals. This is plotted against the estimated heritability of each trait, also from (3). Table A6 gives the estimated values of *b, m* and h^2 from these estimates, as well as the R^2 of the fit, which averages 0.86.

As figure 1 shows, the estimates of b, social status persistence, cluster around 0.80 for all nine measures, even though the measured heritability of traits varies substantially. The high R^2 of the fit implies the Fisher formulas predict well the correlations. Figure 1, which shows the same underlying rate of social mobility from the eighteenth century to the present, suggests that possibly all social status shows the same persistence parameter of 0.8. If this is through additive genetic transmission, then also throughout this period marital partners had to be correlated 0.6 on the relevant genetics.

Figure 2 illustrates how well the assumption of additive genetic determination of social status, with marital assortment at 0.6, describes the data. With the assumption that m = 0.6, we can arrange the various pairs of relatives in terms of their shared genotype on the horizontal axes. Then on the vertical axis we can plot the relevant correlation. In figure 2, this is the log house value 2017. As noted above, the house value is normalized by region to remove regional effects.

House value is serving here as an indicator or the income of the family. The fit is based on 98,117 house value correlations between different sets of relatives.

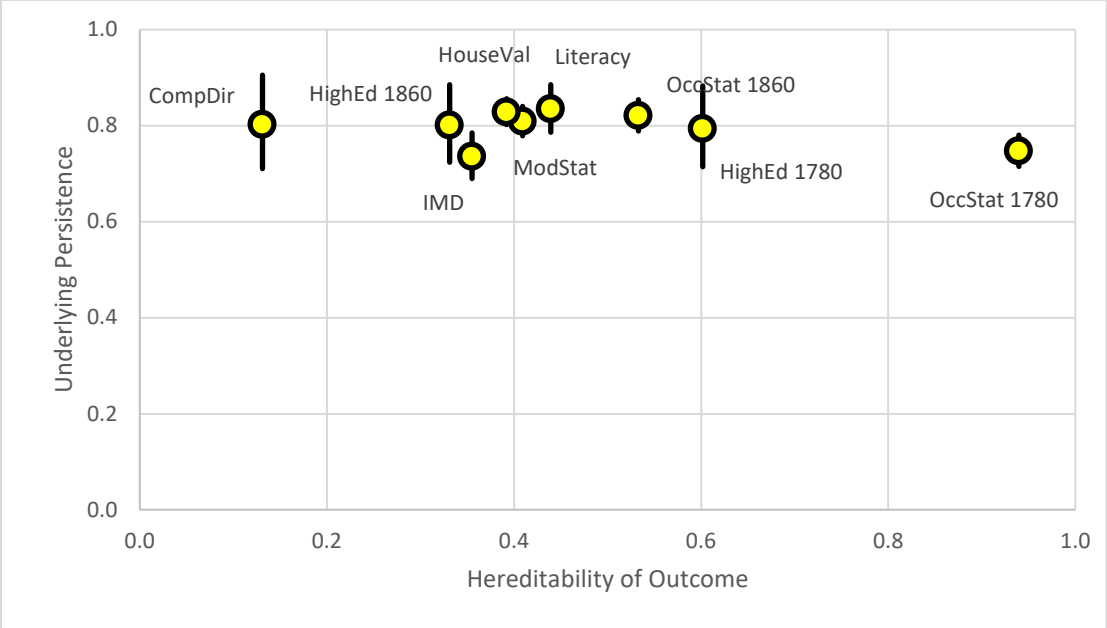
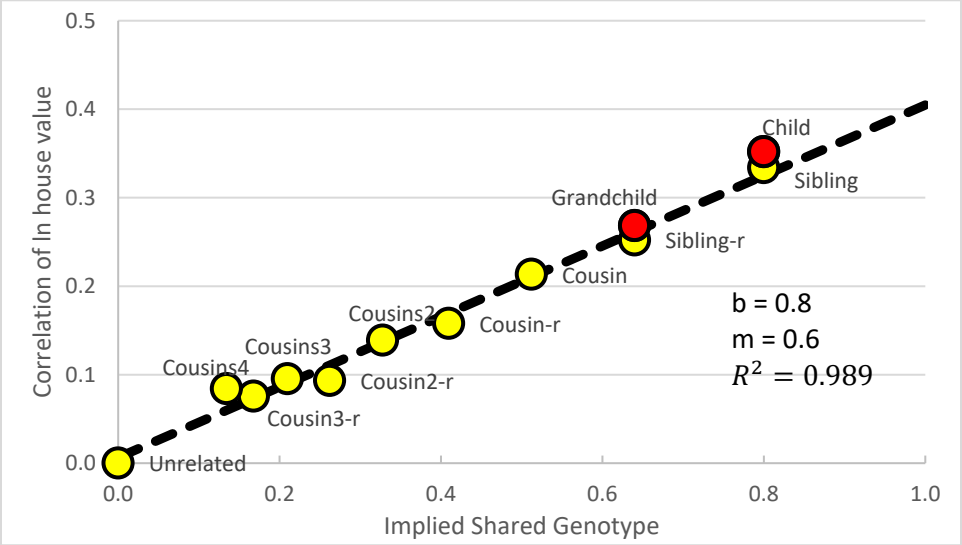


Figure 1: Estimates of persistence versus heritability, births 1780-1996

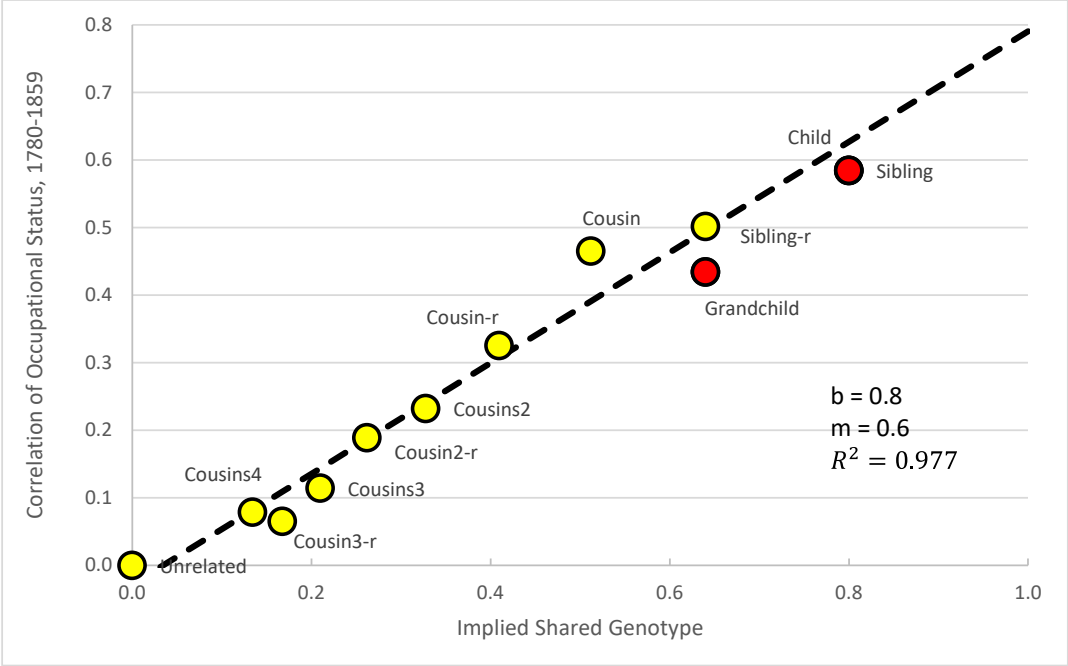
Notes: Company Director (2015-22), "CompDir", Ln House Value, 1999-2022, "HouseVal", Index of Multiple Deprivation, 1999-2022, "IMD", Modern Social Status, 1999-2022, "ModStat", Occupational Status births 1780-1859, "OccStat1780", Occupational Status births 1860-1919, "OccStat1860", Higher Education births 1780-1859, "HighEd1780", Higher Education births 1860-1919, "HighEd1860", Literacy Marriages 1754-1879, "Literacy." Lines indicate 95% confidence intervals for the estimates.

Figure 2: Social Status Correlations and Implied Shared Genotype- Ln House Values, births 1920-1996



Notes: The dashed line shows the OLS fit to this data. The R^2 reported is for this fitted line. The child and grandchild correlations (in red) will potentially deviate from the fitted line.

Figure 3: Social Status Correlations and Implied Shared Genotype - Occupational Status Correlations, births 1780-1859



Notes: As figure 2.

As would be predicted with additive genetic transmission of outcomes, there is a clear linear relationship in figure 2, between the implied genotype share of relatives and the house value correlation. The R^2 of the fitted line here is 0.985, and the intercept with the vertical axis is not statistically different from 0.

This linearity of the relationship in figure 2 further emphasizes the stability of the persistence rate b over generations. If persistence from one generation to the next was much higher in earlier years, then for fourth cousins, where on average in this data the common ancestor was born in 1804, the measured correlation now would be above the fitted line in the figure for more distant relatives. Note also that even for fourth cousins in 1999-2022, who would likely have no social interaction, the correlation in house values within regions is both quantitatively and statistically significant.

Figure 3 similarly shows the close correlation between the implied fraction of shared genes (assuming m = 0.6) and the correlation of occupational status for men born 1780-1859. The fit here is based on 104,584 pairs of occupational status. Again the relationship is linear, as additive genetic transmission would imply. Again the OLS fitted line intercepts the vertical axis close to 0. And again this implies a stability in the persistence of status across generations all the way from 1678 or earlier, when on average fourth cousins had a common ancestor, to 1859.

The appendix shows the correlations of status and implied shared genotype for the other seven status measures, assuming genetic marital assortment of 0.6 (figures A3-A9). The R^2 of the fit varies with the numbers of observations, and the heritability of the trait. But it averages 0.92. The figures in the appendix look similar to the ones presented here, and show the same consistent pattern in inheritance.

The lineage database also contains a large number of observations on wealth at death for men and women dying 1800-2022. This measure was not included in table 1 and figure 1 because it clearly involves the non-genetic transfer of wealth between generations. For richer families that transfer was also affected by social elements such as the number of children in a family, or by the gender of the child.

Wealth inheritance also shows a significant asymmetry between men and women in a way that is inconsistent with additive genetic transmission. The implied persistence of wealth by generation, however, is even higher than for the measures used here, being 0.84 for births 1780-1859 and 0.86 for 1860-1919.

The finding of an intergenerational persistence rate of 0.8 that is stable over time is buttressed by surname status studies carried out by the author and collaborators. Rarer surnames often deviate on average social status from the social mean. Surname inheritance in a society such as England follows the same pattern as the y chromosome. Thus the rate of movement of surname status towards the social average should show a persistence across each generation of about 0.8. For England there is exactly such a surname status persistence, unchanged persistence from the seventeenth century until now.[15,16]

Data

Table 1 above summarizes the correlations of social status outcomes for nine measures of social status. For the current period, births 1910-1995, there is, for both genders, estimated log house value, normed to 2017, the Index of Multiple Deprivation, Company Director, and a combined social status score from these first three measures. For these contemporary correlations all the data is included, but the common ancestor between two individuals must be born 1780 or later. The elite lineages cannot be used for ancestors born before 1780, since they were selected on the basis of the status of ancestors born in the period 1780-1840.

For men born 1860-1919, and 1780-1859 we have both occupational status, and attainment of higher education. Women in England were not admitted to most universities and professional qualifications until 1920 or later, so though there were highly educated women, there is no formal record of that. Middle and upper class women typically did not work outside the home, so occupational status measures for women before 1920 are not very useful indicators of social status. Finally for men and women born 1725-1862 we have literacy measured at marriage. This was recorded for all marriages 1754 and later.

Table 1 also shows that in all cases where individuals were randomly assigned partners from another lineage, and so unrelated, the estimated correlation was statistically indistinguishable from 0. There is nothing in the structure of the data that is spuriously creating correlations even between unrelated individuals.

Additional Tests

Another implication of additive genetic transmission is symmetry of mothers and fathers in transmitting status to children. As noted above, for much of this period 1600-2022 we do not observe social status outcomes (except literacy) for women. But we can proxy the implied status of mothers and fathers by using the status of the maternal and paternal grandfathers. We can then estimate the parameters b_f and b_m in the equation

$$y_c = a + b_f y_{gf} + b_w y_{gm} + e \qquad (4)$$

where y_c is the social outcome for the grandson, y_{gf} the outcome for the paternal grandfather, and y_{gm} the outcome for the maternal grandfather. For births in the period 1780-1919 we have as grandfather outcomes occupational status, higher education, wealth at death. In addition for 1754-1879 we have mother and father literacy at marriage, and child literacy at the child's marriage, which allows a direct estimate of the relative predictive ability of mother versus father literacy for both daughters and sons.

Figure 4 shows the estimated coefficients from (4), and directly for literacy. For literacy, higher education attainment, and occupational status, there is no significant difference in the predictive effect of father versus mother status (or that of their fathers). But the wealth of the paternal grandfather is three times as large as that of the maternal grandfather in predicting child wealth.

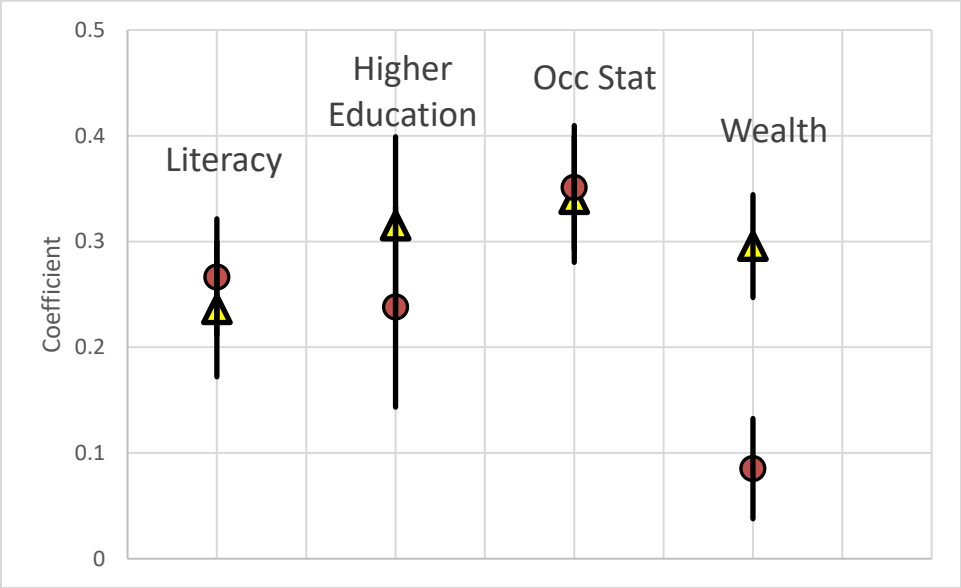
Figure 4 also shows that the focus of the One-Name Studies lineages on the patriline will not exaggerate estimates of status persistence across generations,

except in the case of wealth. For persistence is just as strong in the matriline as in the patriline.

Spousal Genetic Correlation in Marriage

Until recently the finding that an additive genetic model of status determination, combined with the social parameter of a spousal correlation in the underlying genetics of 0.6, would have been dismissed on the grounds that spousal genetic correlations could not be so high. A typical social trait where spousal correlations have been measured is years of education, and here the correlations for the modern UK are typically 0.4-0.5.[6,7,17] On the normal assumption that $m = h^2 r$, this implies a genetic correlation in the relevant genes of less than 0.25.

However, two recent studies of the genetic predictors of Educational Attainment (measured as years of schooling) both imply that the spousal correlation in the genetics relevant to educational attainment is much higher than 0.25. In the first study, based on 7,780 couples in the UK Biobank with measures of educational attainment, the spousal phenotype correlation was only 0.41 (s. e. 0.011). However, the correlation across the same couples at trait associated loci for educational attainment was significantly higher, 0.654 (s. e. 0.014).[6]



Notes: O - mothers, Δ - fathers. 95% confidence intervals indicated by bars. Confidence intervals generated clustering on fathers.

Figure 4: The Comparative Influence of Mothers and Fathers

The second study showed a phenotype correlation in years of education of 0.43 (s. e. 0.017) between 2,465 couples from the UK. There was, however, an unexpectedly high 0.175 correlation (s. e. 0.020) in the PGI for educational attainment.[7] Since the PGI is a noisy measure of the full genetic educational potential, the full correlation will be significantly higher than this measured correlation.

If we take the analogous case of height, also reported in this paper, the phenotype correlation between spouses was 0.290 (s.e. 0.018), but the PGI index correlation was only 0.106 (s.e. .020).[7] Since height has a heritability of 0.8, and is largely genetically determined in high income societies, the true genetic correlation between partners in height would thus be 0.236. This implies the PGI correlation for height between partners has to be multiplied by 1.65-3.27 to estimate the full genetic correlation. If we apply this same adjustment to the

measured genetic correlation between spouses for educational attainment then the implied actual correlation averages 0.39, with a 95% confidence interval of 0.29-0.57. The height PGI is based on larger samples, and height as a phenotype has less noise than educational attainment. So the .175 genetic correlation observed between partners for educational attainment is potentially consistent with a true genetic correlation of 0.6.[18]

Another recent study for Norway, with 26,681 pairs of partners and 2,170 pairs of siblings found a 0.42 phenotype correlation between partners in years of education, but an estimated 0.37 genetic correlation for educational attainment between partners. This is lower than the UK estimates, but the 95% confidence interval for this estimate was 0.21-0.67. In line with this partner correlation, the sibling genetic correlation was estimated as 0.68 (95% CI 0.61-0.75). Comparison of the genetic similarities of partners and siblings implied that assortative mating at the observed level had taken place for at least five generations in Norway.[19]

Thus the evidence, at least for the modern UK, is that parents are matching much more strongly on a latent social abilities phenotype than they are on the observed phenotypes such as years of education, occupational status, or income. This strong matching then makes possible the high observed genotype correlation.

We can find evidence in marital records for England and Wales 1837-2022 for just such strong latent status phenotype matching.[14] As noted above, these marital records, collected again by amateur genealogists, show occupations for grooms and brides and their respective fathers. Suppose that grooms and brides match in marriage to some social status phenotype they observe, with a correlation, r. Suppose also we only have noisy measures of this phenotype, such as years of education, or an occupational status index. In that case the observed phenotype correlation in marriage will be biased downwards by some factor $\theta < 1$. But suppose also that both bride and groom correlate in their true social phenotype with a correlation of β with their respective fathers. This implies that the observed correlation of groom to his father will be $\theta\beta$. The observed correlation of the groom to their father-in-law, if the matching in marriage is just

bride to groom, will be $\theta r \beta$. This implies that the true correlation between bride and groom in their social phenotype can be calculated as

$$\frac{groom-father-in-law\ correlation}{groom-father\ correlation} = \frac{\theta r \beta}{\theta \beta} = r \qquad (5)$$

For marriages in England and Wales, 1837-2022 this underlying correlation is consistently close to 0.79 across all periods (see table A7 for details).[5] It may be objected that if the groom or the groom's father are also matching directly to the father-in-law, the measured marital correlation will be driven upwards. However, this estimation produces the same marital correlation in cases where the father-in-law is dead at the time of the marriage, or in cases where the father is dead. In such cases we would expect less groom-father-in-law matching if that was occurring, and consequently a lower estimated marital correlation. We observe no sign of that.

A marital correlation in a latent social status phenotype of 0.79 is compatible with a correlation in social status genetics of 0.6. It would rely on a heritability of the underlying social status of 0.76, which is high but similar to that for height. Thus the evidence on strong latent phenotype matching in marriage throughout the years 1837-2022 is consistent with the evidence above of strong and stable genetic matching throughout this period.

Materials and Methods

The lineage connections in the database were largely identified by amateur genealogists constructing family trees. Family lineage studies can involve significant problems of selectivity, where more notable ancestors, or those leaving descendants, are more often included. To avoid such problems of selectivity in who gets included in a family tree, the lineages used here are mainly those constructed by the members of the *Guild of One-Name Studies*.[9] *Guild* members aim to include all persons with a chosen rare surname – Argall, Errey, Mitchelmore, etc. - in their lineages. This avoids the problem of selective inclusion, though because surnames are preserved at marriage only for males, it

does focus on the patriline. However, it is shown below that for most outcomes except wealth intergenerational transmission of status is symmetric on the matriline and the partriline. Also comparison of wealth, literacy, and occupational status for the lineages used here, detailed in the supplement, suggests these lineages are only of modestly higher than average status across the years 1800-2022.

To many of these lineages derived from *Guild* members have been added additional information on social outcomes derived from census records 1841-1911, from the 1939 population register, marriage records 1837-2022, ship passenger records, the electoral rolls 1999-2022, registers of company officers, matriculation records for Oxford, Cambridge, Durham and London Universities, the medical register 1857-2022, armed forces appointments, and members of engineering societies.

For people in the most recent years the electoral rolls 1999-2022 reveal the address of many individuals.[10] This makes it possible to estimate the value of the house people were living in, by postcode, using the UK Land Registry data on sales 1995-2017 (where the typical postcode covers 40 houses).[11] Since the data shows that people show strong geographic persistence, and since house values vary substantially by region in England and Wales, we normalize house values in the sample to their deviation from the average house value across 6 regions (North, Midlands, Wales, East and South East, London, South West). From the address we also observe the social status of the local area (around 1,000 households) as expressed by the Index of Multiple Derivation (IMD) for 2019.[12] Independently we can identify if a person alive 2015 and later was a Company Director, from the Director's Register.[13] To get an independent measure of status from the address we included only individuals ages 24 and above who were not at the same address as a parent. Using the three measures – house value, IMD, and Company Director – we derive using Principal Component Analysis a more general measure of social status, "Modstat" for those living 1999 and later.

In earlier years we have two measures of status which apply only to males. The first is occupational status. An index of social status by occupation was estimated from 1.4 million marriage records 1837-1939 which give occupations at marriage for the groom, his father, and his father-in-law.[14] Status is assigned to occupations in such a way as to maximize the father-groom and father-in-law-groom correlations. The second is whether a person had attained higher education such as attending university or a military academy, and/or qualifying as an attorney, doctor, engineer, or clergyman. We have a further measure, literacy at marriage, which applies to both men and women marrying 1754-1879. This is inferred from the ability to sign the marriage register. Because of more modest numbers of observations for literacy we only extend this to sets of relatives of second cousins once removed or closer. The higher education measure will tend to be informative of educational status for those of higher status, while the signature measure will be informative for those of lower status.

References

1. R. A. Fisher. The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh, **52**:* 399-433 (1918).
2. D. S. Falconer. *Introduction to Quantitative Genetics*. 2nd ed. London: Longman. (1981).
3. J. F. Crow & J. Felsenstein. The effect of assortative mating on the genetic composition of a population, *Eugenics Quarterly*, **15**:2, 85-97, (1968).
   DOI: 10.1080/19485565.1968.9987760
4. T. Nagylaki. "The correlation between relatives with assortative mating." *Annals of Human Genetics, **42***: 131. (1978).
5. G Clark, N. Cummins, Assortative Mating and the Industrial Revolution: England, 1754-2021. (2022). https://hub.cepr.org/discussion-paper/138817.
6. M. R. Robinson, et al., 2017 Genetic evidence of assortative mating in humans, *Nature Human Behaviour*, 2016, **1**. Article number: 0016 (2017).
7. Okbay, A., Wu, Y., Wang, N. *et al.* Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nat Genet* **54,** 437–449 (2022). https://doi.org/10.1038/s41588-022-01016-z
8. Lindert, P. *Growing Public: Social Spending and Economic Growth since the Eighteenth Century*. Cambridge: Cambridge University Press. (2004). doi:10.1017/CBO9780511510717.
9. https://one-name.org/
10. https://www.192.com/people/search/
11. https://landregistry.data.gov.uk/app/ppd
12. https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019
13. https://find-and-update.company-information.service.gov.uk/search/officers?
14. https://www.freereg.org.uk/.
15. G. Clark, & N. Cummins. Surnames and Social Mobility: England, 1170-2012. *Human Nature, **25***(4), 517-537. (2014).

16. G. Clark, N. Cummins et al. *The Son Also Rises: Surnames and the History of Social Mobility.* Princeton: Princeton University Press. (2014).
17. Mascie-Taylor, C. G. Nicholas. Assortative mating in a contemporary British population. *Annals of Human Biology*, *14*:1: 59-68. (1987). DOI: 10.1080/03014468700008841
18. One of the authors of this paper, Alex Young, has developed a more sophisticated correction for the PGI attenuation which implies a true genetic correlation of 0.56-0.75.
19. Torvik, F.A., Eilertsen, E.M., Hannigan, L.J. *et al.* Modeling assortative mating and genetic similarities between partners, siblings, and in-laws. *Nat Commun* **13**, 1108 (2022). https://doi.org/10.1038/s41467-022-28774-y.
20. https://mitchelmore.one-name.net/
21. http://auty-1.one-name.net/
22. G Clark, N. Cummins, M. Curtis. The Mismeasurement of Man: Why Intergenerational Mobility is much lower than Conventionally Measured, England, 1800-2021. (2022). https://cepr.org/publications/dp17346.
23. Ancestry.com.
24. Paul S. Lambert, et al. The Construction of HISCAM: A Stratification Scale Based on Social Interactions for Historical Comparative Research, *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 46:2, 77-89, (2013) DOI: 10.1080/01615440.2012.715569
25. Prandy, K., & Lambert, P. S. Marriage, Social Distance and the Social Space: An alternative derivation and validation of the Cambridge Scale. *Sociology*, 37(3), 397-411. (2003)
26. https://www.statista.com/statistics/294729/uk-population-by-region/
27. https://www.gov.uk/government/news/uk-house-price-index-hpi-for-june-2017#:~:text=The%20data%20shows%3A,property%20value%20to%20%C2%A3151%2C672.

Appendix

The materials for this study are a database of 422,215 individuals linked to their parents and spouses who lived, or had ancestors living, in England and Wales 1600-1822 (the Families of England (FOE) database). This database has two components. The majority of the data is from a set of lineages of persons with rare surnames created by members of the *Guild of One-Name Studies*.[9] These lineages incorporate everyone with a rare surname of interest, wherever they reside, as well as spelling variants of the surname. Thus the Mitchelmore lineage, for example, incorporates the surnames Michelmore, Mitchelmore, Mitchamore, Mitchmore, Mouchemore, Muchamore, and Muchmore.[20] Similary the Auty lineage encompasses Auty, Autey, Awty, Otty, and Ottey.[21] In cases where we only had access to the published lineages, these did not typically contain details of any living holders of the surname. In these cases we added that information ourselves from public records of births, marriages and addresses. Lineages were chosen for inclusion based on their completeness, and either the public posting of the lineages, or their creators' willingness to share the data with us for inclusion in the study.

In addition to these existing lineages, we ourselves created a set of lineages for rare surnames that were high wealth for people in the lineage born 1780-1850, for the purposes of better estimating social mobility rates through having more variance in social outcomes in the earlier generations. For the estimates based on the residence addresses in the electoral rolls and on company directorships of people in 1999-2022 we employ all lineages where the common ancestor was born 1780 and later. For estimates looking at educational and occupational status for people born 1780-1859 and 1860-1919 we exclude the high status lineages 1780-1850, because this will bias the estimates of persistence for relatives with ancestors born earlier than 1780. For literacy from marriage records for people marrying 1754-1879 we used again just the non-selected rare surname lineages.

Table A1 shows the outline of the source of the data, and its distribution across time, and between general and elite lineages. Table A2 shows the numbers of relationship pairs in the data, again by lineage type. The reason for

Table A1: Families of England Data Outline

Birth Period	All	General Lineages	Elite Lineages
1600-99	5,720	5,472	248
1700-99	27,033	22,422	4,611
1800-49	62,366	52,373	9,993
1850-99	110,882	96,171	14,711
1900-49	76,167	67,485	8,682
1950-2022	46,554	40,556	5,998
All	421,907	363,685	58,222

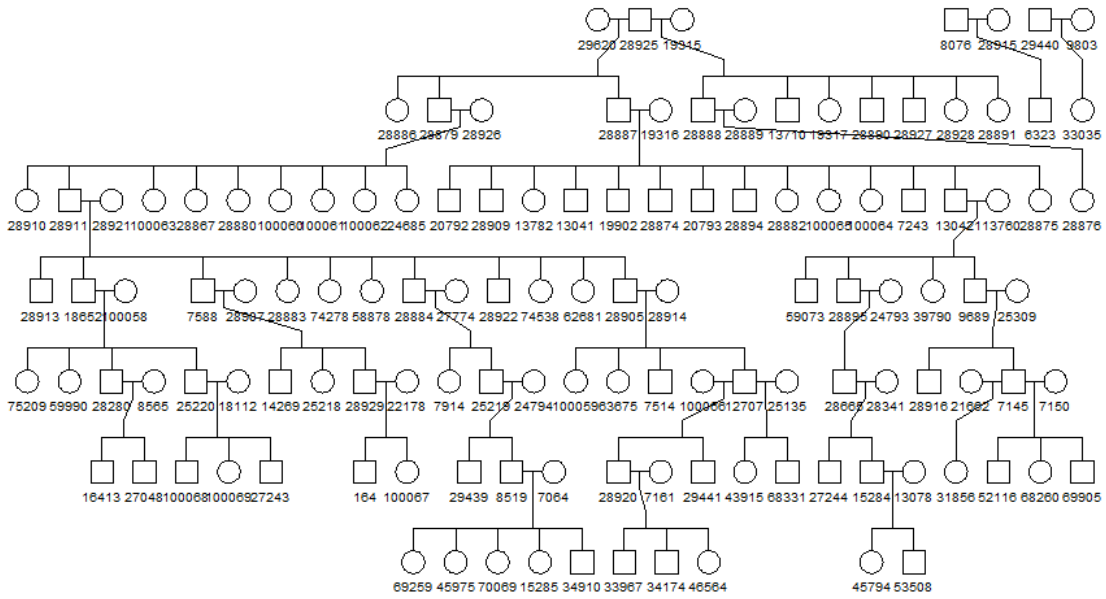
Source: FOE database.

Table A2: Families of England, Numbers of Relationship Pairs

Relationship Pairs	All	General Lineages	Elite Lineages
Spouse	268,957	228,584	40,373
Child-Parent	495,153	431,292	63,861
Sibling (full)	508,144	441,024	67,120
Nephew/Niece	1,029,789	902,283	127,506
Grandchild	465,618	402,834	62,784
Cousin	804,009	726,359	77,650
Cousin Removed	1,713,957	1,563,495	150,462
2nd cousin	1,152,227	1,064,445	87,782
3rd cousins	2,345,628	2,204,766	140,862
3rd cousins	1,425,933	1,359,491	66,442
3rd cousins removed	2,628,002	2,531,016	96,986
4th cousins	1,464,685	1,424,859	39,826

Source: FOE database.

Figure A1: Sample of the FOE database



Notes: A sample section of the FOE database, showing linkages across 7 generations. The squares denote men, the circles women.

the extraordinarily large numbers of pairs of 2nd-4th cousins in table A2 can be seen in figure A1, which shows an illustrative fragment of the genealogy database. Average completed family size in England in the nineteenth century was around 3 adult children, but this varied enormously across families, and the bulk of adults in each generation came from larger than average families, so that average sibship size then was 6. Such demographic processes ensured large numbers of cousins, 2 nd cousins, etc. for adults in each subsequent generation.

Table A3 shows the social outcomes that are available by gender and lineage type. The numbers of any social outcome are much less than the numbers of people in the database because: (1) some outcomes are available for men only, (2) before 1914 a significant number of children die before reaching age 21, (3) for births before 1780 and after 1920 many social outcomes are not observable. But, as table A1 shows, when we consider the numerous pairs of

Table A3: Families of England, Social Outcomes

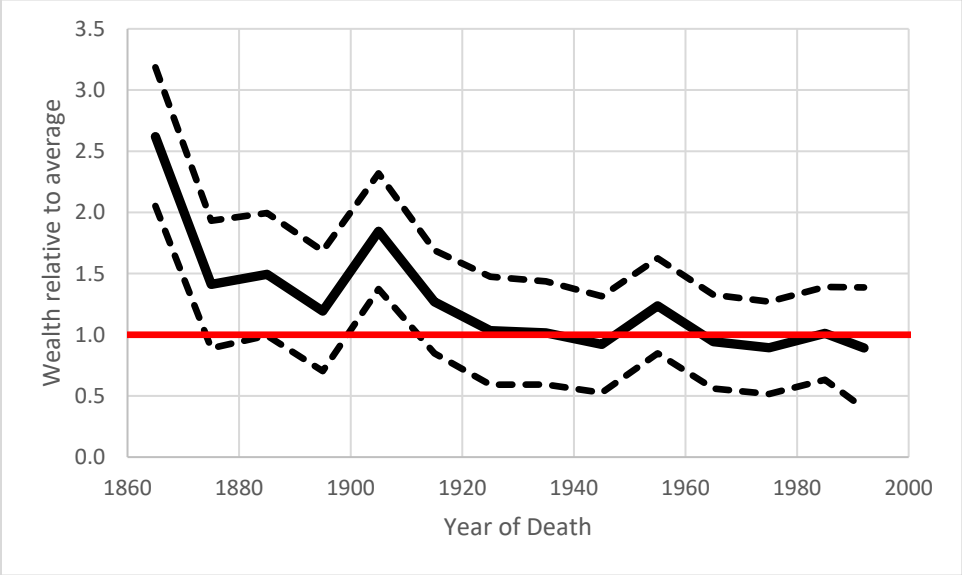
Outcome	Gender	All	General Lineages	Elite Lineages
Literacy	Both	6,700	6,700	0
Higher Education	Male	59,331	42,988	16,343
Occupational Status	Male	55,208	44,186	11,022
At Work 11-19	Both	31,961	26,891	5,070
At School 11-19	Both	31,961	26,891	5,070
Wealth at Death	Both	76,399	55,141	21,258
House Value	Both	24,746	19,864	4,882
Index of Multiple Deprivation	Both	24,898	19,958	4,940
Company Director	Both	32,973	27,112	5,861
Death age	Both	165,630	133,081	32,549

Source: FOE database.

relatives in table A2, there are many pairs observed with the same social outcome, that can be used to estimate the underlying parameters in equation (3). Thus for male occupational status, for men born 1860-1919 the database contains 241,941 pairs of observations.

How representative are the lineages in the FOE database of the general population in England and Wales? One test is average wealth at death in the general lineage 1858-1996 compared to average wealth of all deaths in England and Wales in these same years. Figure A2 shows this ratio by decade 1860-1990. As can be seen for the death decades 1920 and later, and thus the birth decades 1860 and later, the Families of England average lineages seem

Figure A2: Wealth at Death, Average FOE Lineages relative to England



Notes: Average wealth at death by decade of members of the average FOE lineages relative to England and Wales as a whole. Dashed lines show 95% confidence intervals.

representative of the general population in terms of wealth, and thus also in terms of other aspects of social status. For deaths before 1920, and thus births before 1860, average wealth in the general lineages is typically 50% higher than for the general population. The most likely explanation for this is that the processes that generated the rarer surnames used in these lineages were associated with somewhat higher status families in earlier centuries, but that over time slow but steady social mobility has brought these surnames to average social status by the time of births in the 1860s and later. There is also a possibility that lower status holders of the lineage surnames are less likely to appear in the records with a surname recognizable as belonging to these lineages.

Thus for the estimates of familial correlations using the average lineages for births 1860-1919 the sample used is then representative of the general population of England and Wales in its social outcomes. This conclusion is buttressed when we compare the average occupational status of men in the general lineages in the FOE database with the general population. For 16,639 men born in the FOE database who married 1880-1939, the average occupational status for average lineages was 37.2 on a scale 0-100, for those marrying outside the major city London. For grooms in a random sample of 591,000 Church marriages 1880-1939, transcribed by volunteers from the FreeReg organization, but drawn almost entirely from parishes outside London, the average occupation status on the same ranking was 35.5.[14] Thus again for births after 1860 the FOE sample looks very similar in status to the general population.

For men born 1780-1859 we can compare those in the FOE general lineages with the Freereg sample in terms both of occupational status, and of literacy. Men in the FOE general lineages marrying outside London 1837-79 had an average occupational score of 36.9.[21] In contrast 771,000 grooms in the Freereg sample in these same years had an occupational score of only 31.1. Thus in line with the average wealth evidence, men in the average lineages scored nearly 20% higher than the average man in this period. The marriage records, also record if men and women can sign their names at marriage. In the Freereg sample of 484,888 marriages 1837-79 with records on whether brides and grooms signed, 68% of men and 58% of women were literate. For men and women in the FOE general lineages marrying outside London 1837-79, 73% of men and 58% of women were literate. Here the members of the FOE average lineages show only a very modest elevation in social status compared to the average person in England.

Will the modestly elite status of the FOE average lineages for births 1780-1859 bias the estimates of social mobility rates in this period? The answer is that if we draw a sample from the population where the variance of outcomes is different than for the population as a whole we would potentially get a biased estimate of the heritability of traits, h^2. If the variance of the sample outcomes is higher, we will also get a more precise estimate of persistence. But whatever

sample of the population we start from, estimates of the level of persistence should be unaffected.

For the recent period 1999-2022, where we observe house values, the Index of Multiple Deprivation, and Company Directorships table A4 shows the distribution of the observations used here, in terms of geography and house values versus for the general population. The FOE dataset for individuals observed 1999-2022 has a geographic distribution that largely echoes that of the general population distribution, except for being less frequent in London. But the FOE dataset is composed, by design, of long-established English family lineages. London is the area of England with the largest proportion of the population of more recent immigrants. So it is expected that the frequency of the FOE families will be lower in London than for the general population.

The estimated house values observed in the FOE database, adjusted to 2017 prices, again are close to the average across region observed nationally in sales in 2017. The only location with a substantial difference is London, where the FOE house values are higher. But as noted London is the city with the largest share of population of recent immigration to England. Thus there is no reason to expect the FOE house values here to be similar to those of the London population as a whole. Overall house values in the FOE database are 6.6% higher than for England and Wales as a whole in 2017. This is a modest difference.

Table A4: FOE Observations 1999-2022 versus General Population

Region	National Population Share 2020[26]	FOE Address Data	FOE Address Share	National Average Dwelling Value 2017[27]	Average FOE Dwelling Value
London	0.151	1,914	0.078	£481,556	£576,957
East and South East	0.257	7,872	0.321	£306,534	£307,754
South West	0.096	2,883	0.118	£246,519	£254,057
Midlands	0.181	4,722	0.193	£183,773	£171,814
North	0.262	6,413	0.262	£152,318	£163,603
England	0.946	23,804	0.972	£240,325	£254,118
Wales	0.054	688	0.028	£151,672	£149,206

Source: FOE database.

The paper utilizes seven social outcome variables. These were constructed as follows.

1. Literacy. This is inferred for marriages 1754-1889 from whether the bride or groom signed the marriage register. These signature records extend more recently than 1889, but by then signature rates were very high, making the status information content of the measure low. Only a subset of county record offices in England have made images of marriage registers available on-line. So literacy is available only for a subset of men and women marrying in these years.

2. Higher education. This is an indicator variable with a value 1 if the person achieved a higher educational status. Complete records are available for attendees Oxford and Cambridge Universities (1600-2018), Durham University (1837-1939), University of London (1837-1939), the Royal Military Academy Woolwich (1790-1839) and the Royal Military College Sandhurst (1800-1946). Complete records are available for the UK Medical Registers, 1859-2017, UK, Civil Engineer Lists, 1818-1930, UK, Electrical Engineer Lists, 1871-1930, UK, Mechanical Engineer Records, 1847-1930, UK, Articles of Clerkship (attorneys), 1756-1874. This variable is constructed for men born 1600-1919. Many of these records for the years before 1940 are available on Ancestry.com.

In the correlation estimations, correlations are calculated for all pairs of relatives where the younger member of the pair was born 1780-1859 or 1860-1919. To avoid having the outcome variable being measured in completely different epochs for a family pair, pairs were utilized only where they were no more than one generation apart, except for grandfather-grandson.

3. Occupational Status. Occupations are recorded in the censuses of 1841-1911, the population register of 1939, marriage records 1837-2022, probate records, 1858-1939, ship passenger lists, 1870-1959, army enlistment records 1914-1918, and in the professional directories listed above.[23] Where a person had multiple statements of occupation at different times in the life course, the one closest to age 40 was employed. For those born 1780-1919 the very large set of occupation description strings were first assigned to one of 442

categories. Using a large set of 1.6 million marriage records 1837-1939 which give occupations for grooms, their fathers, and the father-in-law an occupational status score 0-100 was derived using Goodman's Association methodology. As can be seen in table 2, the occupational status index derived in this way shows strong parent-child correlations in both 1780-1859 and 1860-1919.[21]

4. House Value. This is estimated from the addresses recorded for people alive 1999 and later in the electoral roll. Since the Roll released 2002 and later records only those who consented to the release of their address, there are potential issues of selectivity. However, we see in table S4 above that the average house value recorded using the electoral register addresses closely approximates to that of England as a whole. There is thus no sign that higher status individuals are less likely to permit publication of their addresses in the Electoral Roll.

The Land registry shows house prices for all property sales 1995 and later. From this we construct an average dwelling value, normed to 2017 prices, for each Postal Code. British Postal codes on average cover only 40 houses. So this gives a good estimate of local house values for the person. Where there was no sale recorded for a postcode, we use the Council Tax Band to estimate the property value. Empirically the log of average house values produces higher correlations between relatives, so we use this measure.

We employ this measure only for men and women aged 24 or above, and living at a different address than their parents.

Property values differ substantially by region in England and Wales, as table S4 shows. London house values, for example, are more than 4 times those in the North of England. Since people show strong persistence by region across generations, we normalize all house values to remove regional effects, dividing England and Wales into 6 regions for this purpose.

5. Index of Multiple Deprivation. This index is a ranking of Lower Layer Super Output Areas (LSOAs), typically with a population of around 1500, in terms of a weighted average of measures of social deprivation. The index is available

by post code. The 2019 index, used here, is a weighted average, with weights indicated, of measures of: Income (22.5%), Employment (22.5%), Education (13.5%), Health (13.5%), Crime (9.3%), Barriers to Housing and Services (9.3%), Living Environment (9.3%). Since the index for Wales is constructed in a non-comparable way, we fix the level for that as the average for the English IMD for the North of England.

6. Company Director. Companies House in the UK maintains a register of the directors of limited companies. Limited companies include commercial enterprises, but also management companies for housing associations, as well as medical and legal practices. The register also includes people who subsequently resigned the position. We classify anyone alive 2015 and later, and aged 24 and above in 2022, as either a company director or not.

7. Status Modern. This is an index of social status which combines the three previous measures, using Principal Component Analysis, into an omnibus modern social status index. The correlation of the normed log house value and the index of multiple deprivation is 0.53, and with being a company director 0.24. The correlation between the index of multiple deprivation and with being a company director is 0.14. The correlation of the Status Modern index with these three components is: normed log house value, 0.86, index of multiple deprivation, 0.82, company director, 0.49.

Details of Estimation

Table A5 reports the details of the estimates of b, m, and h^2 from equation (3) for the 9 different outcomes that underly Figure 1 using the correlations in table 1. In just one case, Higher Education 1780-1859, the observation for 4th cousins was dropped as an outlier. Figures S3-S9 show the correspondence between genetic closeness and outcome correlation under the assumption that $m = 0.6$. In only one case, where IMD is the outcome, does the OLS line fitted to the data show a significant deviation from 0 at the vertical axis.

Table A5: Summary of Estimates of b, m, and h^2

Outcome	Birth Period	Pairs	Gender	b	m	h^2	R^2
Modern Status	1910-96	98,119	both	0.809	0.618	0.409	0.947
Normed L(house value)	1910-96	98,117	both	0.828	0.657	0.392	0.959
Index Mult. Deprivation	1910-96	98,711	both	0.736	0.471	0.355	0.942
Company Director	1910-96	219,552	both	0.802	0.604	0.131	0.656
Literacy	1725-1869	53,047	both	0.832	0.664	0.396	0.844
Occupational Status	1780-1859	104,854	male	0.747	0.494	0.940	0.963
Occupational Status	1860-1919	241,941	male	0.821	0.641	0.533	0.944
Higher Education	1780-1859	103,506	male	0.794	0.588	0.602	0.788
Higher Education	1860-1919	219,507	male	0.801	0.601	0.331	0.781
Average				0.795	0.591		0.859

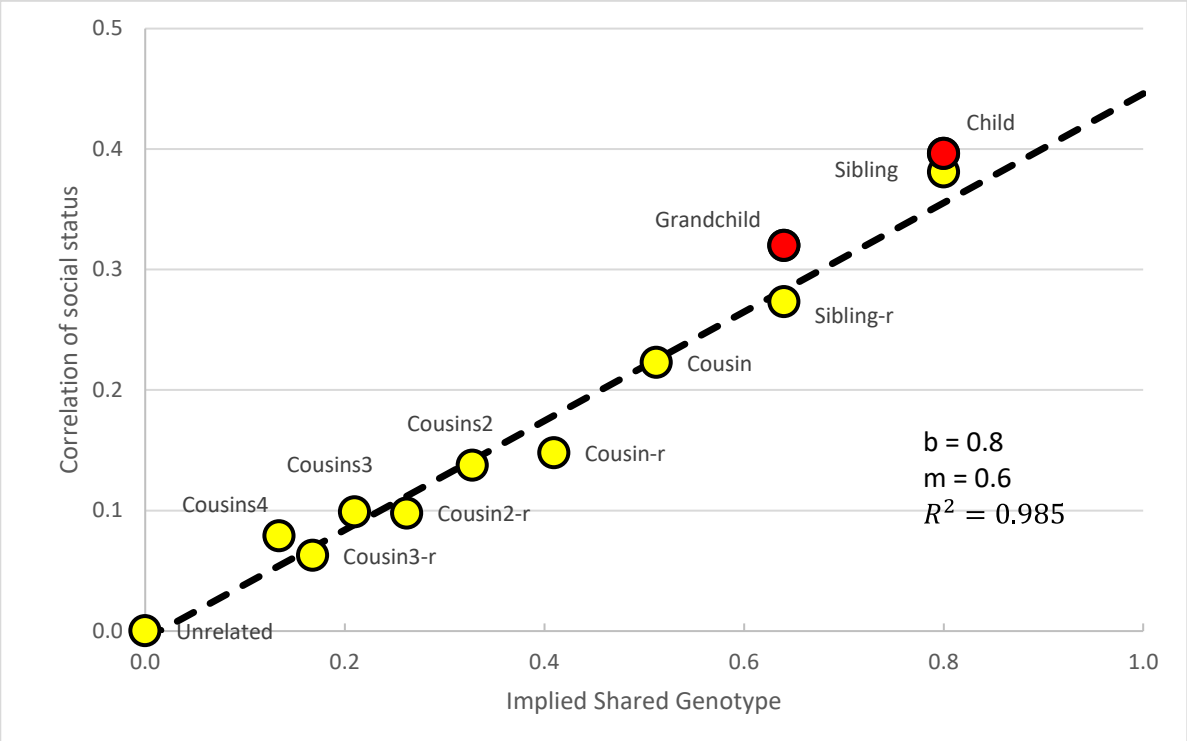
Table A6 reports the details of the estimate of the correlation, *r*, in latent social status between marital partners by period for marriages 1837-2022. In total there were 1,014,299 marriages with information on the occupational status of groom, father, and father-in-law. Bride occupational status was not used to estimate *r* because for most of this interval no occupation was listed for most brides. We can do such an estimation using the comparative correlation of bride literacy with her father in law's occupational status relative to father's status. This gives similar results to those reported in table A6.[21] Occupational status before 1940 was assigned using the HISCAM-GB scale.[24] Occupational status for 1940-2022 was assigned using the CAMSIS 1991 scale for Britain.[25] The estimated latent phenotype correlation for occupational status ranges by period from 0.77 to 0.82, with a mean of 0.79.

Table A6: Implied underlying phenotype correlation in marriage, 1837-2022

Period	N	Status Index	ρ_{gf}	ρ_{gfinl}	r
1837-1859	343,623	HISCAM	0.631	0.480	0.771
			(.001)	(.002)	(.004)
1860-1899	438,725	HISCAM	0.601	0.464	0.772
			(.001)	(.001)	(.004)
1900-1940	174,474	HISCAM	0.498	0.384	0.771
			(.002)	(.002)	(.004)
1940-1979	47,033	CAMSIS	0.424	0.346	0.816
			(.004)	(.004)	(.017)
1980-2021	10,444	CAMSIS	0.339	0.275	0.812
			(.009)	(.009)	(.045)

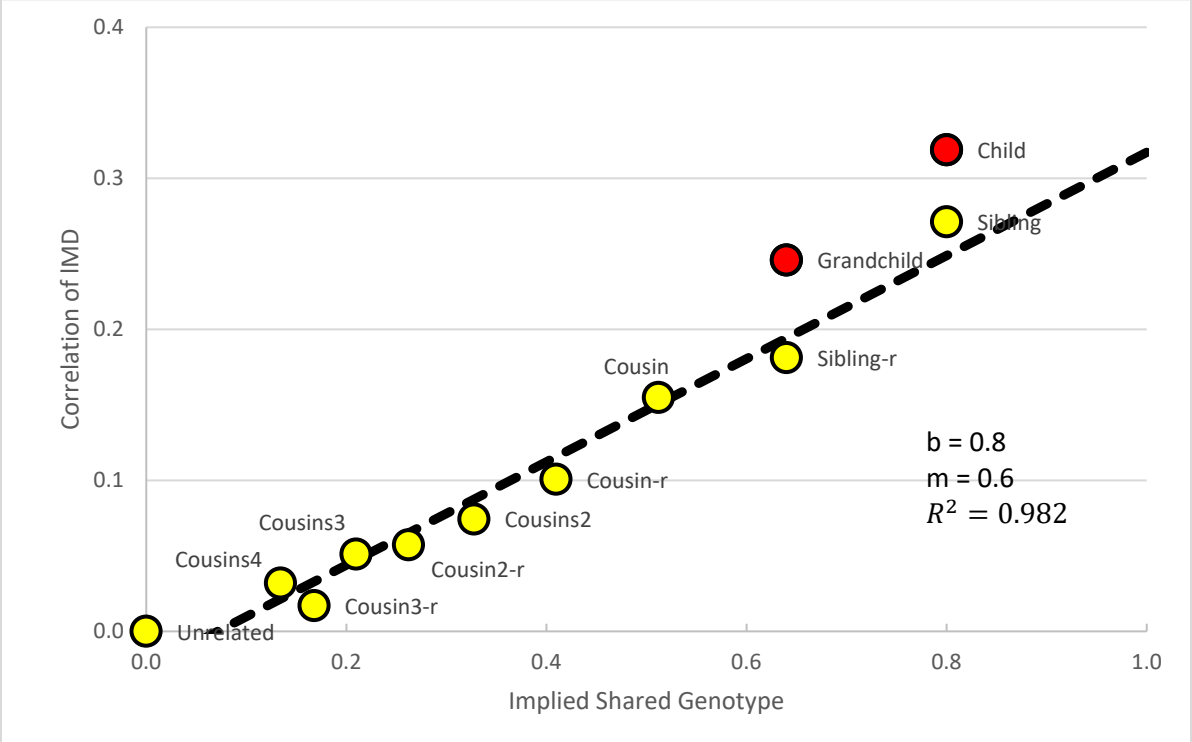
Notes: *gf* = groom-father, *gfinl* = groom-father-in-law. Standard errors in parentheses.
Source: (21), Tables 3 and 4.

Figure A3. Modern Status Correlations and Implied Shared Genotype



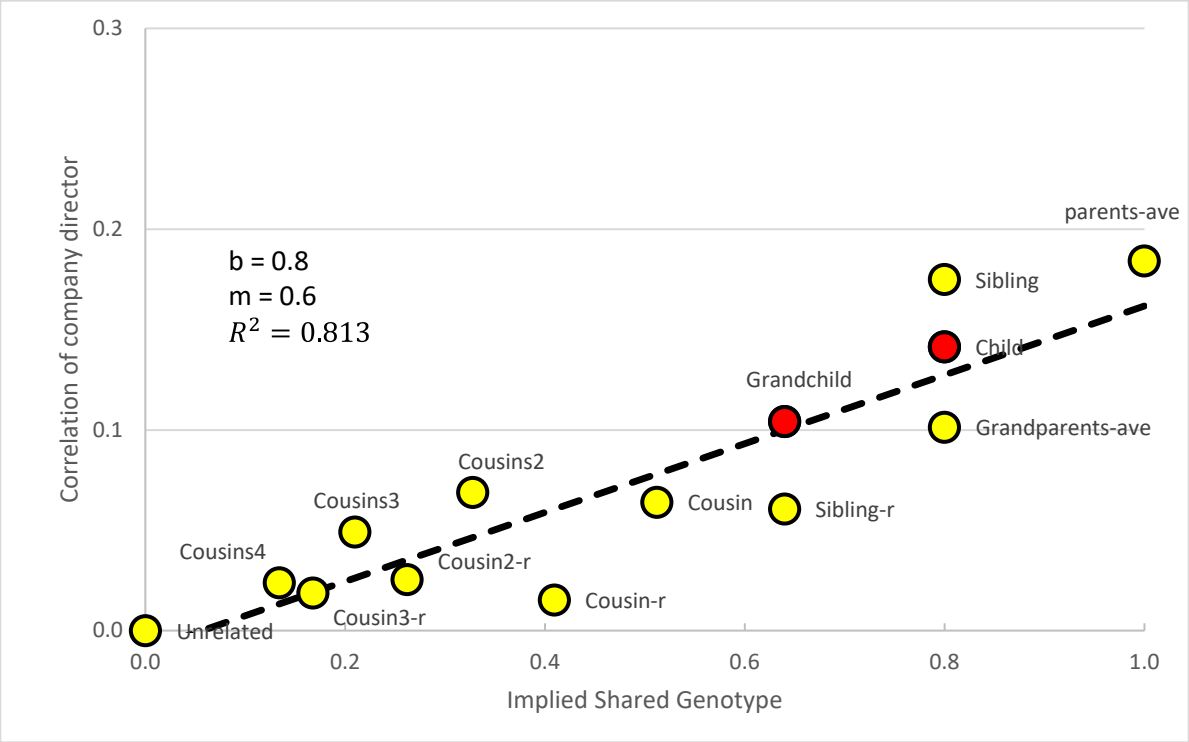
Notes: The dashed line shows the OLS fit to this data. The R^2 reported is for this fitted line. The child and grandchild correlations potentially deviate from this fitted line.

Figure A4. IMD Correlations and Implied Shared Genotype



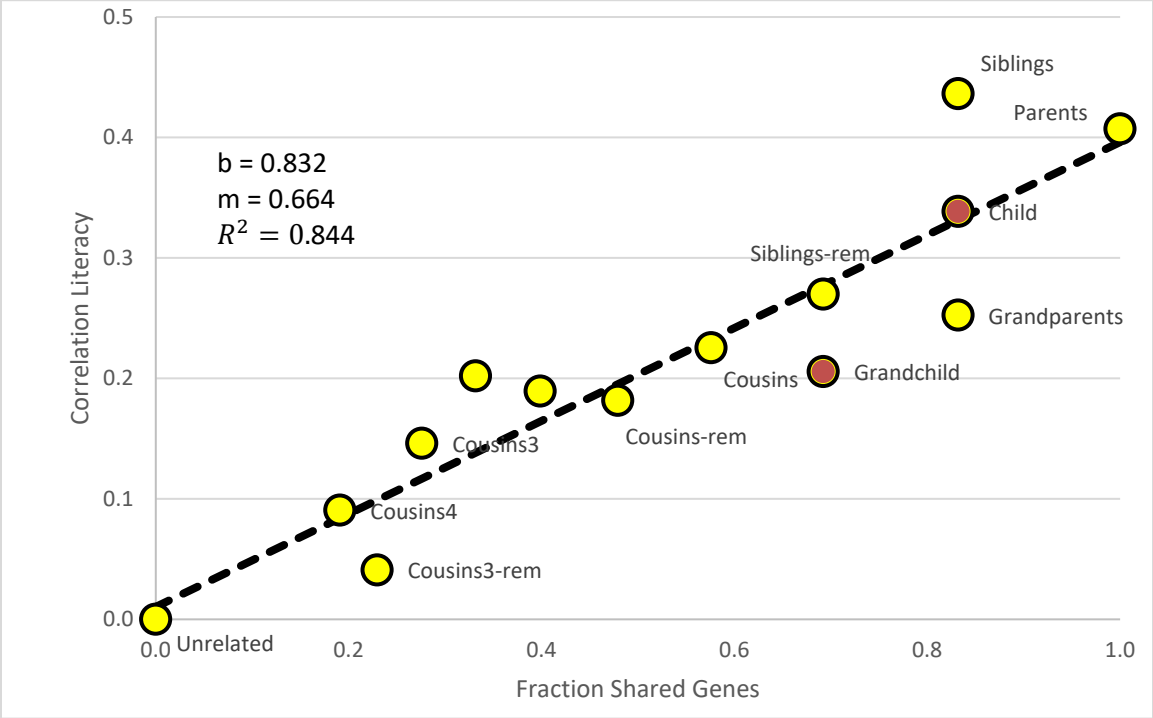
Notes: The dashed line shows the OLS fit to this data. The R^2 reported is for this fitted line. The child and grandchild correlations potentially deviate from this fitted line.

Figure A5: Company Director Correlations and Implied Shared Genotype



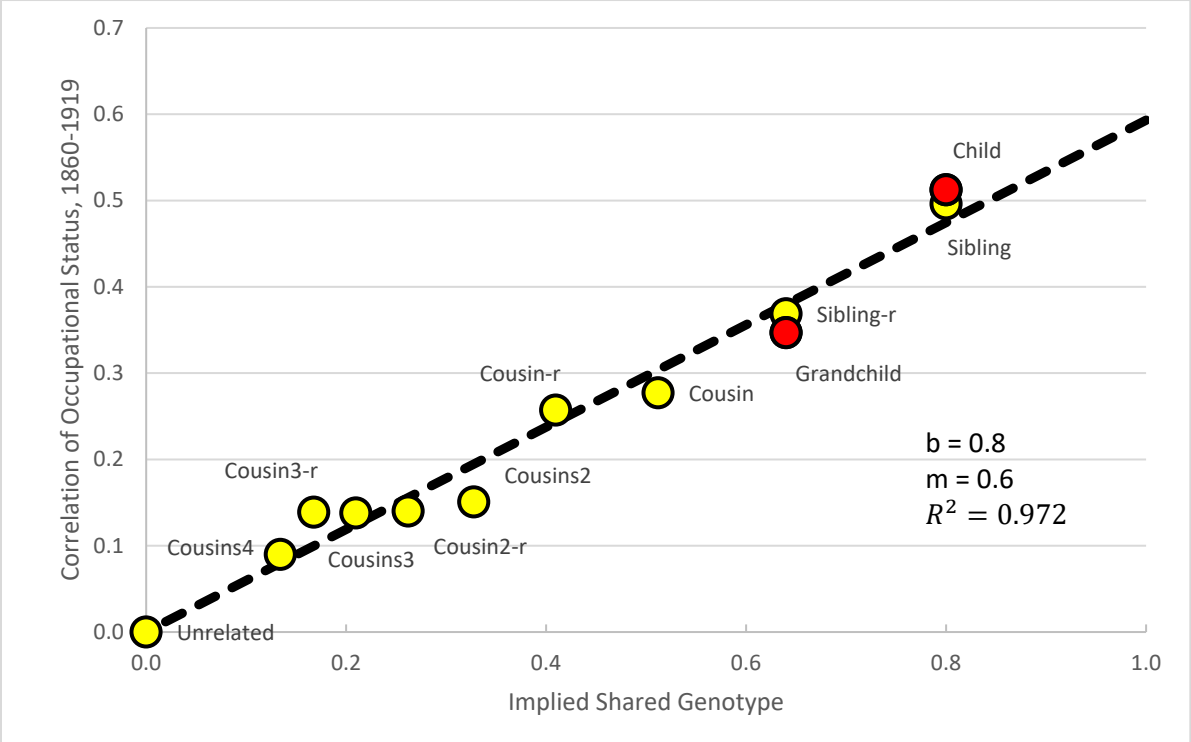
Notes: The dashed line shows the OLS fit to this data. The R^2 reported is for this fitted line. The child and grandchild correlations potentially deviate from this fitted line.

Figure A6: Literacy, births 1725-1869, and Implied Shared Genotype



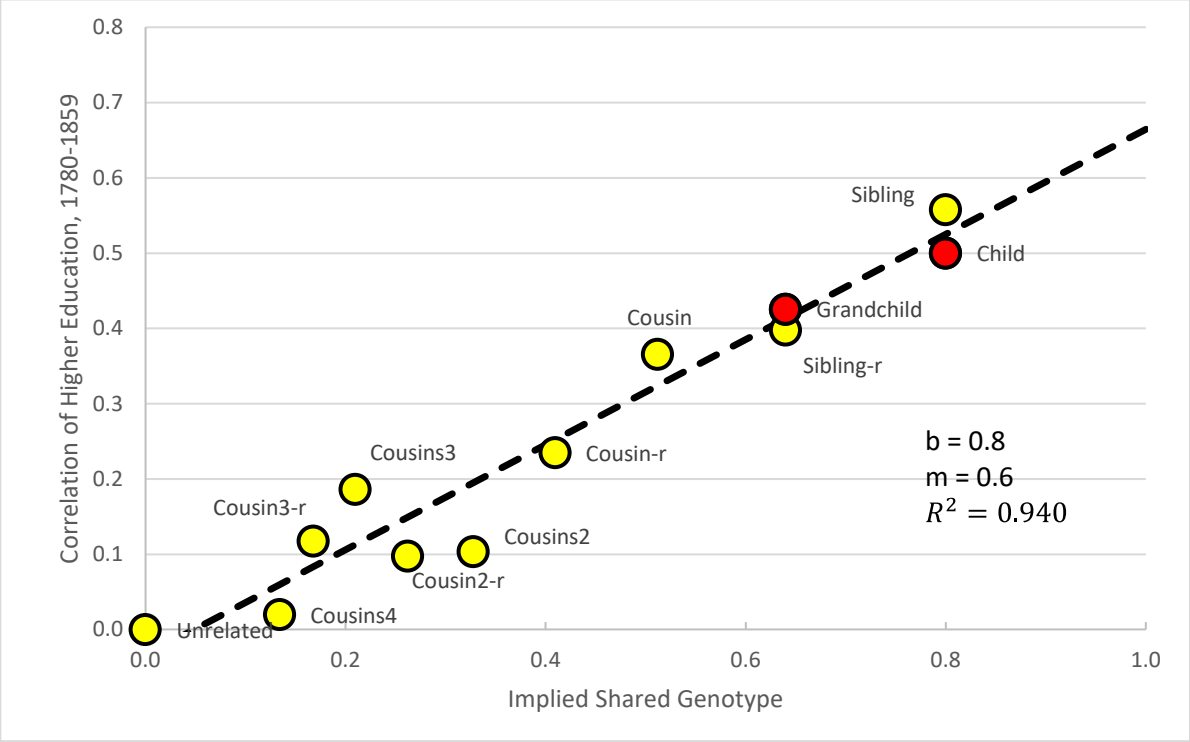
Notes: The dashed line shows the OLS fit to this data. The R^2 reported is for this fitted line. The child and grandchild correlations potentially deviate from this fitted line.

Figure A7: Occupational Status Correlations, births 1860-1919, and Implied Shared Genotype



Notes: The dashed line shows the OLS fit to this data. The R^2 reported is for this fitted line. The child and grandchild correlations potentially deviate from this fitted line.

Figure A8: Higher Education Correlations, births 1780-1859, and Implied Shared Genotype



Notes: The dashed line shows the OLS fit to this data. The R^2 reported is for this fitted line. The child and grandchild correlations potentially deviate from this fitted line.