Nature Versus Nurture in Social Outcomes. A Lineage Study of 62,000 English Individuals, 1750-2016

Gregory Clark, UC Davis and University of Southern Denmark, Neil Cummins, LSE

Economics, Sociology, and Anthropology are all dominated by the belief that while physical traits like height are mainly determined by genetics, child social outcomes are principally created by parental investment and community socialization. This paper shows that with just observational data of social outcomes and knowledge of the degree of relatedness of people in an extensive lineage, we can test whether additive genetic inheritance of social traits can be rejected. Using a lineage of 62,000 people born in England 1750-2016, we show that the pattern of correlations between people in the lineage is mainly consistent with additive genetic inheritance of social status, as is seen with height. The high persistence of status over multiple generations requires a high degree of assortative mating. We show evidence that marriage in the lineage indeed does show a high degree of assortment based on some underlying trait.

It is widely believed that while social status - measured as occupational status, income, health, or wealth – is correlated between parents and children, this correlation is driven by parental investments in children, or by cultural transmission.[1] This belief has profound influence on peoples' perception of the fairness of society, and of the need for government intervention in the lives of children.

It might seem that just from observational data it would not be possible to measure the relative importance of genetics versus nurture in generating inheritance of abilities. The pathways through which nurture and culture can operate are many and varied. The constraints imposed by theory on possible effects are minimal. Small interventions could have profound effects. Mothers can matter more than fathers. Early life experiences can outweigh later ones, or vice versa.

[1] Studies of adoptions and of twins suggest that this belief is not well founded. Such studies suggest that genetic transmission explains the majority of social outcomes, but leave room for substantial social influences. See, for example, Sacerdote, 2007.

In contrast, genetic inheritance as a mechanism, unlike cultural inheritance, has implications for the pattern of correlation across relatives of various degrees of genetic relatedness. A cultural mechanism of transmission of factors leading to higher status is agnostic, for example, on how the correlation of longevity between siblings compares with that between each parent and child. The correlation between siblings could be much lower than that between parent and child, if each child gets some combination of parent characteristics, plus some independent random shock. Or it could be much higher if children get some combination of parent characteristics plus a shared shock from the childhood environment.

But genetic transmission implies that the sibling correlation will have a set value relative to parent-child correlations. What these correlations are depends on the degree and nature of assortative mating. Below we set out all the predictions of additive genetic inheritance of social status within a lineage. We then test whether observed correlations are close to those predicted by additive inheritance for educational status, wealth, employment status, and longevity.

The Predictions of Additive Genetic Inheritance

The predictions draw below are based on a number of assumptions (see Nagylaki, 1978)

1. The traits in question are controlled by many loci in the genome, each of which makes a small contribution.
2. There is an absence of important dominance and epistasis effects.
3. Genes and environment are uncorrelated, or the environment has little independent impact on outcomes.

As noted above, to set out the predictions of additive genetic inheritance across family relationships within a lineage we also need to consider the nature of matching in marriage. Suppose, for example, that matching is based only on the underlying genetic characteristics that determine social status.[2] Suppose the correlation in genetics is *m*. Then the sibling correlation will be

$$h^2 \frac{1+m}{2}$$

[2] That is, there is no correlation based on the accidental elements of the phenotype.

Table 1: Phenotype Correlations for a Genetically Inherited Trait

Relative	Matching on Genotype	Matching on Phenotype
Parental	$h^2 m$	r
Mid-parent - child	h^2	h^2
Single parent – child	$h^2 \frac{1+m}{2}$	$h^2 \frac{1+r}{2}$
Siblings	$h^2 \frac{1+m}{2}$	$h^2 \frac{1+m}{2}$
Uncles/Aunts – child	$h^2 \left(\frac{1+m}{2}\right)^2$	$h^2 \left(\frac{1+m}{2}\right) \frac{1+r}{2}$
Grandparent – child	$h^2 \left(\frac{1+m}{2}\right)^2$	$h^2 \left(\frac{1+m}{2}\right) \frac{1+r}{2}$
Cousins	$h^2 \left(\frac{1+m}{2}\right)^3$	$h^2 \left(\frac{1+m}{2}\right)^2 \frac{1+r}{2}$
Great Grandparent – child	$h^2 \left(\frac{1+m}{2}\right)^3$	$h^2 \left(\frac{1+m}{2}\right)^2 \frac{1+r}{2}$
Second Cousins	$h^2 \left(\frac{1+m}{2}\right)^5$	$h^2 \left(\frac{1+m}{2}\right)^4 \frac{1+r}{2}$

Note: *m* is the correlation of parents in genotype, *r* the correlation in phenotype.

where h^2 is the hereditability of the trait, and will be the same, to a first order, as the parent-child correlation, as shown in table 1.

This result concerning siblings would not be expected if we postulate that social outcomes are mainly the product of family environments or family resources. Siblings share the same family environment and resources. The correlation of their environment will be close to 1. Parents had a family environment in their childhood which is correlated with that of their children. But there has to be variation in family environments across generations. Suppose that social outcomes, *y*, are the product of the family environment, *z*, and some random component, *u*, so that

$$y_t = z_t + u_t .$$

Since, as we will see below, there is constant regression to the mean of social outcomes, where the correlation of parent and child outcomes on any measure of status is typically less than 0.5, it must be that

$$z_t = bz_{t-1} + e_t .$$

Again *e* is a random component that must exist to keep the dispersion of *z* constant across generations. With this structure the average correlation of social outcomes between parent and child will be

$$\hat{\beta} = b\frac{\sigma_z^2}{\sigma_z^2+\sigma_u^2}$$

The correlation between siblings, sharing a common environment, will on average be

$$\hat{\rho} = \frac{\sigma_z^2}{\sigma_z^2+\sigma_u^2}$$

We will see below that b would have to be in the range 0.7-0.8 to fit the observed pattern of correlations between parents, children, grandchildren etc. So the sibling correlations on such a transmission mechanism would be 25-40% higher than the parent child correlations.

Another powerful implication of the additive genetic model is that the familial correlation is a function of the proportion of genes two relatives share by descent, as shown also in table 1 for relatives up to second cousins. These correlations were established by Fisher in a famous 1918 paper.[3] In this case the phenotypic status value of an individual in any generation, can be modelled as

$$y_t = x_t + u_t , \quad (1)$$
$$x_t = bx_{t-1} + e_t \quad (2)$$

where *x* is the genotype value, e_t and u_t are random shocks, and

$$b = \frac{1+m}{2}$$

$$h^2 = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2}$$

Elsewhere when we estimate such a model for the inheritance of social features such as educational status, occupational status, and wealth we consistently find values of b in the range 0.7-0.8.[4] For this to fit with genetic transmission it must be that m, the correlation of genes determining social status, is in the range 0.4-0.6 between marital partners. If mating were random, so that $m = 0$, b would be constrained to be ½.

If, however, the matching of parents is through the phenotype, the predicted set of correlations becomes a more complex. Suppose the correlation in phenotype across parents is *r*. Now the single parent-child correlation depends on r, the correlation of phenotype between parents, and not on m, the correlation of genotype. Now also

$$m = rh^2$$

so that m will be less than r.[5] Column 3 of table 1 shows also the pattern of correlations in this case. Now, for example, the correlation between siblings will be less than that between sibling and parent. But the intergenerational pattern will be very similar.

If parental matching is through the phenotype then the underlying correlation of the parents genotypes, *m*, will necessarily be low. Empirical evidence on the correlation of phenotypes, r, suggests these are relatively modest. Table 2, for example, shows measured correlations of married couples by a variety of characteristics. These correlations are typically in the range 0.2-0.5, suggesting that m would be in the range

[3] Fisher, 1918. Nagylaki , 1978, gives a simpler derivation of this result.
[4] Clark et al., 2014, and Clark and Cummins, 2014, 2015.
[5] With matching on genotype, in contrast, $m = h^2 r$.

0.1-0.25 only, if matching is through the phenotype.

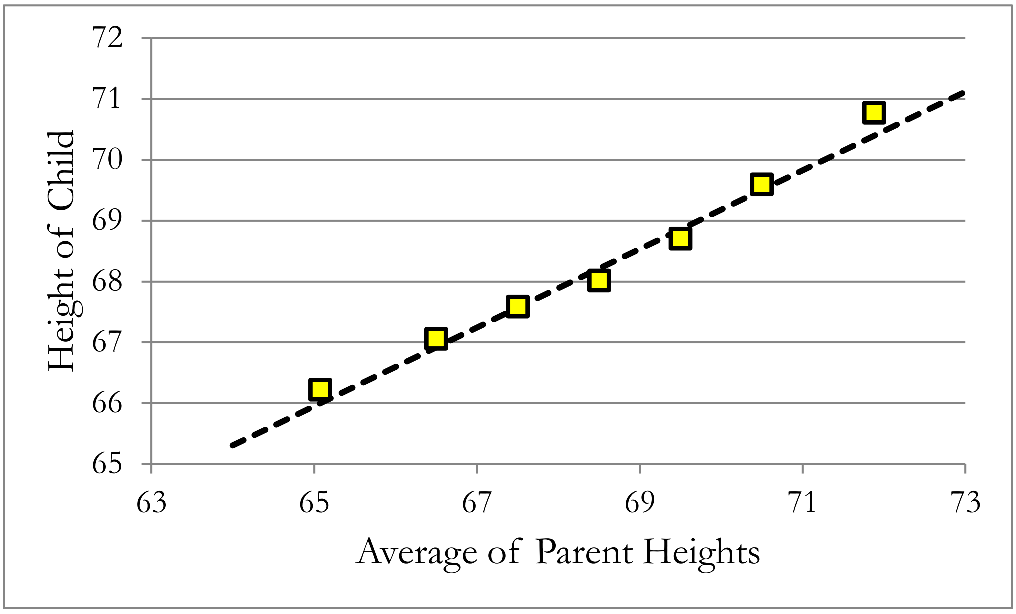
In modern high-income societies height is largely genetically inherited, and is the outcome of at least 300 genes each of which exerts very modest influence. Height inheritance certainly meets the second stipulation for the correlation pattern set out in table 1. Regression to the mean is indeed, at least to a first approximation, linear across the whole range of parent heights as figure 1 shows. The figure shows the heights of parents and children in Galton's pioneering study of the inheritance of heights. Height inheritance thus fits well this additive genetic model. Also, plausibly, mating actually sorts on the height phenotype rather than the underlying height genotype. So for height the long run intergenerational correlation should be close to 0.5.

Table 3 shows the correlation of heights between relatives recorded in a modern health study of a district in Norway 1984-6 (Tambs et al., 1992). The spousal correlation, measuring only 0.18, was taken as correct. From this, and the parent-child correlation, h^2 is estimated at 0.73. If sorting was on the phenotype this implies an underlying correlation of height genotype between parents of only 0.13. Knowing *r*, h^2, and *m* we can predict the other correlations between relatives – siblings, grandparent, avuncular, and cousins – and compare this with the measured correlation. Except for cousins the model predictions correlations are close to the actual. But for cousins the sample size is very small, and the correlation consequently measured with much potential error.

That in turn implies that the long run intergenerational correlation $\left(\frac{1+m}{2}\right) = 0.56$. When we apply this in table 3 to calculate correlations across siblings more distant kin we can see that the model works well, except for cousins.

A further implication of additive genetic inheritance of social status will be that to produce the observed patterns of very slow mobility mating must be assortative with respect to the genotype that generates social phenotypes, and not with respect just to the phenotype. In that case the correlation in genotypes will be greater than the observed correlation in phenotypes. The need for closer genetic correlations of parents implies also that the relevant intergenational correlations will be those of the second column of table 1, corresponding to equations (1) and (2) above as describing the intergenerational mobility process.

Figure 1: Linearity of Regression to the Mean with Height



Source:

Table 2: Phenotypic Correlations between Spouses

Characteristics	Correlation	Source
Height	0.29	McManus and Mascie-Taylor, 1984
Education	0.50	Watkins and Meredith, 1981
Income	0.34	Watkins and Meredith, 1981
Occupational Status	0.12	Watkins and Meredith, 1981
IQ	0.20-0.45	Mascie-Taylor, 1989
BMI	0.28	Abrevaya and Tang, 2011
Personality Traits	0.15	Mascie-Taylor, 1989

Table 3: Height Correlations in Norway, 1984-6

Relation	Number	Measured Correlation	Predicted Value	Fitted Value
Spouses	24,281	0.179	r	(0.179)
Parent-Child	43,613	0.430	$h^2 \frac{1+r}{2}$	(0.430)
Siblings	19,168	0.453	$h^2 \left(\frac{1+m}{2}\right)$	0.412
Grandparent-Child	1,318	0.250	$h^2 \left(\frac{1+m}{2}\right) \frac{1+r}{2}$	0.243
Avuncular	1,218	0.217	$h^2 \left(\frac{1+m}{2}\right) \frac{1+r}{2}$	0.243
Cousins	112	0.209	$h^2 \left(\frac{1+m}{2}\right)^2 \frac{1+r}{2}$	0.137

Source: Tambs et al., 1992.

Another human trait which is almost entirely genetically inherited is Total Ridge Count, which is the number of ridges, measured in a standardized way, on all 10 digits. Table 4 shows the familiar correlations for this measure for a sample of 200 husbands and wives and their children, including disproportionately twins. The familial correlations fit well with the Fisher predictions.

Table 4: Inheritance of Total Ridge Count

Relationship	Number of Pairs	Correlation (s.e.)	Predicted
Mother-Child	405	0.48 (.02)	$h^2\left(\frac{1+m}{2}\right)$= 0.50
Father-Child	405	0.49 (.02)	$h^2\left(\frac{1+m}{2}\right)$= 0.50
Husband-wife	200	0.05 (.03)	*m =0.05*
Sibling-Sibling	642	0.50 (.02)	$h^2\left(\frac{1+m}{2}\right)$= 0.50
Monozygotic Twins	80	0.95 (.01)	h^2=0.95
Dizygotic Twins	92	0.49 (.04)	$h^2\left(\frac{1+m}{2}\right)$= 0.50

Source: Holt, 1961.

How can we test if mating is assortative based on underlying genetics as opposed to individual phenotype characteristics? Suppose, for example, that the various characteristics associated with status – education, wealth, occupation and longevity – all derive from the same underlying genetics, so that

$$y_i = x + u_i , \tag{3}$$

Let M indicate males and F females. Then if we regress for couples

$$y_{iM} = \lambda y_{iF} + \varepsilon_i , \tag{4}$$

we can measure the phenotype correlation, λ, but not the genotype correlation. Suppose, however, we have multiple measures of the phenotype of the parents, wealth and education for example, y_i and y_j. In this case if we regress

$$y_{iM} = \lambda y_{iF} + \varepsilon_i , \tag{5}$$

but instrument for y_i with y_j , then $E(\hat{\lambda}_{IV})$ = correlation of the underlying characteristic, x.

For women in earlier generations we often do not have status. In this case use status of a brother. We can regress status of the husband on that of the brother and instrument for characteristic i of the brother with characteristic j. For the rare surnames database in England (high wealth surnames) discussed below we have information on wealth and education of sons in law, and of the married brothers of the wives.

Another implication of additive genetic inheritance is that elements that can make a difference to the social treatment of children, such as birth order, or the gender composition of siblings, will have no effect on social outcomes. In particular, family size itself will have no effect on social outcomes. In the modern world it is very hard to establish from observational data the effects of family size on child outcomes. Family size is correlated with status. But for marriages before 1880 in England there is an absence of any such correlation, and indeed good evidence that family size was also not a choice by the parents, but the result of biological accidents uncorrelated with social status.

A further implication of the structure described by equations (1) and (2) is that again if we estimate the correlations between relatives in table 1 for one measure of social status, but instrument with another measure, then when the OLS correlation is

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

the IV correlation will be predicted to be

$$\left(\frac{1+m}{2}\right)^n$$

Thus with multiple measures of social status we can estimate again whether the lineage data is consistent with additive genetic inheritance.

Testing Genetic Inheritance against an English Lineage

We propose to test the various additive genetic inheritance predictions outlined above using a lineage under construction for English families 1750-2018, that will show all familial links, plus a variety of social outcomes. So far we have the basic information for 62,000 people in the lineage. Figure 2 shows a sample lineage for one couple and some of their descendants from the database. The figure illustrates the richness of the set of family links that the database contains. In this case the lineage covers 7 generations. But what matters is the set of social outcomes we can associate with the members of the lineage. Table 5 summarizes the data currently available.[6] The social status indicators we have are wealth at death, occupation, educational attainment, schooling and training 11-20, age at death, and first name.[7]

Wealth at Death: The Principle Probate Registry records whether someone was probated, and the value of their estate for all deaths in England 1859-2016. 1799-1858 there is information on wealth at death of higher status individuals through the records of the Prerogative Court of Canterbury. This information is the most comprehensive and unusual outcome that we have for this lineage.

Occupation: The censuses of 1841-1911 record occupations. In addition the 1939 population register also records occupations, and for those dying before around 2010 such records are available. Thus we can estimate adult occupations for the cohorts born 1920 and before. The absence of female employment in earlier years means that we can only get meaningful occupation rankings for men. We can rank occupations in a number of ways. One is the average wealth at death of men probated in 1858. We can then rank occupational status by the logarithm of this wealth. Another ranking of occupational status is the child mortality rate associated with each occupation as recorded in the 1911 census. That census asked mothers both the number of birth they had experienced, and the number of child deaths. From this was calculated a standardized child mortality rate by husband's occupation, which ranged from 68/1000 (Clergy, Church of England), to 239/1000 (Dock Laborers). Since 1911 is closer to the center of the occupational data we use the 1911 ranking of occupations, expressed as surviving children per 1,000 born.

Birth certificates record the occupation of fathers, and from 1995 on that of mothers also. Marriage certificates record the occupations of bridge and groom, and of their fathers. The time and money cost of collecting marriage certificates for everyone born post 1886 makes this infeasible. But we propose to collect this information for a

[6] We expect to be able to add much more information on occupations and schooling.
[7] In recent years in England first names are a strong indicator of social status.

Figure 2: An Illustrative Portion of a Family Lineage, Lineage Database

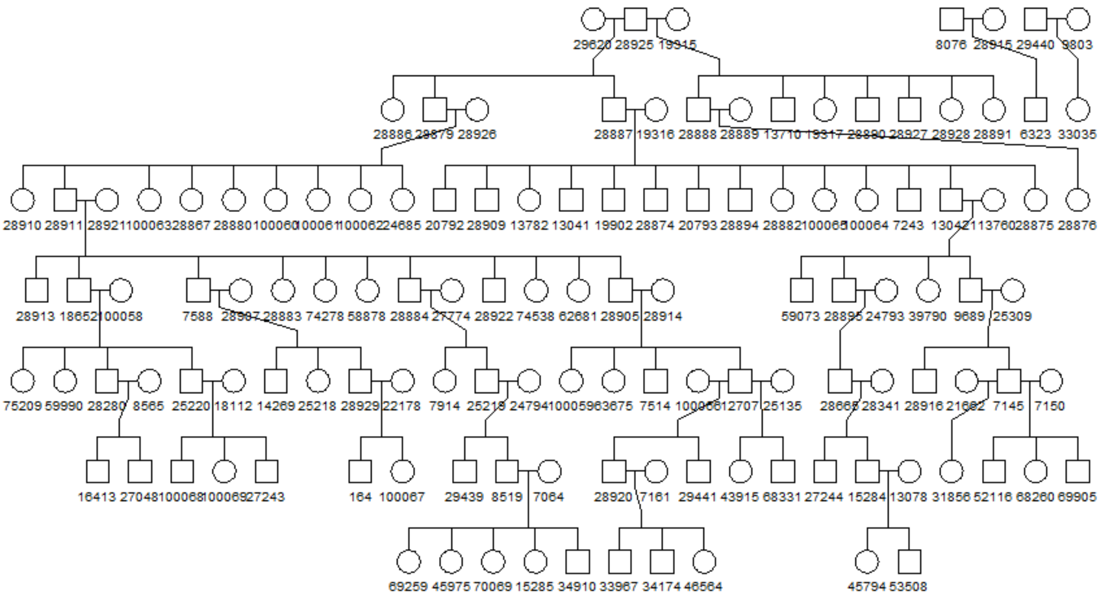


Table 5: Information Available on Relatives Characteristics, males

Relationship	All	Higher Education	Wealth at death	Occupation
Father-Son	17,557	10,101	8,315	5,003
Brothers	21,154	11,113	5,887	4,765
Grandson	12,996	7,021	5,655	3,203
Uncle-Nephew	34,532	17,182	12,857	6,821
Great-Grandson	8,673	4,154	2,983	1,481
Uncle-GNephew	14,171	6,268	3,882	3,440
Cousins	17,074	8,487	6,825	3,519
GG-Grandson	4,845	2,023	1,022	347
Uncle-GGNeph.	8,500	3,141	1,510	977
Uncle-GGGNeph				124
GGG-Grandson	2,309	1,029	217	34
2nd Cousins	12,307	5,220	4,319	1,835
3rd Cousins	5,145	1,710	1,167	433

subset of people who are in the fourth and fifth generations of family trees, to examine long run transmission of occupational status.

Schooling and Training: The censuses of 1841-1911, and 1939 register, record whether anyone aged 11-20 is still attending a school or in some kind of training, which gives us a measure of education for the earlier year. This measure is available for both men and women.

From a previous NSF project we have a database of all students who attended Oxford or Cambridge, 1750-2015. But this constitutes only 1-2% of each cohort nationally. Complete records are available for attendees at the Royal Military Academy Woolwich (1790-1839) and Royal Military College Sandhurst (1800-1946). Complete records are available for the UK Medical Registers, 1859-2015, UK, Civil Engineer Lists, 1818-1930, UK, Electrical Engineer Lists, 1871-1930, UK, Mechanical Engineer Records, 1847-1930, UK, Articles of Clerkship, 1756-1874. From all these measures we can construct indices of educational attainment for people in the database born before 1900. This measure, however, again applies only to men.

Location: From the electoral census of 1998 we have the address where adults are living in 1998, from which we can infer the property value of their dwelling from later real estate listing services.

Children's Names: Children's first names are a good proxy for family social status in modern generations. Using records of Oxbridge attendance and property values we can assign status measures to parents based on their child name choices.

The first pattern we would expect with genetic inheritance of social status would be linearity in regression to the mean of social status from fathers to sons, all across the status distribution. For wealth, shown in figure 3, we do see exactly the predicted pattern. There is no asymmetry between mobility at the top or bottom of the distribution, and no difference in mobility at the extremes compared to the center. This is surprising given that the social processes are very different at the top and bottom of the wealth distribution. At the bottom children receive no bequest, but tend to accumulate some wealth. At the top children receive substantial bequests and then have to decide how much to spend in their lifetimes.

For occupational status, using the 1911 morality ranking, there is also near linearity. This is shown in figure 4.

Figure 3: Son Wealth relative to Father Wealth, by decile

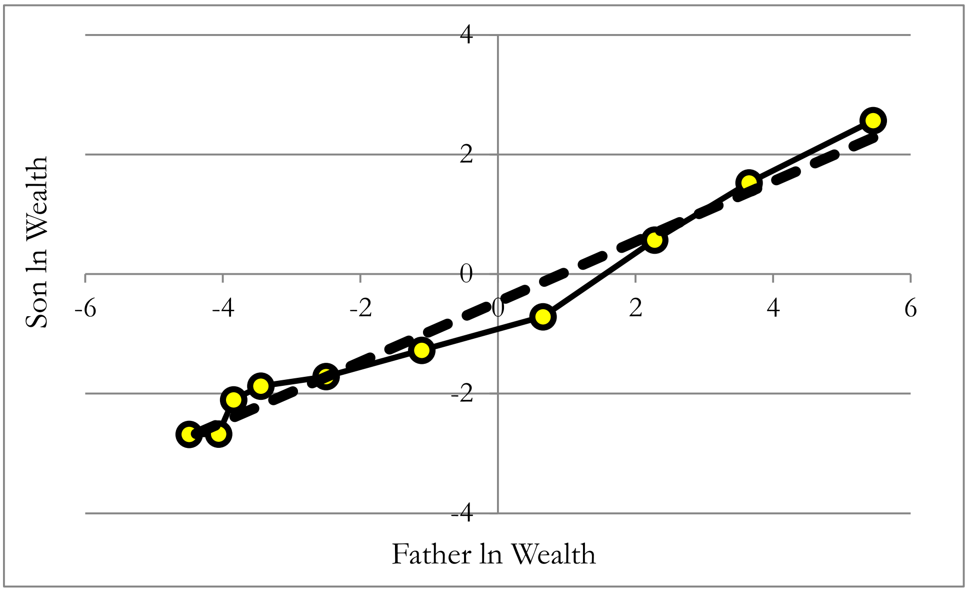
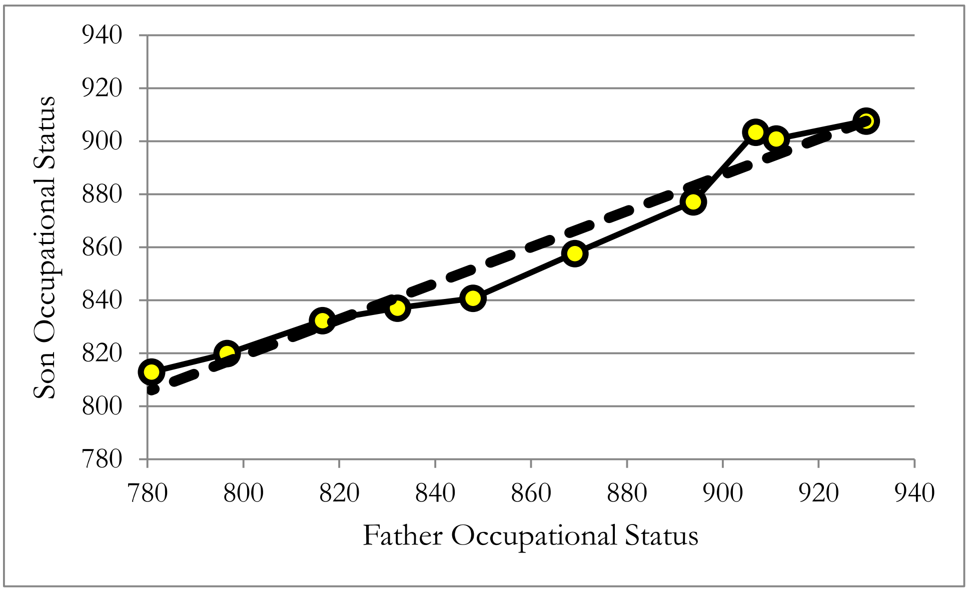


Figure 4: Son Occupational Status relative to Father's Status, by decile



Note: The occupational status score is the survival rate of children (per 1,000) by occupation in 1911.

The next prediction of multifactorial additive genetic inheritance would be the set of correlations between relatives in the pattern shown in table 1. Table 6 shows the correlations of relatives of different degrees on three of the status characteristics, as well as the closeness of the genetic connection. Males only were used since for much of our period most women did not have formal educational status, or occupations. For wealth and occupational status we can even get estimates of the correlation in wealth for 3rd cousins. Their genetic correlation would be

$$\left(\frac{1+m}{2}\right)^{7}$$

Assuming m is 0.6 (see below), this would be 0.21, while for brothers it will be 0.8.

Table 1 implies that the logarithm of the intergenerational correlation of status on any measure will be linear with respect to genetic distance, as shown in figure 5, if the transmission is through additive genetic effects.

Figure 5: Expected pattern of intergenerational status correlations with genetic distance

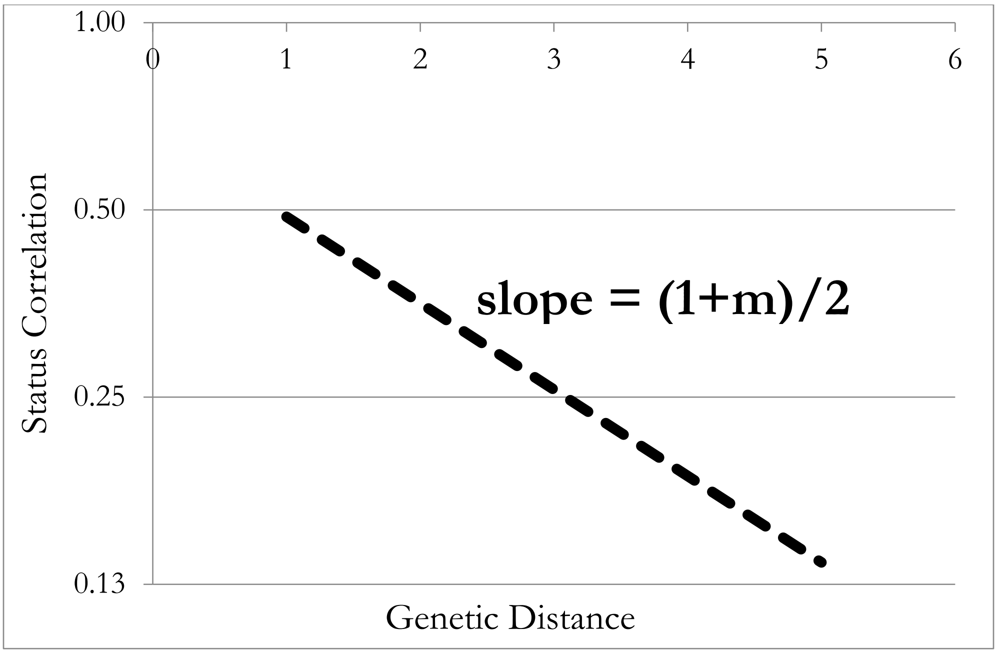


Table 6: Intergenerational Correlations, Males

Relationship	Genetic Distance	Wealth	Occupational Status	Higher Education
Father-Son	1	0.628 (.012)	0.703 (.012)	0.352 (.016)
Brothers	1	0.553 (.013)	0.697 (.016)	0.329 (.018)
Grandson	2	0.520 (.016)	0.639 (.019)	0.246 (.020)
Nephew	2	0.465 (.019)	0.642 (.019)	0.259 (.018)
Great-Grandson	3	0.434 (.022)	0.566 (.032)	0.163 (.029)
Great-Nephew	3	0.432 (.035)	0.598 (.028)	0.210 (.020)
Cousins	3	0.385 (.025)	0.638 (.026)	0.235 (.027)
GG-Grandson	4	0.325 (.035)	0.420 (.070)	0.078 (.059)
GG-Nephew	4	0.314 (.051)	0.602 (.052)	0.142 (.037)
GGG-Grandson	5	0.238 (.078)	0.239 (.216)	0.186 (.161)
Second Cousins	5	0.294 (.053)	0.524 (.056)	0.145 (.041)
GGG-Nephew	5	-	0.529 (.117)	0.047 (.055)
Third Cousins	7	0.170 (.053)	0.446 (.098)	0.119 (.060)

Notes: For educational status only men born before 1915 used (because of absence of information for later cohorts on some higher educational categories).

The first issue raised above was the relative correlation of siblings versus parent-child in outcomes. These correlations are shown in the second and third lines of table 6. Figure 6 shows the relative father-son and brother correlations for the three attributes in table 6, as well as for lifespan. As can be seen the relative sibling compared to father-son correlations are strongly consistent with status mainly being transmitted genetically.

Figure 7 plots all the relative correlations in table 6 for wealth, against the expected genetic distance between relatives. Since genetic distance has a predicted multiplicative effect in reducing correlations the correlation is shown on a logarithmic scale on the vertical axis. As can be seen the correlations for the entire set of relatives fall very close to the predicted linear pattern, with $(1+m)/2$ = 0.79 (the standard error on this estimate is 0.013).

Figure 6: Comparative Father-Son and Brother Correlations

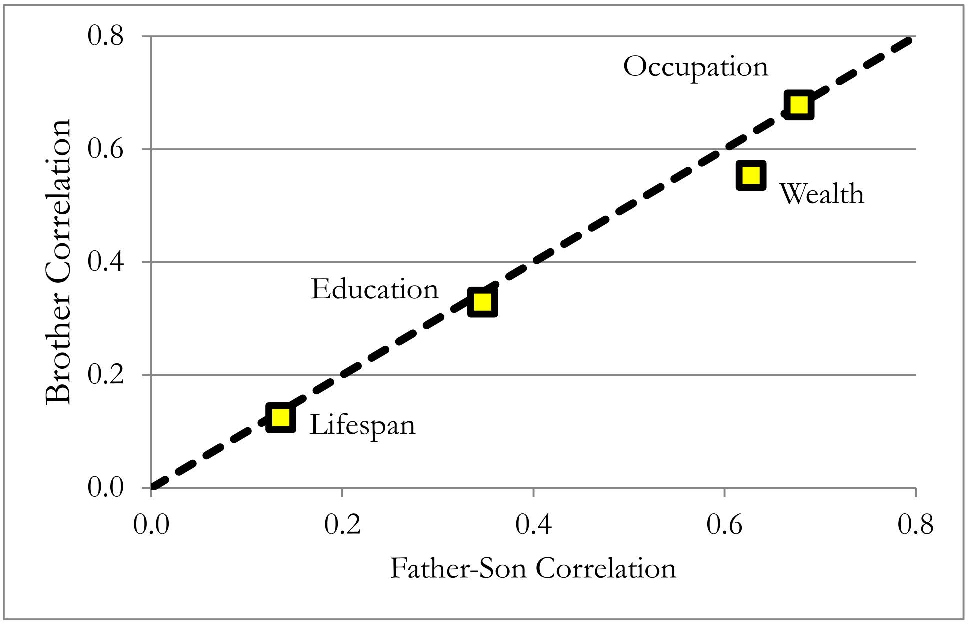


Figure 7: Wealth Correlations

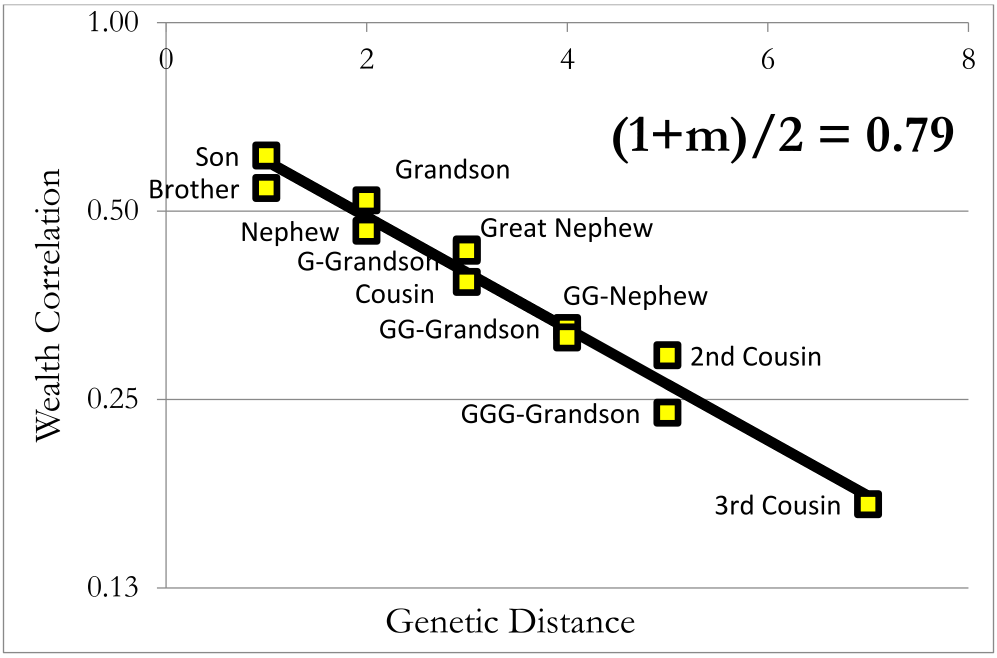


Figure 8: Occupational Status Correlations

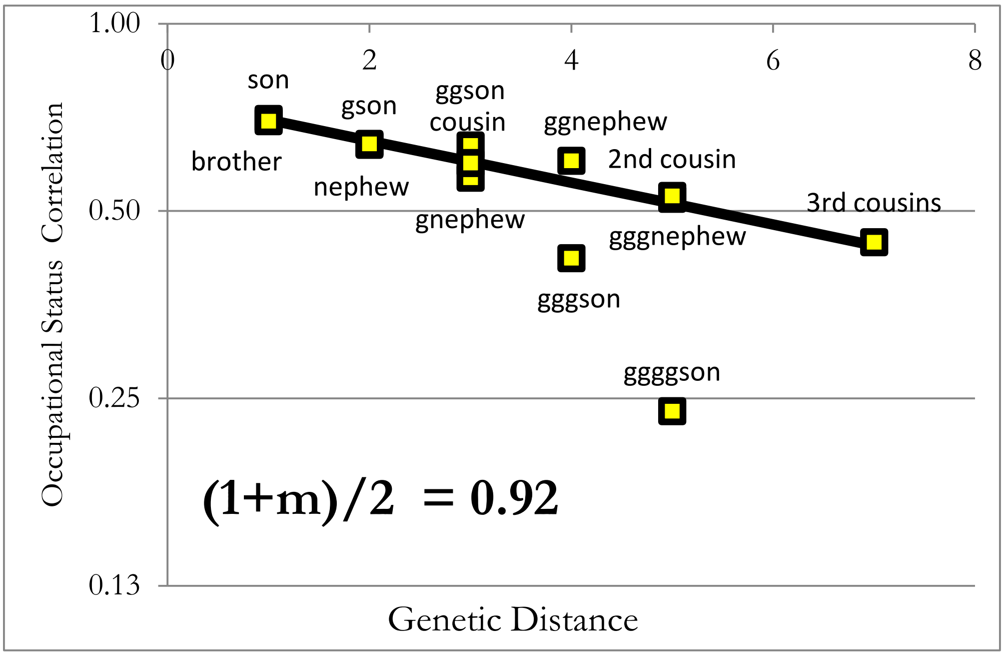


Figure 9: Educational Status Correlations

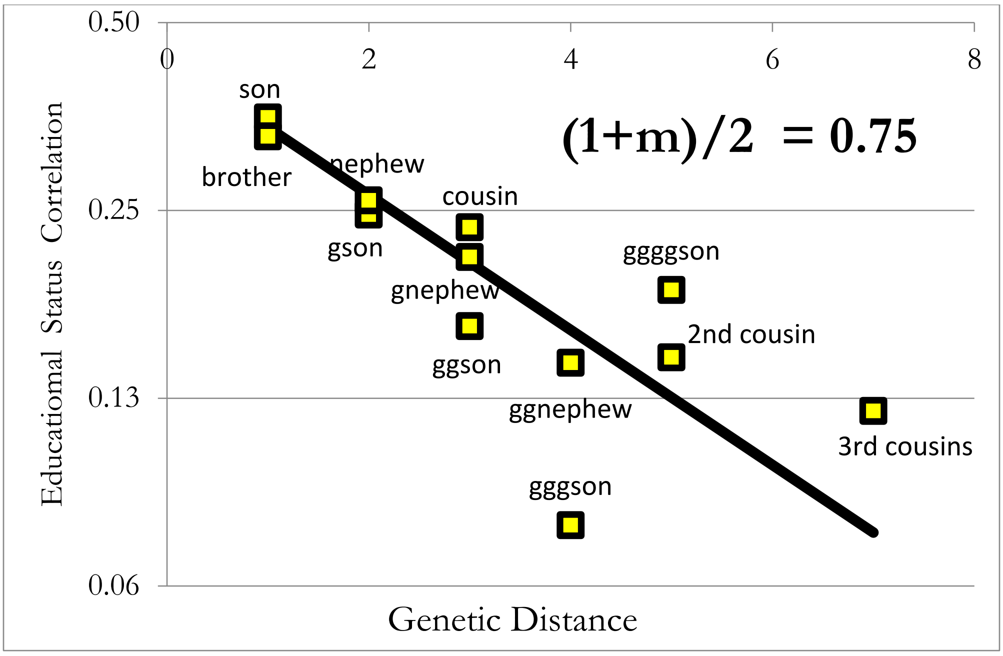


Figure 8 shows the same graph, but now for occupational status. Now the correlations again fall relatively closely to the predicted linear pattern. The R^2 for the fitted line is 0.81. However, the estimated value for *(1+m)/2* = 0.92 (the standard error on this slope estimate is 0.011).[8] So the two estimates of *(1+m)/2* from wealth and occupational status are statistically significantly different. There is thus violation here of the simple additive genetic model posited above, where all aspects of social status stem from the same set of genes.

Figure 9 shows again the pattern of correlations, this time for educational status (measured as an indicator variable for attaining higher educational qualifications. Here the estimated value for *(1+m)/2* = 0.75 (the standard error on this estimate is 0.025). The R^2 for the fitted line, estimated as above by weighted least squares, is 0.90. The slope here is very consistent with that found for wealth, but different from the slope for occupational status.

The pattern of correlations found across relatives in all three cases is consistent with additive genetic inheritance. However, if social status is a unitary inherited trait, then the underlying long-run correlation, *(1+m)/2 ,* should be the same in all three cases. But the data suggests that occupational status was more persistent across the generations than either wealth or educational attainment. Whether this would be consistent with a model where social status has multiple independently inherited dimensions we do not know.

[8] Here we fit the line using weighted least squares, with the weights the standard errors of the estimated correlations.

Instrumental Variable Estimates of Father-Son Correlation

We saw above that an implication of the genetic model of status transmission will be that if we estimate the correlation between father and son on any status measure (phenotype), but instrument for that measure using another aspect of social status, then we should uncover the underlying correlation of genotype across generations $(1+m)/2$, which based on the above should be around 0.8. This is assuming that all measures of status derive from the same underlying status genotype, but with independent random components for each element of the status phenotype. If the random elements in the phenotype are linked – a shock to education creates also a shock to wealth, for example – then the instrument will move us closer to estimating $(1+m)/2$, but the instrumented estimate will still be below the true correlation across generations.

Table 8 shows the OLS and IV estimates of the father-son correlation, where we instrument for the particular outcome with one of four aspects of status: ln wealth at death, occupational status, an indicator for having completed some type of higher education, and an indicator for being at work aged 11-20. In each case the IV coefficient rises significantly relative to the OLS coefficient. The average IV coefficient is 0.77, close to the expected value of around 0.8 based on the pattern of intergenerational correlations. Thus the father-son data is very consistent with the patterns above indicating a strong intergenerational correlation in genotype, but modest hereditability of any particular status phenotype element. This data also implies an underlying genetic correlation in mating of $m = 0.54$.

Table 8: Instrumental Variable estimates of father-son coefficient

Outcome	OLS	IV Lnwealth	IV Occupation	IV Education
Ln Wealth	0.506 (.010)	-	0.608 (.015)	0.608 (.024)
Occupational Status	0.681 (.011)	0.944 (.020)	-	0.803 (.017)
Higher education	0.338 (.014)	0.935 (.046)	0.703 (.026)	-

Note: Standard errors in parentheses. Standard errors clustered by father.

Assortative Mating

The underlying correlations in social status observed in table 6, and figures 4 and 5, to be genetic in origin, require a high degree of assortment in marriage on the genotype itself. But as noted above, the correlation in any particular aspect of the status phenotype between spouses tends to be modest, in the range 0.2-0.5. If the matching is on the basis on some element of the phenotype, such as wealth, then the underlying genotype correlation would be only half as strong, and the observed underlying persistence of status could not stem from genetic transmission.

But if there was marital matching based on underlying genetic characteristics (such as if the parties match on the basis of the average of a host of observed and unobserved status phenotypes that approximates then the status value of the genotype), then instrumenting for one aspect of status in estimating the correlation λ in

$$y_{iM} = \lambda y_{iF} + \varepsilon_i , \qquad (5)$$

will reveal the correlation in underlying genetic characteristics. Our data for spouse status is for marriages 1802-1946, largely in a period before women had good independent measures of status. We have to proxy for the status of women using either that of their brothers, or the status of their father. This will be a noisy measure of the underlying status of women, with an expected correlation of the underlying genetics of father and brother with daughter being $(1+m)/2$. So in this case if we use an IV estimate, the coefficient we would expect to recover would be $m(1+m)/2$.

Table 9 shows the brother-brother in law correlations in wealth, occupation and higher education attainment. These, as expected are less than for the father-son pairing in each case. However, once we instrument in this case the underlying coefficient averages $0.80 = m(1+m)/2$, implying an estimate of m of 0.86. However, there is more error associated with this estimate because of the smaller numbers of son-in-law observed. So it need not be inconsistent with the previous estimate of $m = 0.54$ from the father-son correlations.

These estimates do support, however, the idea that even in the nineteenth century marriage was highly assortative despite the absence for women of educational qualifications, and formal occupations. Their husbands match almost as closely to their brothers as the brothers do to each other.

Table 9: Instrumental Variable estimates of brother-brother in law correlation

Outcome	OLS Brother-Brother in law	IV lwealth	IV Occupation	IV Education
Ln Wealth	0.413 (.021)	-	0.905 (.040)	0.785 (.061)
Occupation Rank	0.627 (.037)	0.927 (.049)	-	0.838 (.055)
Higher education	0.184 (.020)	0.701 (.032)	0.603 (.046)	-

Table 10: Instrumental Variable estimates of father-son in law correlation

Outcome	OLS Father-Son in law	IV lwealth	IV Occupation	IV Education
Ln Wealth	0.459 (.020)	-	0.624 (.030)	0.578 (.060)
Occupation Rank	0.539 (.037)	0.795 (.049)	-	0.734 (.070)
Higher education	0.132 (.025)	0.780 (.086)	0.604 (.074)	-

Another way we can estimate $m(1+m)/2$ is to regress son in law's status on the father's status, again instrumenting with another measure of status. Table 10 shows the results of this procedure. Sons in law are almost as closely correlated with fathers as are sons, but always with a somewhat lower correlation. However, if we instrument the implied correlation rises again to an average this time of 0.69. This implies that m has a value of 0.77. Thus again there is evidence consistent with strong underlying assortment in marriage based on the genotype.

Family Size and Birth Order

Another implication of additive genetic status transmission would be the unimportance of family size and birth order to outcomes. The lineage discussed here also allows us to test for family size and birth order effects, because it can be shown that for marriages initiated 1780-1880 family size is random with respect to social status, and not the product of individual decisions. The strong evidence for this is the absence of any correlation in family size between brothers, and also between fathers and sons, for marriages falling in this period. We will just summarize the effects of the family size tests here, since we have another paper devoted to this substantial topic (Clark and Cummins, 2016). Family size has significant effect of the probability of achieving higher education, on the probability of being at work ages 11-20, on occupational status, or on adult longevity. It also has no effect on child mortality, for the children of the next generation. The one social outcome that family size has clear effects on is wealth at death. Children from larger families have lower wealth at death. However, while an extra child in generation 0 effects the wealth of generation 1, that effect quickly dissipates so that by generation 2, the shock to family size in generation 0 has no significant effect on wealth. Thus in the long run wealth seems to depend more on underlying abilities and attitudes whose inheritance is not affected by shocks to family size. So while in the course of two generations wealth is dependent on non-genetic factors, in the longer run wealth dynamics are possibly still genetic.

Conclusions

It is generally assumed that the elements that define social status – occupational status, educational attainment, wealth, and even health – are transmitted across generations in important ways by the social environment. Above we show that the patterns of correlation of social status attributes in an extended lineage of 62,000 people in England are mainly those that would be predicted by simple additive genetic inheritance of social status in the presence of highly assortative mating around status genetics. Parent-child correlations for a train equal those of siblings, regression to the mean is linear for traits, and the patterns of correlation of relatives of different degrees of genetic affinity is mainly consistent with that predicted by additive genetics. Further family size and birth order have little effect on adult outcomes. The underlying persistence of traits is such that people who have likely never interacted socially, such as second and third cousins, remain surprisingly strongly correlated in terms of occupational status and wealth. The patterns observed imply that marital sorting must be strong in terms of the underlying genetics. But we find evidence, through IV estimates, that there is indeed such sorting, with an underlying correlation in latent social status estimated to be in the range 0.6-0.9.

If this interpretation is correct then aspirations that by appropriate social design, rates of social mobility can be substantially increased will prove futile. We have to be resigned to living in a world where social outcomes are substantially determined at birth.

Bibliography

Clark, Gregory and Neil Cummins. 2016. "The Child Quality-Quantity Tradeoff, England, 1780-1880: A Fundamental Component of the Economic Theory of Growth is Missing." Working Paper, March.

Clark, Gregory and Neil Cummins. 2015. "Intergenerational Wealth Mobility in England, 1858-2012. Surnames and Social Mobility." *Economic Journal, 125*(582): 61-85.

Clark, Gregory and Neil Cummins. 2014. "Surnames and Social Mobility: England, 1170-2012." 2014. *Human Nature, 25*(4), 517-537.

Clark, Gregory, Cummins, Neil et al. 2014 *The Son Also Rises: Surnames and the History of Social Mobility.* Princeton: Princeton University Press.

Falconer, D. S. 1981. *Introduction to Quantitative Genetics*. 2nd ed. London: Longman.

Fisher, R. A. 1918. "The Correlation between Relatives on the Supposition of Mendelian Inheritance." *Transactions of the Royal Society of Edinburgh, 52:* 399-433.

Holt, Sarah B. 1961. "Quantitative Genetics of Fingerprint Patterns." *British Medical Bulletin*, 17: 247-50.

Mascie-Taylor, C. G. Nicholas. 1987. "Assortative mating in a contemporary British population." *Annals of Human Biology*, *14*:1: 59-68. DOI: 10.1080/03014468700008841

Mascie-Taylor, C. G. N. 1989. "Spouse Similarity for IQ and Personality and Convergence." *Behavior Genetics, 19(2)*: 223-7.

McManus, I. C. and C. G. N. Mascie-Taylor. 1984. "Human Assortative Mating for

Height: Non-Linearity and Heteroscedasticity." *Human Biology, 56(4)*: 617-623.

Nagylaki, Thomas. 1978. "The correlation between relatives with assortative mating." *Ann. Hum. Genet., Lond.*, 42: 131.

Plomin, R., Fulker, D.W., Corley, R., and DeFries, J. C. 1997. "Nature, nurture, and cognitive development from 1-16 years: A parent-offspring adoption study." *Psychological Science*, *8*: 442-447.

Sacerdote, Bruce. 2007. "How Large Are the Effects from Changes in Family Environment? A Study of Korean American Adoptees." *Quarterly Journal of Economics, 122*(1): 119-157.

Silventoinen, Karri, Jaakko Kaprio, Eero Lahelma, Richard J. Viken, and Richard J. Rose. 2003. "Assortative Mating by Body Height and BMI: Finnish Twins and Their Spouses." *American Journal of Human Biology 15*:620–627.

Tambs, Kristian, Torbjørn Moum, Lindon J. Eaves, Mike C. Neale, Kristian Midthjell, Per G. Lund-Larsen and Siri Næss. 1992. "Genetic and Environmental Contributions to the Variance of Body Height in a Sample of First and Second Degree Relatives." *American Journal of Physical Anthropology 88: 285-294.*

Watkins MP, Meredith W. 1981. "Spouse similarity in newlyweds with respect to specific cognitive abilities, socioeconomic status, and education." *Behavior Genetics, 11(1)*: 1-21.