

Surnames and Social Mobility

Gregory Clark¹

Neil Cummins²

To what extent do parental characteristics explain child social outcomes? Typically, parent-child correlations in socioeconomic measures are in the range 0.2-0.6. Surname evidence suggests, however, that the intergenerational correlation of overall status is much higher. This paper shows, using educational status in England 1170-2012 as an example, that the true underlying correlation of social status is in the range 0.75-0.85. Social status is more strongly inherited even than height. This correlation is constant over centuries, suggesting an underlying social physics surprisingly immune to government intervention. Social mobility in England in 2012 is little greater than in pre-industrial times. Surname evidence in other countries suggests similarly slow underlying mobility rates.

KEYWORDS: Social Mobility, intergenerational correlation, status inheritance

Since the pioneering work of Francis Galton and Karl Pearson, there has been interest in how strongly children inherit parental characteristics, the “Laws of Inheritance” [1-3]. In this paper we tackle this issue afresh, using status information from surnames to estimate the intergenerational correlation of social status. The data we use is for educational status in England from 1170 to 2012, but similar results can be found for other measures of status and other countries. By social status we mean the overall ranking of families across aspects of status such as education, income, wealth, occupation and health.

Conventional estimates put the correlation between parents and children of the components of status at 0.3-0.5 in England, both in recent generations and in the

¹Department of Economics, University of California, Davis, CA 95616.
gclark@ucdavis.edu.

²Department of Economics, CUNY, Queens College, Queens, NY 11367.
neilcummins@gmail.com

nineteenth century [4-10]. The intergenerational correlations of income and education in England fall at the average of those observed internationally [5,9]. These correlations imply rapid regression to the mean of family socioeconomic characteristics across generations. They also imply that parental characteristics explain only a quarter or less of the variance in child outcomes. These correlations have been assumed to represent overall social mobility rates. If the process of social mobility is Markov, the same across each generation, these intergenerational correlations imply that the expected status of most elite and disadvantaged families will converge within 3-5 generations. Class structure does not persist across generations in modern societies.

Here we estimate from surnames the intergenerational correlation of educational status in England over the course of the years 1230 to 2012, 27 generations of 30 years. Since the medieval period, surnames in England in any generation were mainly derived from inheritance. Thus if family statuses quickly regress to the mean, so should surname statuses. But they do not. Surnames reveal the intergenerational correlation of educational status in England to be in the range 0.73-0.83, even for the most recent generations. Measured in this way educational status is even more strongly inherited than height.³ Initial status differences in surnames can persist for as many as 20-30 generations.

We postulate that the surname correlations are much higher than conventional estimates because families have an underlying social status that is changing slowly. In practice we observe aspects of status such as education, occupation and income. These individual aspects of status are linked to underlying status through random components. A family of high underlying social status can for accidental reasons appear high or low in status in terms of the individual aspects such as education. The surname estimates measure the correlation of underlying social status across generations. Because of the random components, aspects of social status have less intergeneration correlation than underlying social status, and give biased estimates of true rates of underlying social mobility. An implication of this postulate is that social mobility rates measured from surnames will be the same for any aspect of status. We

³ The intergenerational correlation for height is 0.64 [11]. Grönqvist et al. estimate however that in modern Sweden the intergenerational correlation of cognitive ability is as high as 0.77 [12].

show that the intergenerational correlation of wealth for surnames 1830-1966 is indeed 0.78, similar to that for education.

Surname Status

To measure the average social status of surnames we use as an indicator the frequency of surnames among students at Oxford and Cambridge Universities, hereinafter referred to as *Oxbridge*, compared to the frequency of these surnames in the general population. For the average surname this ratio, the relative representation, will be 1. For high status surnames it will be greater than 1, and for low status surnames less than 1. We utilize a database with the surnames of most of those who attended Oxbridge 1170-2012. These were England's only universities until 1832, and thereafter the most elite English universities, enrolling typically only one percent of the eligible population.

We have information on the relative frequency of surnames in the population from 1538-2005 from a variety of sources: censuses, and records of births and marriages. These sources are described in the Supplementary Material.

In England in 1300 surnames varied substantially in average social status. Surnames were first adopted by the upper classes. The Domesday Book of 1086 records surnames for many major landholders, these being mainly the Norman, Breton and Flemish conquerors of England in 1066. These surnames derived mainly from the home estates of these lords in Normandy. They have remained a distinctive class of surnames throughout English history. They include many still well known: *Baskerville, Darcy, Mandeville, Montgomery, Neville, Percy, Punchard, and Talbot*.

Another, later, vintage of high status surnames were those of landholders listed in the *Inquisitions Post Mortem* of 1236-1299. The *Inquisitions* were enquiries into successors of the feudal tenants of the king. Among these property owners were many with relatively rare surnames of more recent English origin, again mainly deriving from the location of their estates: *Berkeley, Pakenham*, etc.

Lastly locative surnames, those which identified a person by their place of origin such as *Atherton, Puttenham, Beveridge*, were typically of higher status in 1300. At the time of creation such locative surnames, such as *Roger de Perton* (later *Roger Perton*),

implied the possessor operated in the larger world outside the rural villages that dominated medieval life. They were thus merchants, traders, attorneys, priests, civil servants, and soldiers. Although such surnames must originally have been a modest share of all surnames, they now constitute at least a quarter of all surnames of English origin. We utilize a sample of these surnames whose endings, such as *..ton*, imply they are locative surnames.

Surname spelling was not standardized in England before the late eighteenth century. The modern *Smith*, for example, evolved from one of four medieval spellings – *Smith*, *Smithe*, *Smyth*, and *Smythe* - only in the seventeenth and eighteenth centuries. But also surnames mutated from their original forms when the earlier meaning was lost. This stems partly from elite surnames moving down the social ladder across generations because of social mobility, to be borne eventually by illiterates ignorant of the surname meaning. The occupational surname, *Arbalistarius*, for example, recorded in the Domesday Book, and derived from the Latin *Arcus* (bow) and *Ballista* (catapult), has no meaning to those without a Classical education. Thus it mutated into the modern forms *Arblaster* and *Alabaster*. So in looking at the frequencies of these medieval surnames across generations, we include spelling variants and derived surnames.

The process of social mobility, however, means medieval high status surnames lost most status information over generations. Long established surnames at high frequency in the population were average in social status by 1800. For later periods we can, however, identify rarer surnames that just by chance had acquired an average high or low status. We thus form, for example, a sample of the rare surnames of the successful by selecting the surnames of those matriculating at Oxbridge 1800-1829, where 40 or fewer people held the surname in the 1881 census. The surnames on this list appear similar in character and perceived status as those not on the list, as table S2 illustrates. Such surnames themselves would not help determining the social position of bearers. Also high status individuals were not selectively adopting these surnames as a more socially fitting appellation.

Estimating Intergeneration Status Correlations

We assume that there is a normal distribution of underlying family competence or social status across families i , of surname group j , in generation t , indexed by x_{ijt} , and that $x_{ijt} \sim N(0, 1)$. We also assume that there is an intergenerational correlation of this status b , such that $x_{ijt+1} = bx_{ijt} + e_{ijt}$. Assume also that there is a measure of educational attainment, y_{ijt} , such that $y_{ijt} = x_{ijt} + u_{ijt}$. Educational status is linked to underlying social status, but with a random component. In this case for individual families the correlation of educational status across generations will be $= \frac{b}{1+\sigma_u^2}$. The greater the variance of the random component linking underlying social competence of families to educational status, the lower will be the correlation of educational status for individual families across generations. But for surname groupings of sufficient size the intergenerational correlation of average educational status \bar{y}_{jt} will be b , the underlying correlation of social status. This is because for such surname groupings the average random component will be close to zero, so that $\bar{y}_{jt} = \bar{x}_{jt} + \bar{u}_{jt} \cong \bar{x}_{jt}$.

We assume that Oxbridge represent the top of the education distribution for England. We further assume that elite surname groups have the same variance of educational status as the population as a whole.⁴ These assumptions imply that the share of these surnames at Oxbridge will decline over time for elite groups in a predictable way, given any value of b . The design here is thus to measure b from the rate of decline of the share of elite surname groups over generations at Oxbridge, as illustrated in figure 1. The key statistic we focus on is the relative representation of any group of surnames among the elite where this is given for surname group z as

$$\text{relative representation of } z = \frac{\text{Share of } z \text{ in elite group}}{\text{Share of } z \text{ in general population}}$$

For a given path across generations of relative representation of a surname at Oxbridge, we fit the initial mean status and b value that minimizes the sum of squared deviations of fitted relative representation from actual, measured in logarithms.

⁴ If this is incorrect it will appear when we try and model the observed relative representation over generations with a single b . We will not be able to find a good fit.

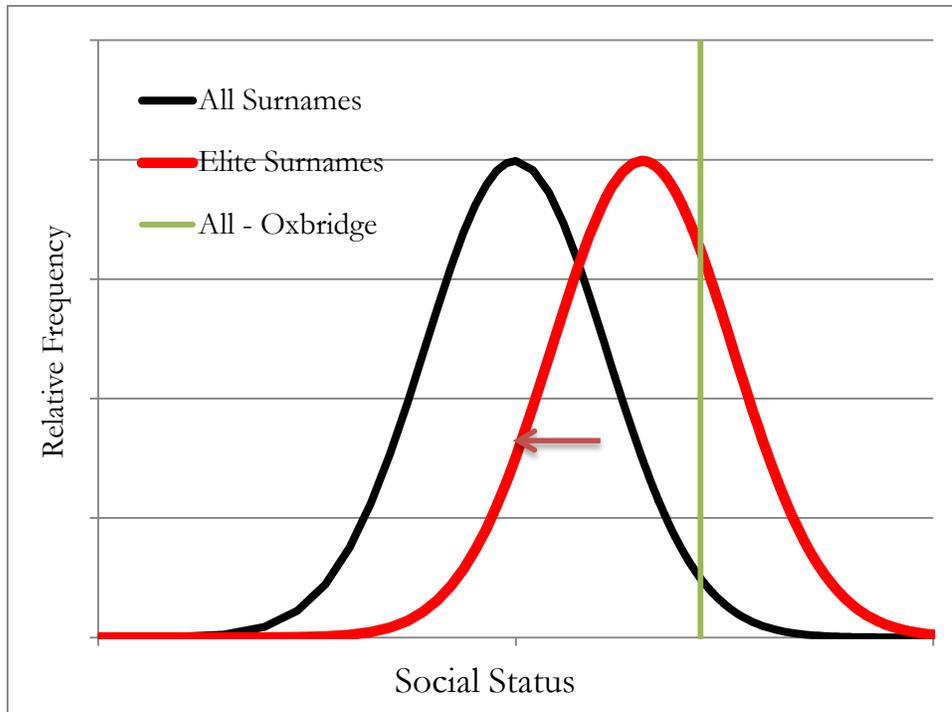


Figure 1. Regression to the Mean of Elite Surnames. The strength of the intergenerational correlation, b , can be measured by the rate of decline of elite surnames to the average of all surnames.

Social Mobility, 1830-2012

We define elite surname groups in 1800-29 by selecting rare surnames found at Oxbridge 1800-29. Taking surnames found 0-40, 41-100, 101-200, 201-300, and 301-500 times in 1881 defines sets of surnames of on average high educational status. The rarer the surname group the higher the average educational status. Figure 2 shows the relative representation of these surnames for thirty year student generations 1830-59, ..., 1980-2009, and 2010-2. We do not use this measure for the generation 1800-29 which is used to group the surnames. The measure in that generation will be upward biased by the random element linking educational status to underlying social status. So for 1800-29 to 1830-59 the intergenerational correlation will be much lower. But in later generations that random component will average 0.

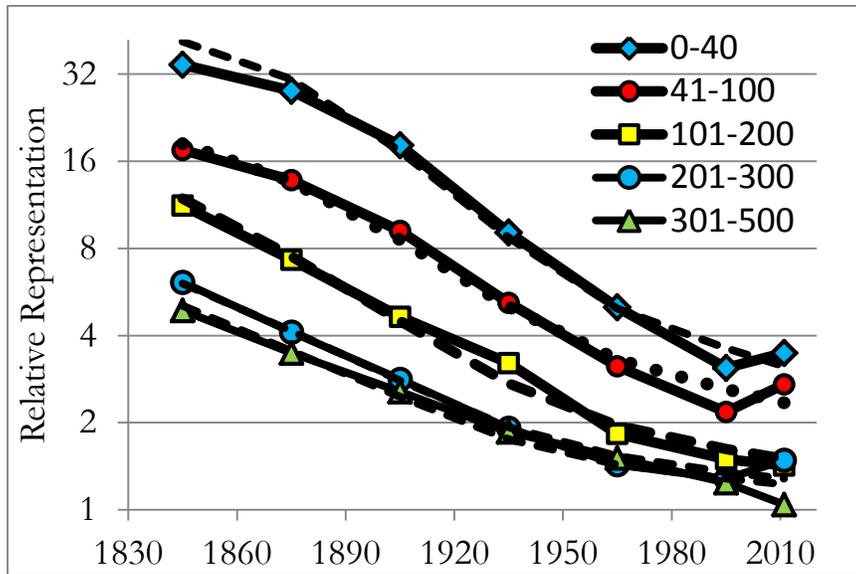


Figure 2: Relative Representation of Rare Surname Groupings, Oxbridge, 1830-2012. Relative representation will equal 1 if surnames have a representation at Oxbridge equal to that expected from the general population. The rare surname sample is taken from Oxbridge lists 1800-29. The different frequency bounds show that rarer names are more elite. Their decline over time is equal to the degree of intergenerational educational mobility.

As expected, the rarer the surname, the higher the implied average status. All surname groups show a steady regression towards a relative representation at Oxbridge of 1. But three things stand out. First the rate of regression to the mean is very slow. As table 1 shows the average estimate for b , following the procedures outlined above, is 0.73. This is much higher than conventional estimates for any type of status persistence. It means that even in 1980-2009, 150 years later, all these surname groups have a statistically significantly higher than average representation among Oxbridge students. Social status persists strongly.

The second striking feature is that the process seems to be Markov. The average status of the next generation depends only on that of the current generation, not on the earlier history.

The third striking feature is that the implied intergenerational correlation of status seems constant 1830-2012. Social mobility does not increase with the emergence after the Industrial Revolution of modern social institutions, such as public education, mass democracy, and redistributive taxation. We see this clearly if we amalgamate the rare elite surnames into one group, surnames held by 0 to 500 people in 1881. This is

shown in figure 3. Relative representation across generations now lies along a smooth curve. One b , 0.73, predicts well the individual observations. The R^2 of the fit is 0.995. There is no increase in social mobility in later generations. Also shown in figure 3 are the 95% confidence intervals for the relative representation stemming from random factors determining whether someone entered Oxbridge. The confidence bounds are narrow, hard to distinguish in the figure, because of the large sizes of the student samples in each generation.

Table 1: b estimates, 1830-2012

Group	Surname Holders 1881	1830-2012 b	Relative population share 2010 versus 1880
High Status			
0-40 1881	12,948	0.77	0.61
41-100 1881	7,838	0.79	0.60
101-200 1881	8,050	0.71	0.76
201-300 1881	11,703	0.69	0.72
301-500 1881	136,925	0.68	0.81
0-500 1881	177,464	0.73	0.78
Low Status			
Rare 2001-5000 1881	501,773	0.64	0.82

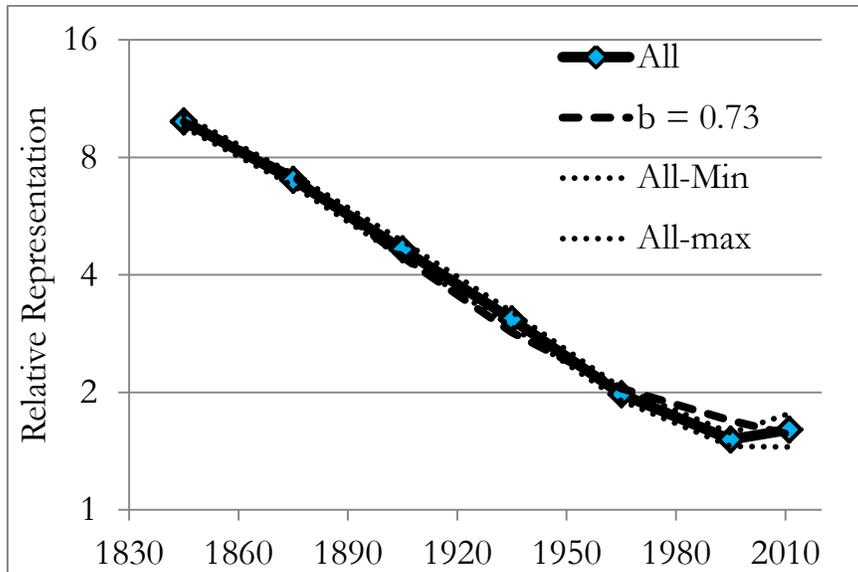


Figure 3. Relative Representation at Oxbridge, All Rare Surnames, 1830-2012. Relative representation will equal 1 if surnames have a representation at Oxbridge equal to that expected from the general population. The rate of decline of all rare surnames, sampled in 1800-29, is constant from 1830 to 2012.

There seems to be a simple law of social mobility, $x_{ijt+1} = bx_{ijt} + e_{ijt}$, that operates largely independently of the social institutions of the society. In England between 1830 and 2012 public provision of education expanded greatly. Publicly provided education was only introduced in 1870, but education to age 10 only became compulsory in 1880. The school leaving age was raised to 11 in 1893, to 14 in 1918, and 15 in 1944.

Local schools, however, played little role in Oxbridge entry in earlier years. Entry to Oxbridge was limited by a number of barriers for lower class students before the 1980s. Oxbridge had its own special entrance exams until 1986. The entry exams for Oxford, for example, until 1940 included a test in Latin. Preparation for these exams was a specialty of a small number of elite secondary schools in England, many of them private fee-paying institutions. In 1900-13 nine schools, including Eton, Harrow and Rugby, supplied 28% of Oxford students [13]. Only in the 1980s did the entry process equalize opportunities to students from all secondary schools.

Another barrier for lower class students was that before 1902 was lack of public support for university education. Oxbridge supplied some financial support, but most

scholarships went to students from the elite schools that prepared them to excel in the scholarship exams. From 1920 to the 1980s, state support for secondary and university education greatly expanded.

We would thus expect more regression to the mean for elite surname frequencies at Oxbridge in the student generations of 1950 and later. There is no evidence of this in figure 3. The earlier surname elite persisted just as tenaciously after 1950 as before.

Above we observe only downwards mobility. Another class of surnames are those which do *not* appear at Oxbridge 1800-29. For a very rare surname, not appearing at the university in this window reveals little about its average educational status. But for more common surname, having not even one holder appear at Oxbridge implies low average educational status.

We thus form a group of surnames held by 2001-5000 people in 1881 which did not appear at Oxbridge 1800-29. In 1830-59 these had a relative representation at Oxbridge only about one third of the average. Even by 2010-2 these names had a relative representation of only 0.94. Figure S3 shows the path to the average of these names. Again there is an implied constant rate of regression to the mean across the generations, though with a somewhat lower estimated b of 0.64.

A fourth feature that emerges in table 3 is that elite surnames have been in relative population decline since 1880. The more elite, the greater the decline. Fertility was lower for upper class families, particularly 1880-1960. Did upper social groups maintain their social position by greater family limitation, and consequent greater child investments, than lower class families? However, the persistence of elite surnames is as strong in the generations 1830-89 when fertility was as high for social elites as for the lower classes [14]. Again changes in the correlation of fertility with social class have no effect on mobility rates.

Social Mobility, 1170-1800

We can estimate surname shares at Oxbridge back to 1170 for the three medieval elite surname groups. To estimate b we need the surname population shares also. We estimate these from marriage records 1538-1800. In pre-industrial England, elite

surnames tended to increase population share over time as a result of the greater fertility of wealthier families [14, 15]. For 1170-1537 we thus project the surname share backwards from that of 1538-1559. We assume the same average percentage change by generation as from 1560-89 to 1650-89. As table S3 shows, the population share of these surnames increased between 1560 and 1680. So we are projecting a smaller share for 1290 than for 1530. That projection may be high or low, creating greater uncertainty about the earlier mobility estimates.

Figure 4 shows the estimated relative representation of a set of locative surnames: those ending in “ton” “ham” “dge” “bury” “land” and derivatives. These at their peak represented 7.1% of all English surnames. These surnames rose in relative representation from 1170 to their peak in 1290-1319, when they were five times as common among Oxbridge attendees than in the general population. That representation declined to the present, and was within 10% of their population share by 1860-89.

Assuming a constant intergenerational status correlation 1290-2012 the best fitting b is 0.83. This is remarkable status persistence by modern standards. Remarkable again is the stability of b across different social eras. It is the same in the Middle Ages, when the universities were dominated by the Catholic Church, as after the English Reformation of 1534-58, when a new more Protestant theology prevailed. There is no sign of enhanced mobility in the Industrial Revolution era of 1760-1860, despite the rise of new industries, and new wealth. For the modern period, mobility may be greater, but these names are so close to average status by 1860 that we cannot measure this.

A more elite set of medieval surnames is identified from a sample of the rarer surnames held by men dying 1236-99, whose estates were subject to an Inquisition Post Mortem (IPM). Though identified purely through their wealth, these surnames peak in their relative representation at Oxbridge at the same time, in the years 1230-59. Then they are 30 times as common at the universities as their population share. Again one b fits the IPM group 1230-2012 reasonably well, as figure 4 shows, though this one is even higher at 0.90. These surnames are still statistically significantly overrepresented at Oxbridge as recently as 1980-2009, 750 years after their peak.

Figure 4 suggests that b for the IPM surnames may be lower 1800-2012. Estimated just for these years it is 0.81. This however, is still higher than the intergenerational correlation estimated for rare surnames at Oxbridge 1830-2012. However, the IPM surnames declined in relative population share less than expected for elite surnames 1880-2012 (S3). Possibly there has been adoption of these surnames by upwardly mobile families because of their elite connotations. Such adoption by entrants to the elite would slow the measured rate of social mobility. This suggests the more status neutral locative surnames likely give better estimates of the true rates of social mobility before 1800.

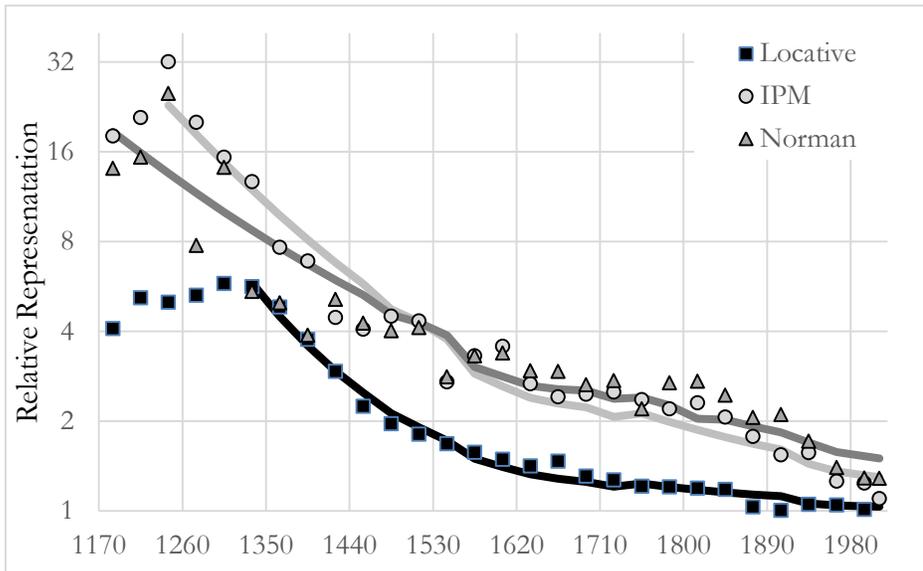


Figure 4. b Estimates, 1170-2012. Elite medieval rare surnames (sourced from the Domesday book (Norman), Inquisition Post Mortem (IPM) and those of Locative origin) rate of decline in representation at Oxbridge. The underlying intergenerational mobility this suggests is consistent with the analysis of the rare names from 1800-29.

The Norman surname sample shows even stronger persistence. These surnames persisted so strongly at Oxbridge, with a b of 0.93, that even in 2010-2 they are statistically significantly overrepresented. Again there is sign of less persistence post

1800, with a b of 0.82. Once more, however, there is an unexpected maintenance of population shares for these surnames 1880-2012 (table S3). Locative surnames' population share declined 20% over this interval, but Norman surnames declined only 6%. Selective adoption of these surnames by entrants to the elite may have maintained the status of the surnames more than the status of the actual descendants of the original bearers. Again the more status neutral locative surnames likely indicate the true rates of persistence in England 1300-1800.

Overall the rate of regression to the mean of these elite surnames suggests that there has been modest improvement in social mobility rates between the medieval era and the modern world, with that change occurring around 1800. But what is remarkable in both periods is the very high implied intergenerational correlation. 0.73 since 1800, 0.83 before 1800.

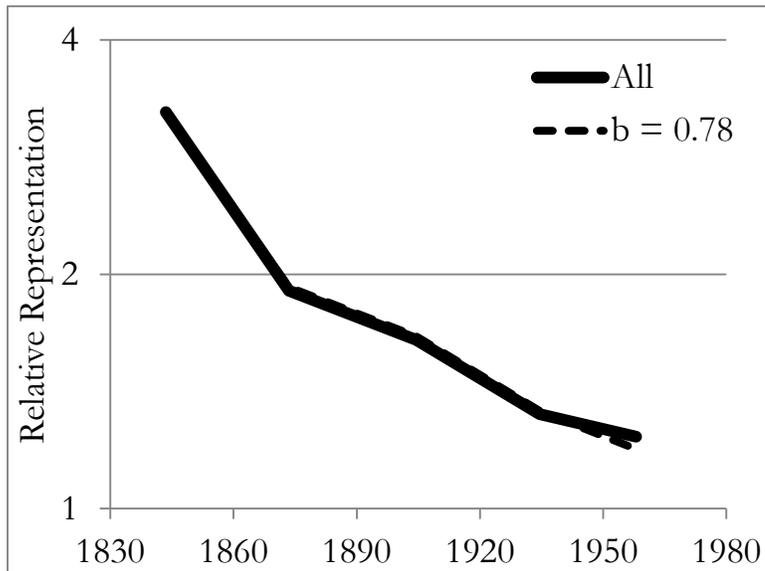
Why are Social Mobility Rates so Low?

We can dismiss a couple of possible reconciliations of the low b from surnames with conventional estimates. One is that the high degree of persistence applies only to the most elite families, with most families display higher rates of educational mobility. Another is that there is a special barrier concerning entry to Oxbridge. There was an Oxbridge “club” that families and their descendants belonged to.

The evidence that there is nothing special about the persistence of high status families, or about Oxbridge as a measure of general status, comes if we look at another more democratic measure of status, the fraction of people whose estates were probated at death. There is a national probate register for England 1858-2012. But only a fraction of the population, those with estates above a minimum value, was legally obliged to be probated. The fraction of all adults probated at death was thus 15% in 1858-89, rising to 47% by 1950-66. When we measure wealth mobility using the fraction of surnames of a given type probated we thus measure mobility across a large share of the wealth distribution. If social mobility rates are higher outside elite families, the b derived from probates will be lower. If entry to Oxbridge is unusually persistent compared to less “clubby” measures of status, such as wealth, again the wealth b will be lower.

Figure 5 graphs the relative representation of rare surnames (500 or fewer in 1881) found at Oxbridge 1800-29, in the national probate records 1858-1966. Also indicated is the relative representation of these surnames among the earlier Canterbury Prerogative Court probates 1830-1858. Under the earlier ecclesiastical probate system the Canterbury court represented the richest probates, with about 4% of all adult males probated here. People dying 1830-1858 would include many from the generation attending Oxbridge 1800-29, since life expectancy at 25 in England was then 30 years. Figure 5 also shows the best fitting b for these five generations. That b is 0.78, and once again shows remarkable stability across these generations. In a related paper using similar methods and the Canterbury Prerogative Court probates 1710-1858 we show that the implied b for wealth mobility in Industrial Revolution England is 0.77-0.82, little if at all higher than for the modern era [16].

Figure 5: Mobility Measured by Relative Probate Frequencies, Oxbridge Elite 1800-29



This wealth b of 0.78 shows that the remarkable status persistence found using Oxbridge attendance as the status measure is found just as strongly with a more general and democratic measure of status such as asset ownership. There is no special

persistence at Oxbridge, or in education, or only in the upper reaches of status. The high and stable wealth b also shows again the remarkably irrelevance of institutions to social mobility. Over these generations there were substantial increases in the rate of taxation of wealth and income, especially after 1910. Yet this did nothing to increase rates of wealth mobility [17].

The similar magnitude of the estimated b for educational status and wealth is consistent with the hypothesis above that there is a deeper latent social status of families that correlates much more highly across generations than any individual status component. This implies also that if we find surname groupings with high status on any aspect of social status at one time, they will be equivalently high status on any other measure of social status. What is being measured in this way is generalized social mobility.

The relative constancy of the intergenerational correlation of underlying social status across very different social environments in England from 1800 to 2012 suggests that it stems from the nature of inheritance of characteristics within families. Strong forces of familial culture, social connections, and genetics must connect the generations. There really are quasi-physical “Laws of Inheritance.” This interpretation is reinforced by the finding of Clark in work with other co-authors that all societies observed – including the USA, Sweden, India, China and Japan - have similar low rates of social mobility when surnames are used to identify elites and underclasses, despite an even wider range of social institutions [18].

References

1. Galton F (1869) *Hereditary Genius: An Enquiry into its Laws and Consequences*. London: Macmillan.
2. Galton F (1889) *Natural Inheritance*. London: Macmillan.
3. Pearson K, Lee A (1903) On the Laws of Inheritance in Man: I. Inheritance of Physical Characters. *Biometrika*, 2: 357-462.
4. Atkinson A, Maynard A, Trinder C (1983) *Parents and Children: Incomes in Two Generations*. London: Heinemann.
5. Corak M (2012) Inequality from Generation to Generation: The United States in Comparison, in Robert Rycroft (editor), *The Economics of Inequality, Poverty, and Discrimination in the 21st Century*, ABC-CLIO.
6. Dearden L, Machin S, Reed, H (1997) Intergenerational mobility in Britain. *Economic Journal*, 107: 47-66.
7. Ermisch J, Francesconi M, Siedler, T (2006) Intergenerational mobility and marital sorting, *Economic Journal* 116: 659-679.
8. Harbury C and Hitchens D (1979) *Inheritance and Wealth Inequality in Britain*. London: Allen and Unwin.
9. Hertz T, Jayasundera T, Piraino P, Selcuk S, Smith N, Verashchagina A (2007) The inheritance of educational inequality: International comparisons and fifty-year trends. *The B.E. Journal of Economic Analysis & Policy* 7, Article 10.
10. Long J (2013) The Surprising Social Mobility of Victorian Britain. *European Review of Economic History*, 17(1): 1-23.
11. Silventoinen K, et al. (2003) Heritability of adult body height: A comparative study of twin cohorts in eight countries, *Twin Research*, 6: 399-408.
12. Grönqvist E, Öckert B, Vlachos J (2011) The intergenerational transmission of cognitive and non-cognitive abilities. IFN Working Paper No. 884. Available at <http://dx.doi.org/10.2139/ssrn.2050393>.
13. Greenstein D (1994) The junior members, 1900-1990: a profile. In Brian Harrison (ed.), *The History of the University of Oxford, Volume VIII*. Oxford: Clarendon Press.
14. Clark G, Cummins N (2012b) Malthus to Modernity: England's First Demographic Transition, 1760-1800. Working Paper, University of California, Davis.
15. Clark G, Hamilton, G (2006) Survival of the Richest. The Malthusian Mechanism in Pre-Industrial England. *Journal of Economic History*, 66(3) (September): 707-36.

16. Clark G, Cummins N (2013) Inequality and social mobility in the Industrial Revolution Era. Forthcoming in Roderick Floud, Jane Humphries, and Paul Johnson (eds.), *The Cambridge Economic History of Modern Britain*. Cambridge: Cambridge University Press.
17. Clark G, Cummins N (2012a) Intergenerational Mobility in England, 1858-2012. Wealth, Surnames, and Social Mobility. *Under review*.
18. Clark G, Cummins N, Diaz Vidal D, Hao Y, Ishii T, Landes Z, Marcin D, Mo Jung K, Marek A, Williams K (2014) *The Son Also Rises: 1,000 Years of Social Mobility*. Book Manuscript in preparation for Princeton University Press.

Supplementary Information

The Oxbridge Surnames Database

The sources for this database are listed in the references [1-16].

For the years before 1500 the database includes the names of faculty as well as students. Also for Oxford 2010-2, the structure of the e-mail directory makes it impossible to exclude some faculty names. The incompleteness and informality of records at Oxford and Cambridge in earlier years, and the imperfect sources in later years such as exam results lists, means that the database is necessarily always just a sample of those attending the universities.

Table S1 shows the total stock of people identified as attending Oxbridge in each generation, assumed to be 30 years. In earlier years this is just a sample of those attending the universities. From 1530 to 1892 this is a nearly complete list of all matriculating students. 1892-2009 the data is once more just a sample of all attendees. The third column shows the estimated total numbers of students in each generation. For 1170-1469 the share attending Oxbridge is assumed to be 0.8% of each male cohort. This is similar to the shares observed for 1470-1499, and is 4-5 times the observed shares pre 1440. But the source limitations in these years mean that only a fraction of attendees were observed.⁵ The fourth column gives the population of those surviving to age 16 in each generation from which the student population was drawn from. Before 1870 this population is assumed to be males only. Thereafter an increasing number of females attended the university, until it is assumed that by 1990 the all males and females aged 16 are potential Oxbridge attendees.

⁵ Ashton estimates that students recorded for Oxford 1170-1500 were only 20-25% of actual numbers [17].

Table S1: Surnames at Oxbridge

Generation	Oxbridge Students observed	Estimated Total Oxbridge Students	Assumed Domestic Share	Population students drawn from	Oxbridge cohort share (%)
1170-99	107	-	1.00	-	0.80
1200-29	260	7,510	1.00	853,400	0.80
1230-59	386	8,742	1.00	993,407	0.80
1260-89	787	9,514	1.00	1,081,095	0.80
1290-1319	1,317	11,934	1.00	1,356,162	0.80
1320-49	2,284	12,590	1.00	1,430,674	0.80
1350-79	1,746	9,991	1.00	1,135,318	0.80
1380-1409	3,332	7,241	1.00	822,842	0.80
1410-39	2,115	6,333	1.00	719,703	0.80
1440-69	5,454	5,744	1.00	652,724	0.80
1470-99	6,146	6,146	1.00	628,280	0.89
1500-29	5,684	5,684	1.00	654,964	0.79
1530-59	6,477	6,477	1.00	789,152	0.71
1560-89	19,349	19,349	1.00	849,960	2.01
1590-1619	22,327	22,327	1.00	1,009,277	2.06
1620-49	24,232	24,232	1.00	1,273,656	1.85
1650-79	23,908	23,908	1.00	1,462,187	1.75
1680-1709	17,042	17,042	1.00	1,479,698	1.13
1710-39	16,021	16,021	1.00	1,492,885	1.00
1740-69	10,519	10,519	1.00	1,583,707	0.61
1770-99	11,994	11,994	0.99	1,793,974	0.55
1800-29	18,649	18,649	0.99	2,246,609	0.64
1830-59	24,415	24,415	0.99	3,245,746	0.62
1860-89	38,678	38,678	0.96	7,085,936	0.53
1890-1919	30,962	47,526	0.93	9,265,992	0.48
1920-49	67,927	92,854	0.88	11,589,095	0.70
1950-79	156,645	192,254	0.86	14,209,853	1.16
1980-2009	221,196	314,956	0.76	18,838,670	1.27
2010-12	41,489	41,489	0.62	6,526,919	1.19

In later generations increasing numbers of Oxbridge students have been drawn from outside England and Wales. For 1980-2012 the Oxford University Gazette summarizes the fraction of students drawn from outside England and Wales [18]. Cambridge has similar statistics for 2000-10 [19].

Thus in 2012 only 62.3% of Oxford students were domiciled in England and Wales. In 2010 the equivalent numbers for Cambridge are 61.9%. However, many students from outside England and Wales were drawn from populations that contained substantial numbers of immigrants from England and Wales: Scotland, Northern and Southern Ireland, the USA, Canada, Australia, New Zealand, South Africa. These students constituted 14.4% of the Oxford student population in 2012. The equivalent numbers for Cambridge in 2010 are 10.5%.

We thus took the “English” surname share at Oxbridge as 62% in 2010-2, and 76% in 1980-2009. We project these foreign surname shares backwards by measuring the share of typically German, Swedish, Dutch, Spanish, Italian, Chinese and Indian surnames at Oxbridge 1800-1979.

The final column of table S1 shows the implied share of the eligible population attending Oxbridge. From 1470 to 2012 this has varied. At its peak in 1560-89 it was 2.2%, at its minimum in 1890-1919 it was 0.5%.

A generation is taken to be 30 years. Some studies have assumed a generation as short as 20 years for pre-industrial society. But in England from 1538 onwards the average women gave birth to her first child at age 25 or later, and the average man at 27 or later, so that the average interval for a generation would be around 30 years. If the generation length is actually shorter than this then true social mobility rates will be slower.

Surname Elites

Surnames were written with many spellings before the nineteenth century. Figure S1 shows this for the surname *Smith*. Thus for all the earlier surname samples we take all possible spelling variants of the surname. The English also had the practice from the nineteenth century onwards of creating new surname by compounding surnames. Thus we get Cave-Brown-Cave, Fox-Strangways and so on. We include for the selected surnames also any surnames derived from these by compounding.

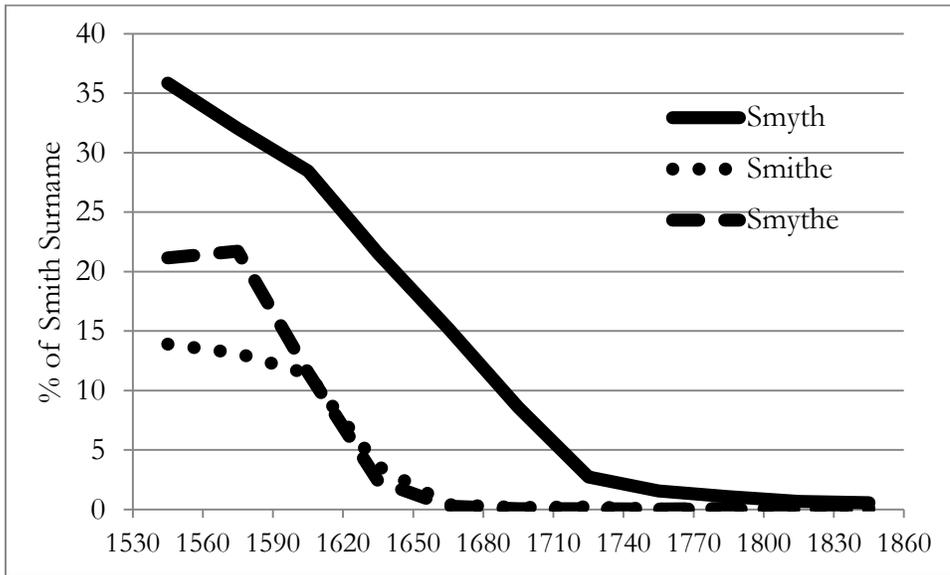
Normans

“Norman” surnames were identified as a sample of the surnames of landlords in the Domesday book identified by Keats-Rohan as deriving from place names in Normandy, Brittany or Flanders [20]. All possible derivations from these original surnames were included.

Medieval Wealthy

The IPM surnames are a sample of rarer surnames that appeared with high frequency in the *Inquisitions Post Mortem* 1236-1299 [21,22]. Rarer in this case meant surnames held by less than 10,000 people in 1881.

Figure S1: “Smith” Variants among Marriages, 1538-1859



Locative Surnames

Location surnames were identified as all those ending in *..ton, ..don, ..dge, ..ham, ..land, bury*, and variants such as *..tone, ..tonn, ..tonne, ..tun*. In this case hyphenated surnames containing one of these surnames as a component were included only if the location surname was the last component.

Rare Surnames, 1800-29

These samples were surnames that appeared at Oxbridge 1800-29 which were rare in the 1881 census. For the list of surnames occurring 0-40 times in the 1881 census Table S2 shows 24 randomly chosen surnames from the beginning of this list of surnames occurring at Oxbridge 1800-29, compared to 24 randomly chosen surnames from the beginning of the surnames of frequency 1-40 in 1881 not on this list.

Table S2: Rare Oxbridge versus non-Oxbridge Surnames, 1800-29

Oxbridge		Non-Oxbridge	
Agassiz	Brickdale	Agnerv	Bodgett
Anquetil	Brooshooft	Allbert	Boolman
Atthill	Bunduck	Arfman	Bradsey
Baitson	Buttanshaw	Bainchley	Breckill
Barnardiston	Cantis	Bante	Callaly
Bazalgette	Casamajor	Barthorn	Capildi
Belfour	Chabot	Bavey	Carville
Beridge	Charretie	Bedborne	Cavet
Bleek	Cheslyn	Bemond	Chanterfield
Boinville	Clarina	Berrton	Chesslow
Boscawen	Coham	Bideford	Chubham
Bramston	Conyngham	Bisace	Clemishaw

Table S3: Population Share by Surname Type

Population Share	Locative (%)	IPM (%)	Norman (%)
1290-1319	(4.59)	(0.203)	(0.176)
1530-59	5.72	-	-
1560-89	5.89	0.372	0.329
1680-1709	6.37	0.482	0.432
1770-99	6.64	-	0.453
1881	7.04	0.535	0.508
2002	5.67	0.482	0.475

Notes: (..) indicates projected population share based on the rate of growth of the share 1560-1680.

Candidate surnames on these lists that showed an unusual increase in frequency between 1881 and 2002, and where the surname was of foreign origin, including in this case Scottish and Irish surnames, were excluded. The aim was to have a set of surnames where most of the holders in England and Wales in 2012 descended from the holders of 1800-29.

Population Shares

In the period 1830-2012 population shares of surnames groups for the rare surnames of 1800-29 were estimated for 4 benchmark periods, 1837-57, 1877-97, 1965-85, and 1985-95. The 1837-57 and 1877-97 benchmarks were estimated from the national register of marriages for these years, since child mortality was still significant in these years and differed by social class. The 1965-85 and 1985-95 benchmarks came from the birth register. The population share for 1830-59 for Oxbridge was taken as the 1837-57 benchmark, and that 1860-1919 from the 1887-1897 benchmark. The population share 1980-2009 came from the 1965-85 benchmark, and for 2010-2 from the 1985-95 benchmark. Population shares 1920-1979 were linearly interpolated from the shares 1877-97 and 1965-85.

For the earlier surname elites population shares 1560-89, 1680-1719 and 1770-99 were estimated from parish marriage records as recorded in the International Genealogical Index [23]. For 1881 the share was estimated from the census [24]. For 2002 the share was derived from the Office of National Statistics database of surname frequencies in England and Wales [25]. Population shares were linearly interpolated between these dates. Table S3 shows the resulting implied shares for the medieval surname elites.

Estimating b for Education

Table S4 details how b was estimated for the rare surnames appearing 500 times or less in 1881 that were enrolled in Oxbridge 1800-29. The share of the surnames at Oxbridge was calculated from the assumed share of the students at Oxbridge in each generation from England, as in table S1, but with an allowance for some share of foreign students coming from countries such as New Zealand where many

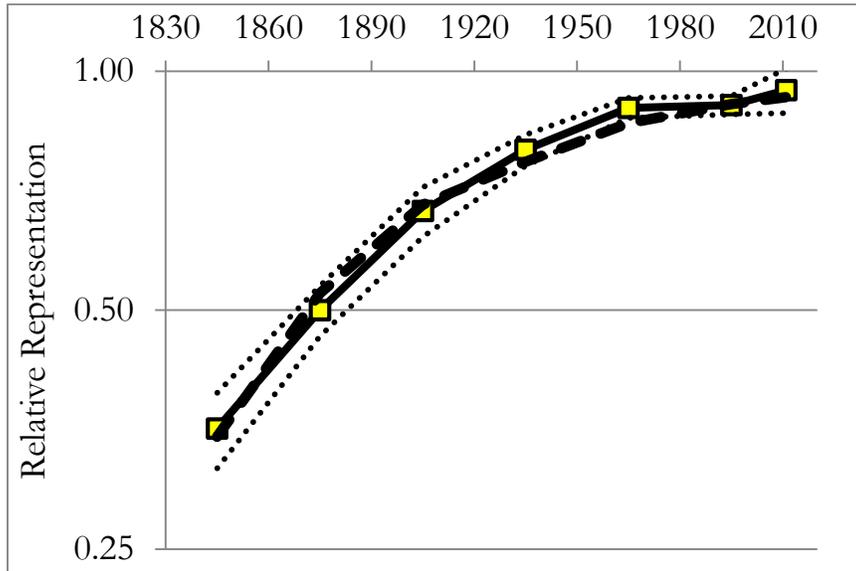
Table S4: Calculating b for the 0-500 Rare Surnames

Period	Share Oxbridge (English Surnames)	Share Population	Relative Representation	Oxbridge Elite (%)	Implied Mean Status	Implied b
1830-59	11.86	1.18	10.04	0.62	1.05	-
1860-89	8.18	1.15	7.11	0.53	0.76	0.72
1890-1919	5.23	1.11	4.72	0.48	0.58	0.76
1920-49	3.24	1.06	3.06	0.70	0.43	0.75
1950-79	1.96	1.01	1.94	1.16	0.26	0.60
1980-2009	1.38	0.86	1.60	1.27	0.19	0.72
2010-2	1.42	0.86	1.65	1.19	0.20	1.12

surnames are of English origin. From the ratio of their share of Oxbridge graduates to their share of the population we get their relative representation in the Oxbridge elite.

We also know what share of each eligible cohort attends Oxbridge, which is assumed to be the top of the educational distribution. Given the relative representation, and the size of the Oxbridge elite, we calculate where the implied mean of the educational status of this group lies relative to the population, in standard deviation units. This is shown in the sixth column on table S3. From this we can calculate a period by period implied b value, as is shown in the last column. Here the average b is 0.78. But this weights equally the observations in the early and later generations. Since the implied group mean of educational status is close to the social average, the estimates in later generations have less precision. So we fit the average implied b by minimizing the sum of squared deviations of the actual log relative representation from the fitted log relative representation, assuming one b throughout, which gives $b = 0.73$.

Figure S2: Regression to the Mean of Low Status Surnames, 1830-2012



The path of relative representation for the surnames of frequency 2001-5000 not found at Oxbridge 1800-29 is displayed in figure S2. Here the estimated b is lower at 0.64, but again fits well for the entire period.

Probate Rates

Probate frequencies for rare surnames 1858-1966 were found from the Calendar of the Principle Probate Registry, as recorded on Ancestry.com. Probate frequencies for the years 1830-1857 were obtained from the Indexes of Wills and Administrative Grants of the Prerogative Court of the Archbishop of Canterbury [26]. The share of deaths in each generation from the rare surname group was taken to be the same as the shares of the population reported in table S4.

References

1. Brasenose College (1909) *Brasenose College Register, 1509-1909*. Oxford, Basil Blackwell.
2. Cambridge University (1954) *Annual Register of the University of Cambridge, 1954-5*. Cambridge: Cambridge University Press.
3. Cambridge University (1976) *The Cambridge University List of Members, 1976*. Cambridge: Cambridge University Press.
4. Cambridge University (1998) *The Cambridge University List of Members, 1998*. Cambridge: Cambridge University Press.
5. Cambridge University (1999-2010) *Cambridge University Reporter*. Cambridge: Cambridge University Press.
6. Elliott, Ivo (ed.) (1934) *Balliol College Register, 2nd edition, 1833-1933*. Oxford: John Johnson.
7. Emden, Alfred B. (1957-9) *A Biographical Register of the University of Oxford to AD 1500* (3 vols.) (Oxford: Clarendon Press.
8. Emden, Alfred B (1963) *A Biographical Register of the University of Cambridge to 1500*. Cambridge: Cambridge University Press.
9. Emden, Alfred B (1974) *A Biographical Register of the University of Oxford AD 1501 to 1540*. Oxford: Clarendon Press.
10. Foster, Joseph (1887) *Alumni Oxonienses: the Members of the University of Oxford 1715-1886: their parentage, birthplace and year of birth, with a record of their degrees: being the Matriculation Register of the University*. 4 Vols. Oxford: Parker and Company.
11. Foster, Joseph (1891) *Alumni Oxonienses: the Members of the University of Oxford 1500-1714: their parentage, birthplace and year of birth, with a record of their degrees: being the Matriculation Register of the University*. 2 Vols. Oxford: Parker and Company.
12. Foster, Joseph (1893) *Oxford Men and Their Colleges, 1880-1892, 2 Volumes*. Oxford: Parker and Co.
13. Venn, John and Venn John. A (1922-7) *Alumni Cantabrigienses, a biographical list of all known students, graduates and holders of office at the University of Cambridge, from the earliest times to 1751*, 4 vols. Cambridge: Cambridge University Press.
14. Venn, John and Venn John. A (1940-54) *Alumni Cantabrigienses, a biographical list of all known students, graduates and holders of office at the University of Cambridge, 1752-1900*, 6 vols. Cambridge: Cambridge University Press.
15. Oxford University (1924, 1972, 1981, 1996, 2000, 2004-8, 2010) *The Oxford University Calendar*. Oxford: Clarendon Press.
16. E-mail Directories (2010-12):

- a. Oxford: http://www.ox.ac.uk/applications/contact_search/
- b. Cambridge: <http://jackdaw.cam.ac.uk/mailsearch/>
- c. Women at Cambridge, 1860-1900. <http://venn.lib.cam.ac.uk/acad/search.html>
17. Ashton, T. S. (1977) "Oxford's Medieval Alumni." *Past & Present*, 74: 3-40.
18. Oxford University Gazette (1980-2012) <http://www.admin.ox.ac.uk/ac-div/statistics/student/>,
<http://www.ox.ac.uk/gazette/statisticalinformation/studentnumberssupplements/>
19. Cambridge University Reporter, special issue, 'Student Numbers': 1999-2000 to 2009-10 <http://www.admin.cam.ac.uk/offices/planning/sso/reporter/index.html>
20. Keats-Rohan, K. S. B. (1999) *Domesday People: A Prosopography of Persons Occurring in English Documents 1066-1166*. Woodbridge, Suffolk: The Boydell Press.
21. Public Record Office (1904) *Calendar of Inquisitions Post Mortem and other Analogous Documents preserved in the Public Record Office, Vol. 1 Henry III*. London: Public Record Office.
22. Public Record Office (1906) *Calendar of Inquisitions Post Mortem and other Analogous Documents preserved in the Public Record Office, Vol. 2 Edward I*. London: Public Record Office.
23. England Marriages, 1538–1973. Index. FamilySearch. <https://familysearch.org> : accessed 2013
24. Schurer, Kevin and Woollard, Matthew, 1881 Census for England and Wales, the Channel Islands and the Isle of Man (Enhanced Version) [computer file]. Genealogical Society of Utah, Federation of Family History Societies, [original data producer(s)]. Colchester, Essex: UK Data Archive [distributor], 2000. SN: 4177, <http://dx.doi.org/10.5255/UKDA-SN-4177-1>
25. Office of National Statistics database of surname frequencies in England and Wales, as listed at <http://www.taliesin-arlein.net/names/search.php>.
26. Indexes of Wills and Administrative Grants of the Prerogative Court of the Archbishop of Canterbury, from the Public Record Office (series PROB 12).