Social Mobility Rates in the USA, 1920-2010: A Surname Analysis

Gregory Clark*, Daniel Marcin**, Firas Abu-Sneneh*, Wilfred M. Chow*, Kuk Mo Jung*, Ariel M. Marek*, Kevin M. Williams*

*University of California, Davis, **University of Michigan, Ann Arbor

How rapid is social mobility in the USA? Using the AMA *Directory of Physicians in the United States*, which lists doctors by date of medical school graduation, and State Bar Association attorney listings, we measure the rate of social mobility in the USA 1920-2012. We use surnames as markers of social groups under- and over- represented in these two elites. The high status groups are descendants of Ashkenazi Jews, descendants of the 1923-4 wealthy, and descendants of pre-1850 Ivy League graduates. The underclass groups are descendants of Native Americans, Blacks, and the inhabitants of New France. Estimated social mobility rates are much lower than conventionally estimated for all groups. Advantages persist for 5 generations or more. However, these underlying mobility rates are no lower than equivalent estimates for Sweden. Nor is there any sign that mobility rates are declining.

Introduction

Conventional measures suggest that social mobility is substantial in the US, even though the US is now regarded as having lower social mobility rates than in much of Europe. A convenient summary measure of social mobility rates is the b in the expression

$$y_{t+1} = by_t + e_t$$

where y_t is a measure of socio-economic status such as income, wealth, or education in generation t.[1] For income or education, b in the US would be estimated at 0.4-0.5.[2] Figure 1, for example, shows the b for years of education across a variety of

[1] The y here is normalized in each generation to a mean of 0, and to the same variance.
[2] See, for example, Black and Devereux, 2010, Corak, 2006 and Corak, 2012.

countries in recent years as estimated by Hertz et al. (2011), versus the Gini coefficient for income inequality. The USA shows up as having a b for education, 0.46, that is relatively high for high income economies, but in the middle for countries as a whole.

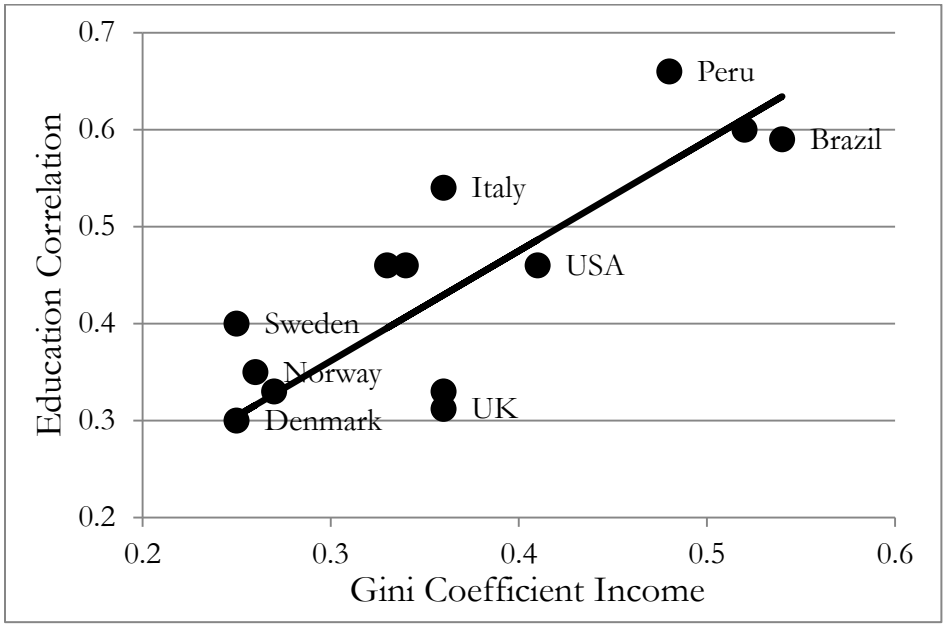
This paper measures b in the US through tracking the social status over generations of elite and underclass groups, using surnames as the markers of these groups. We are able to identify three elite groups in this way – the descendants of Ashkenazi Jews, the descendants of the rich of 1923-4 with rare surnames, and the descendants of Ivy League graduates with rare surnames, 1650-1850. We can also identify three underclass groups – Native Americans, Blacks, and, surprisingly, the US descendants of the French settlers of New France of 1604-1759.

We examine mobility rates across three generations, 1920-1949, 1950-1979, and 1980-2012 using two sources. First is the AMA *Directory of Physicians in the United States*, which lists close to a million doctors licensed to practice in the USA (including Osteopaths who are members of the AMA). As a guard against fraudulent impersonation of retired or deceased physicians, the directory includes many physicians completing medical school in the 1930s and 1940s who are now inactive or dead. The directory thus displays the surname composition of the US medical profession from the 1930s to the 2000s.

The second source is state lists of licensed attorneys, as well as their year of licensure. Unfortunately, unlike the case with doctors, there is no national licensing of attorneys, so this information is contained in 50 state web sites. Using a selection of 25 of these, however, it is possible to get similar measures of social mobility rates for attorneys, who are an elite but less exclusive population than doctors.

The bs estimated from the frequency of surname types among doctors and attorneys over time are systematically much higher than those generated by conventional methods, in the range 0.65-0.85. We show that this implies that generalized and long-run social mobility is much lower than conventionally estimated in the US. However, similar studies for Sweden and England suggest that these US mobility rates, while low, are not any lower than in other societies (Clark, 2012, Clark and Cummins, 2012). Nor is there any sign that these mobility rates are declining.

Figure 1: Intergenerational Education Correlation and Inequality



Sources: Hertz et al., 2011, table 2. Gini for Income, World Bank.

Elite and Underclass Surnames

To measure social mobility using surnames we need estimates of the stock of surnames in the US by cohort. Our basic source on this is a file produced by the US Census Bureau giving the frequency of surnames of a population count of 100 or above in the 2000 Census, recording as well the fraction of holders of each surname declaring themselves as white, black, Asian/Pacific Islander, Native American, and Hispanic. To infer the frequency of surnames with less than 100 bearers in 2000 we use the Social Security Death Index (SSDI), which lists those dying in the US in recent years by name and year of birth, and so gives some idea of the distribution of surnames. To estimate surname frequencies by birth cohorts we can either assume surnames were a stable share of the population, or we can correct earlier surname frequencies by using the SSDI, or using more general information on the frequency of ethnic groups over time from the US censuses. The SSDI correction, however, is biased by the differential death rates of social groups at each age.

In the analysis below we use the following surname groups:

Ashkenazi Jews - This group we identify from the surnames Cohen, Katz, Levin, Levine, Levinsky, Levinson, Levinstein, and Levinthal. These surnames are found heavily in New York City, the area of greatest Jewish population share. However, in the 2000 census 3.6% of the people bearing these surnames declared themselves Black (5.5% for Cohen). Given the different percentage Black across these surnames, this most likely stems not from intermarriage, but from Black Americans independently adopting these surnames because of their Biblical resonance. As can be seen in table 1, these names appear among doctors at a rate 18.7 per 1000 bearers of the surname in 2000, the highest frequency of any surname group.

1923-4 Rich: These surnames were chosen from those appearing in the *New York Times* lists of income tax payers in 1923 and 1924. Congress passed a provision for public inspection of income tax returns in 1924. Before its effective repeal in 1926, major newspapers across the country ran thousands of names and tax payments for the tax years 1923-1924. The Times alone reported the tax payments of more than 34,000 people over these two years. The sample was formed of rarer surnames where there was at least one taxpayer reported per 10 births recorded prior to 1900 in the SSDI. The modal surname was held by less than 100 people in 2000. The frequency of the rarer surnames in 2000 was estimated from SSDI births 1900-2012. The ten most common of the names in 2000 were Vanderbilt, 1,717, Roosevelt, 961, Winthrop, 727, Colgate, 616, Guggenheim, 512, Sonn, 480, Bloomingdale, 467, Plaut, 455, Kempner, 436, and Pruyn, 421. This group of surnames registers among doctors at a rate of 8.7 per thousand, well above the population average of 2.85.

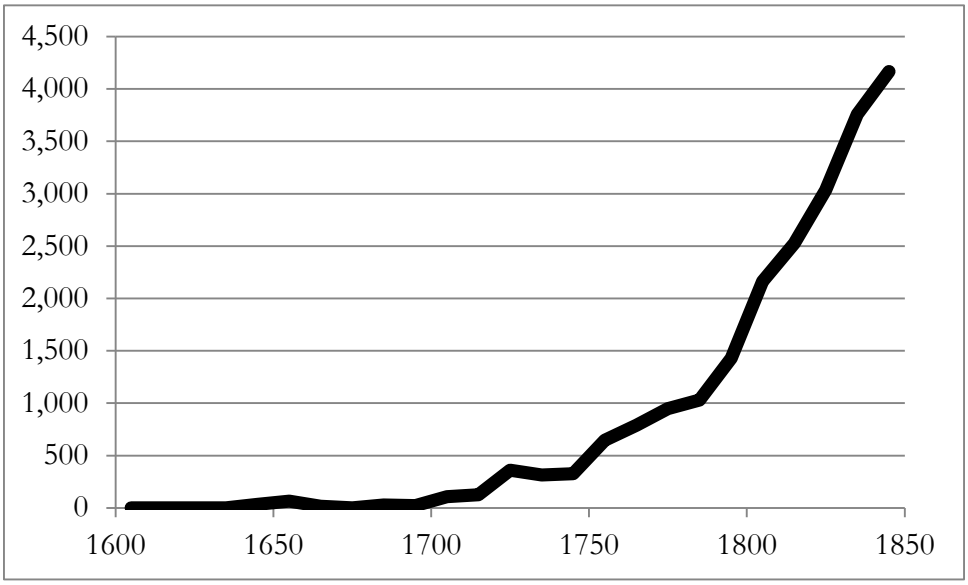
"Ivy League" Graduates 1650-1850: We assembled a dataset of those graduating Brown, Columbia, Dartmouth, Harvard, Rutgers, University of Pennsylvania, William and Mary, and Yale 1850 and earlier. The bulk of these graduates, 71%, are from 1800-49. Figure 2 shows the decadal time path of these observations. We then identify again a set of relatively rare surnames within this set, with the selection criterion being that the surname was held by less than 300 people in the 2000 census, and less than 200 in the 1850 census, per surname holder in the Ivy League sample.[3] This produced 1,085 surnames, with an average estimated frequency in 2000 of only 83. Of these only a very few have any resonance – for example, Rutgers and Rensselaer – the rest being obscure surnames of largely English, Dutch, German and

[3] We also required that for surnames with 100 or more holders in the 2000 census, the name be at least 80% white, and less than 10% black (since blacks have a much lower representation among doctors).

Table 1: The Surname Groups

Group	Examples	Number 2000	Doctors	Rate (per 1000)
Ashkenazi Jewish	Katz, Cohen, Levin	178,442	2,338	18.7
1920s Rich	Vanderbilt, Guggenheim	111,790	972	8.7
"Ivy League"	Rutgers, Rennselaer	79,133	416	5.3
Scandinavian	Olson, Sundberg	935,918	3,395	3.6
French Canadian	Hebert, Gagnon, Cote	667,471	1,257	1.9
Black	Washington, Smalls	475,574	482	1.0
Native American	Begay, Yazzie	77,603	11	0.14

Figure 2: Ivy League Graduates by Decade



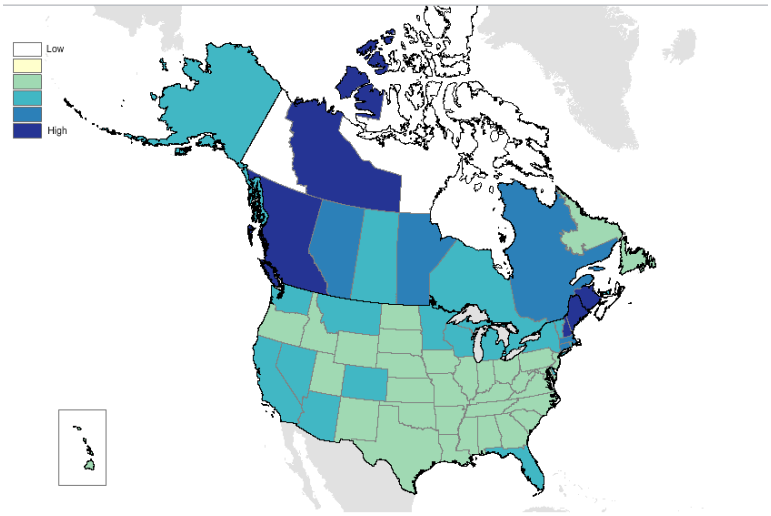
Irish origins. These surnames still have a frequency among doctors, however, of 5.1 per thousand, about 60% greater than average.

Scandinavian: These are surnames mainly of Swedish origin, from the migrations from Scandinavia to the upper Midwest in the years before 1914. It includes patronyms such as Olson which show up with a lower rate per 1000 than names based on geographic features such as Lundquist.

New France: These are surnames derived mainly from the descendants the colonies of New France. They arrived in the US through the takeover of parts of Acadia by the English, the expulsion of Acadians to Louisiana, and the movement of French Canadians to New England in the years 1865-1920 to take employment in New England factories. The surnames chosen were those more common in Canada than in France, and where at least 90% of holders in 2000 declared themselves as white, and less than 5% declared themselves Black. Examples include *Gagnon*, whose distribution across Canada and the USA is shown in Figure 3. *Gagnon* is reported at its highest rate in the population in New Brunswick, Canada (part of the old French colony of Acadia), where it constitutes 0.2% of surnames. But its next most frequent locations are New Hampshire and Maine in the US, where it is again 0.2% of surnames. While its frequency in Canada overall is 633 per million, in France it is a rare surname of frequency 15 per million.

Black - This group is identified as surnames where 87% of more of the holders identified as Black in the 2000 census, and where the surname had English or German origin (to exclude surnames belonging to more recent immigrant groups such as Haitians, or Africans). About a third of this group comes from one name, Washington, presumably adopted on a large scale by emancipated slaves lacking surnames after the Civil War. Many of the other surnames in this group are classical English in sound, and presumably were obtained in the slave era from master's whose one families died out. On average 91.1% of people with these surnames declared themselves as black in the 2000 Census, and 4.2% declared themselves white (with the majority of the rest mixed race). These surnames appear among doctors at one third their surname frequency in 2000.

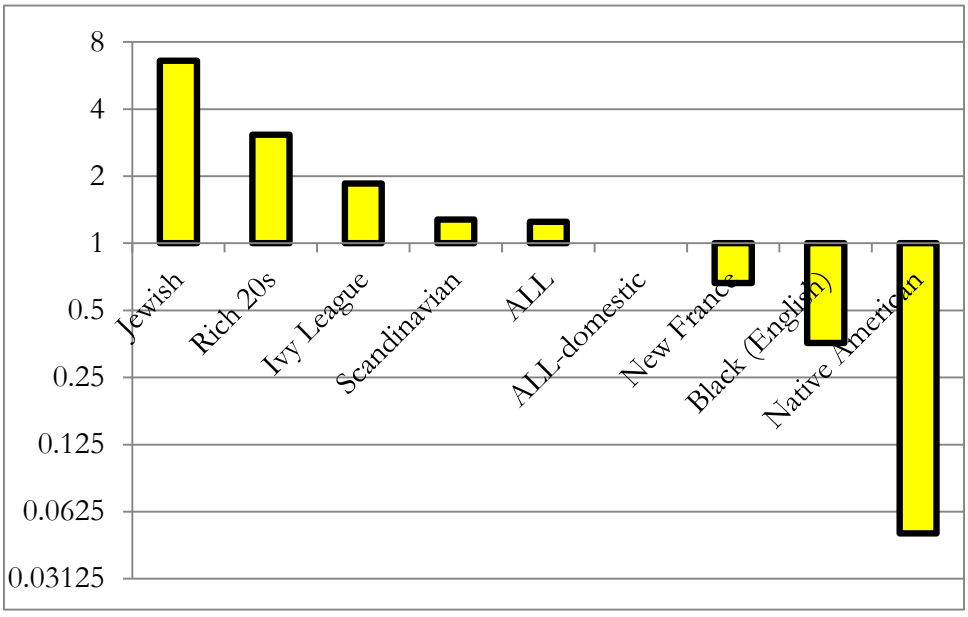
Figure 3: Map of the Distribution in North America of *Gagnon*.



Notes: Deeper shading indicates higher frequency of the surname.

Source: World Names Profiler.

Figure 4: Relative Representation of Surname Types Among US Doctors, 2010 (Log Scale)



Native American – These are surnames where 90% of more of the holders in 2000 identified themselves as Native American. Two names, Begay(e) and Yazzie account for about 40% of this population. But many of the lower frequency Native American surnames are quite distinctive: Manygoats, 650, Roanhorse, 459, Goldtooth, 384, Fasthorse, 367, Yellowman, 280, Twobulls, 249, Bitsilly, 222, Smallcanyon, 219. These surnames are concentrated in the southwest of the USA.

As table 1 shows the rate of occurrence of doctors across these surname groups varies widely. Below we measure social mobility rates using the Relative Representation Rates of surnames: their share among doctors relative to their share in the general population. There are 3.59 doctors per 1,000 in the 2010 AMA directory relative to the stock of surnames that can be identified in the 2000 US census.[4] However, roughly 20 percent of the US stock of doctors for many years has been foreign born, with large contingents of doctors from India, the Philippines, and Pakistan in particular.[5] We thus measure relative representation using as a benchmark the domestic population occurrence rate among doctors of 2.85 doctors per 1,000 of the census surname stock of 2000. On this measure the Jewish, 1920s Rich and Ivy League surnames are overrepresented, and for names of French Canadian origin, for Black and for Native American surnames they are underrepresented. Figure 3 shows the relative representation of each surname type, which is the share of the surname amongst doctors divided by the share of the name in the population (times 0.8). For the domestic population as a whole this ratio is 1. For the Jewish surname group it is 6.6, compared to 0.05 for the Native American surname group.

[4] The census records produced only 269,768,216 surnames since some forms had unreadable or incorrect entries.
[5] AMA, 2006.

Measuring Social Mobility Rates

The measure we have of status at any time for various surname groups is their share in the elite represented by doctors compared to their population share. We thus define the *relative representation* of each surname or surname type, z, in an elite group as

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

With social mobility any surname which in an initial period has a relative representation differing from 1 should tend towards 1, and the rate at which it tends to 1 is determined by the rate of social mobility.

The measure we derive of social mobility is the b in the equation

$$y_{t+1} = by_t + e_t$$

where y is some measure of socio-economic status such as income, wealth, or education.

To extract implied bs from information on the distribution of surnames among elites we proceed as follows. Assume that social status, y, follows a normal distribution, with mean 0 and variance σ^2. Suppose that a surname, z, has a relative representation greater than 1 among elite groups. The situation looks as in figure 4, which shows the general probability distribution function for status (assumed normally distributed) as well as the pdf for the elite group.

The overrepresentation or underrepresentation of a surname in this elite could be produced by a range of values for the mean status, $\bar{y}_{z0}$, and the variance of status, σ^2_{z0}, for this surname. But for any assumption about ($\bar{y}_{z0}$, σ^2_{z0}) there will be an implied path of relative representation of the surname over generations for each possible *b*. This is because

$$\bar{y}_{zt} = \bar{y}_{z0} b^t$$

$$\sigma^2_{zt} = b^{2t}\sigma^2_{z0} + (1 - b^{2t})\sigma^2$$

Figure 4: Initial Position of an Elite

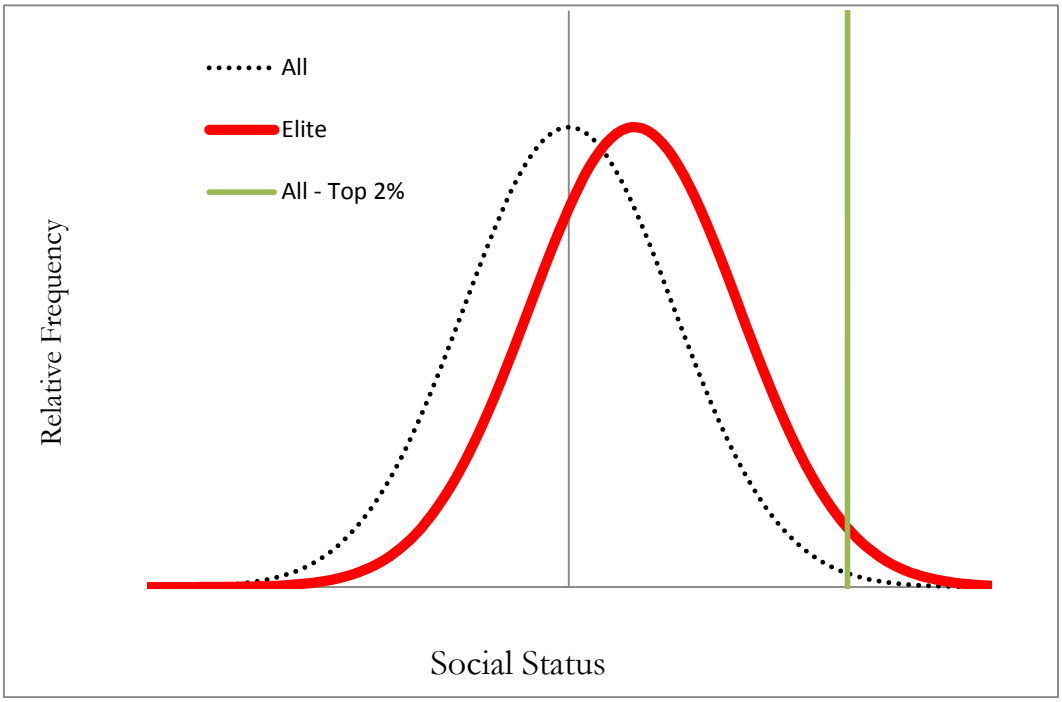
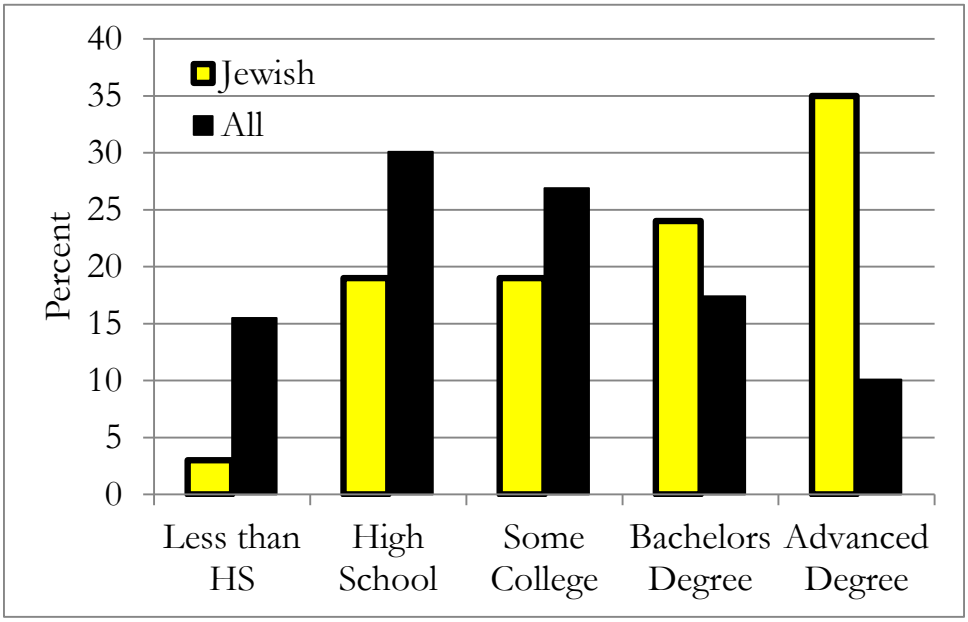


Figure 5: Educational Attainment, Jewish versus US, 2007



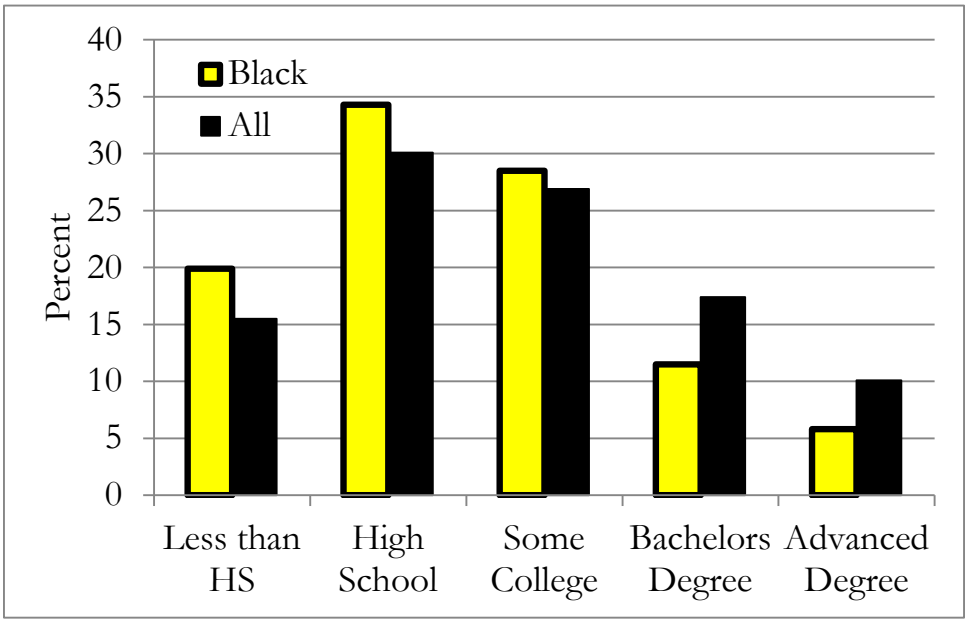
Notes: Jewish results from those identifying as Jewish on a survey on religious affiliation of those 18 and older in 2007. The educational attainment for the general population is for those 25 and above.
Source: Crissey, 2009, table 1, Pew Forum, 2008, 56.

With each generation, depending on *b,* the mean status of the elite or underclass surname will regress towards the population mean, and its variance increase to the population variance (assuming that $\sigma_{z0}^2 < \sigma^2$). Its relative representation in the elite will decline or increase in a particular pattern.

Even though we cannot initially fix $\bar{y}_{z0}$ and σ_{z0}^2 for any surname just by observing its overrepresentation among an elite in the first period, below we assume that the variance equals the population variance. This is because for groups that we can observe, such as the Jewish and Black populations in the US, their distributions of education attainment look as dispersed as the population as a whole, just shifted to the right or left as in figure 4. Thus figure 5 shows educational attainment of the Jewish population (aged 18 and above) compared to the US population in 2007, and figure 6 the educational attainment of the Black population. Even though the Jewish population constitutes an elite, there is still substantial variance in attainment. And even though the Black population falls below average, again there is substantial variance.

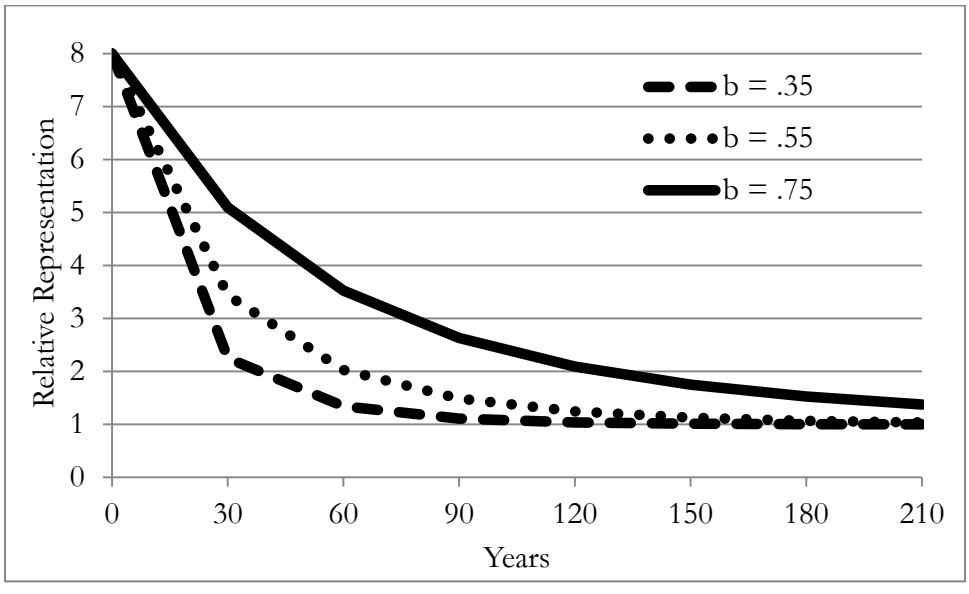
Figure 7 shows what we would expect the relative representation of a surname, which had a relative representation of 8 times its share in the population in the first year, to have in each subsequent 30 year interval with different assumptions about b. If b = .35, the kind of b we expect from standard studies of social mobility, then within two generations surnames heavily overrepresented among the elite should have close to a proportional representation.

Figure 6: Educational Attainment, Black versus US, 2007



Source: Crissey, 2009, table 1.

Figure 7: Relative Representation by Generation with Different bs



Relative Representation by Generation

To measure relative representation by generation we need to estimate for each surname of its representation in an elite by cohorts. We obtain this from the AMA Directory, because it records for each doctor listed the date of completion of medical school. In the directory there are doctors listed as completing medical school from the 1920s on, even though most of these older entrants will now be retired or even dead. Thus we are able to form three generations of doctors from the AMA data: those completing medical school 1920-49, 1950-79 and 1980-2009. Table 2 shows the basic data on surname frequencies by decade 1920-2009, by surname type. For each surname type we can measure their share of all doctors graduating in the same period and recorded in the directory (counting 0.8 of graduations as being from domestic doctors).[6]

However, to estimate relative representation we also need to estimate the share of each surname type in the population aged 25 at the time graduation of each cohort of doctors. We do this in three ways. The first is to just apply the 2000 surname frequencies to all earlier cohorts. The 2000 frequency of all surnames with more than 100 holders is reported by the census in its surnames master file. For rarer surnames we estimate their frequency from the number of people recorded as born with this surname 1900-2012 in the Social Security Death Index. The ratio of such recorded births to the stock of the surname in 2000 is estimated from the ratio of names for the same surname group held by 100-200 people in 2000. The third column of table 3 shows the relative representation of six of the seven surname types by generation on this basis. We cannot do this exercise for Native American surnames because there are too few doctors to get any reliable results.

The demography of these surname groups is not the same, however. In particular the white non-Hispanic population has grown more slowly in the US than the population as a whole. The Ashkenazi Jewish population is also believed to have had low fertility rates in the late twentieth century. Using the census we can estimate cohort sizes for the white non-Hispanic and the Black populations by decade 1900-1980, and use these to derived corrected population shares for each group in each generation. The fourth column shows the relative representation on this measure.

---

[6] The overall frequency of doctors by year was estimated by taking a sample of the directory constituting the first name on each column of each page.

Table 2: Surnames by Type and Decade among Doctors

Decade	All	Jewish	20s Rich	Ivy League	Olson/ Olsen	New France	Black
1920-9	0	1	0	0	0	1	0
1930-9	3,434	36	4	2	4	4	1
1940-9	25,184	100	58	21	10	32	6
1950-9	59,024	353	97	44	46	88	10
1960-9	106,816	522	137	41	83	118	22
1970-9	162,477	729	205	80	138	181	74
1980-9	220,643	836	241	101	155	315	111
1990-9	208,552	511	166	69	131	323	143
2000-9	170,132	355	147	58	120	195	115

Notes: Total for all surnames by decade based on a 1.4% sample of surnames.

The correction applied to all groups other than the Black population is for the white non-Hispanic population.[7] Figure 8 shows these census corrected measures of relative representation across the three generations for each group. As can be seen all 5 groups regress towards the mean, but maintain their relative position over the three generations. The ordinary Scandinavian surname Olson, as a reference, maintains a relative representation close to 1 across all three generations on this measure. The maintenance of a relative representation above 1 for the Jewish surnames, the rare 1920s rich surnames, and the pre-1850 "Ivy League" surnames, implies that there must be slow rates of long run social mobility. For example, those completing medical school in 1980-2009 would be typically the great-grandchildren of the rich taxpayers of 1923-4, 3 generations later. Similarly those completing medical school 1980-2009 are at least 4-5 generation descendants of those graduating college pre 1850. At the b of 0.3-0.5 thought to apply to educational and occupational mobility they should show little trace of any advantage of their earlier forbears.

[7] For the years before 1980 whites were not distinguished as Hispanic and Non-Hispanic.

Table 3: Relative Representation by Generation, Doctors

Group	Period	Relative Representation Pop. Share 2000	Relative Representation Pop. Share adjusted from censuses	Relative Representation Pop. Share from SSDI
Jewish	1920-49	13.64	12.31	6.73
	1950-79	9.68	8.71	8.69
	1980-2009	5.24	5.09	7.52
1920s Rich	1920-49	5.86	5.25	3.82
	1950-79	4.72	4.25	4.29
	1980-2009	3.10	3.04	3.20
Ivy League	1920-49	3.31	2.97	-
	1950-79	2.12	1.91	-
	1980-2009	1.43	1.40	-
Olson	1920-49	1.18	1.06	0.82
	1950-79	1.21	1.09	1.11
	1980-2009	1.02	1.01	1.14
French Canadian	1920-49	0.62	0.55	0.57
	1950-79	0.62	0.56	0.51
	1980-2009	0.69	0.68	0.63
Black	1920-49	0.19	0.22	-
	1950-79	0.20	0.22	-
	1980-2009	0.44	0.44	-

Figure 8: Relative Representation by Surname Type by Generation

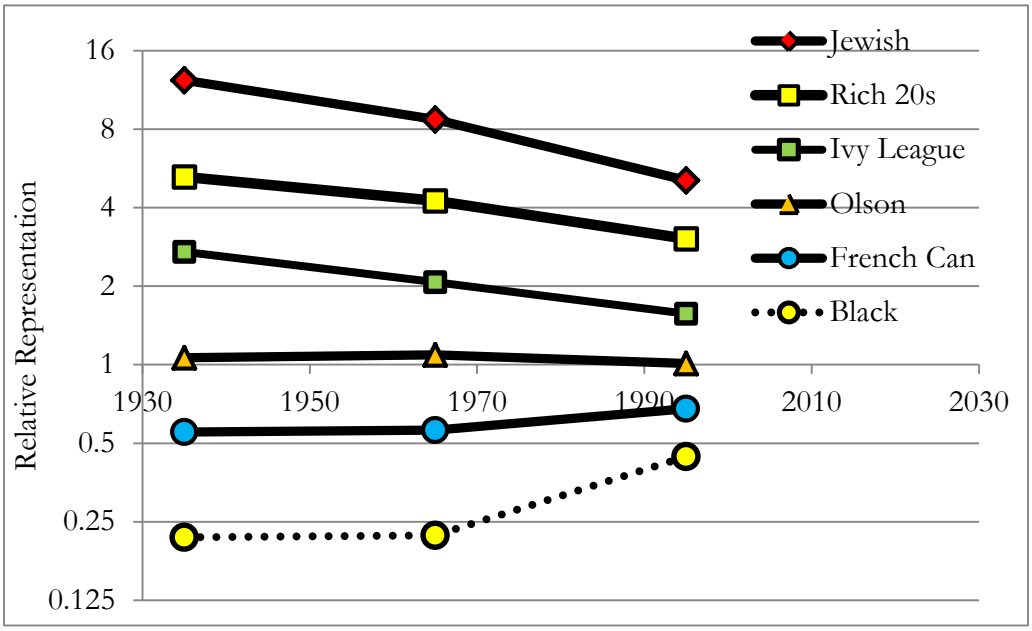


Table 4: Implied bs, 1950-79 to 1980-2009

Surname Group	Pop. Share adjusted from censuses Foreign Doctors 20% in 1950-79	Pop. Share from censuses Foreign Doctors 15% in 1950-79	Pop. Share from SSDI Foreign Doctors 20% in 1950-79
Jewish	0.72	0.74	0.92
1920s Rich	0.75	0.79	0.78
Ivy League	0.61	0.67	-
French Canadian	0.68	0.62	0.70
Black	0.56	0.54	-
Average	0.66	0.67	0.80

The relative representations of Table 3 for the Jewish and Black surnames, assuming both these groups show a normal distribution of social status, imply that the Jewish mean is 0.62 standard deviations above the mean, and the Black mean 0.27 standard deviations below the mean. This is shown in figure 9.

We can also get measures of the numbers of births per year for the particular surnames used here from the Social Security Death Index, and estimate relative cohort sizes. This produces the relative representation by generation shown in the last column. However, because elite groups have lower mortality rates, the SSDI tends to overestimate the population share of these groups in earlier years, and underestimate it for recent years. For a birth in 1900, for example, to get recorded in the SSDI the person would have to die 1962 or later. And for a birth in 1980 to get recorded the person would have to have died at age 32 or younger. But the SSDI provides a bound on the possible relative representation of these surname groups by generation. Thus while the census correction underestimates the declining population share of the Jewish surname group, the SSDI estimates overstate this decline. The true movement of the relative representation will lie between these bounds.

Table 4 reports the b implied by the measures of relative representation in table 2 for the movement between the generation of 1950-79 and 1980-2009, under three different assumptions. The first is that the share of these surnames in the population is measured over time as in the census for non-hispanic whites and blacks, and that in both generations foreign doctors were 20% of those completing medical school in these periods. The second assumes that foreign doctors were only 15% of those completing medical school in 1950-79, compared to 20% in 1980-2009. The third assumes that the foreign doctor share was constant, but that the movement of the surname shares in the population was represented by births recorded in the SSDI.

The persistence displayed by all these groups is high relative to conventional measures of social mobility. The average for the three elite groups in the middle column, for example, is 0.73.

Figure 9: Implied Status Distributions, Jewish and Black Names, 1980-2009

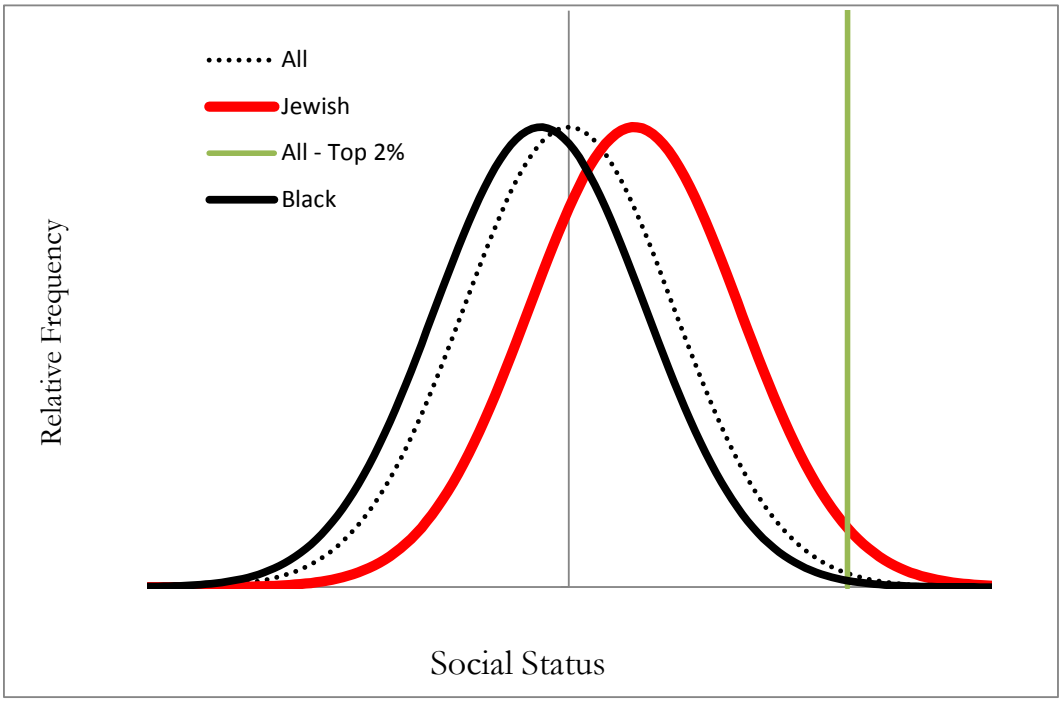
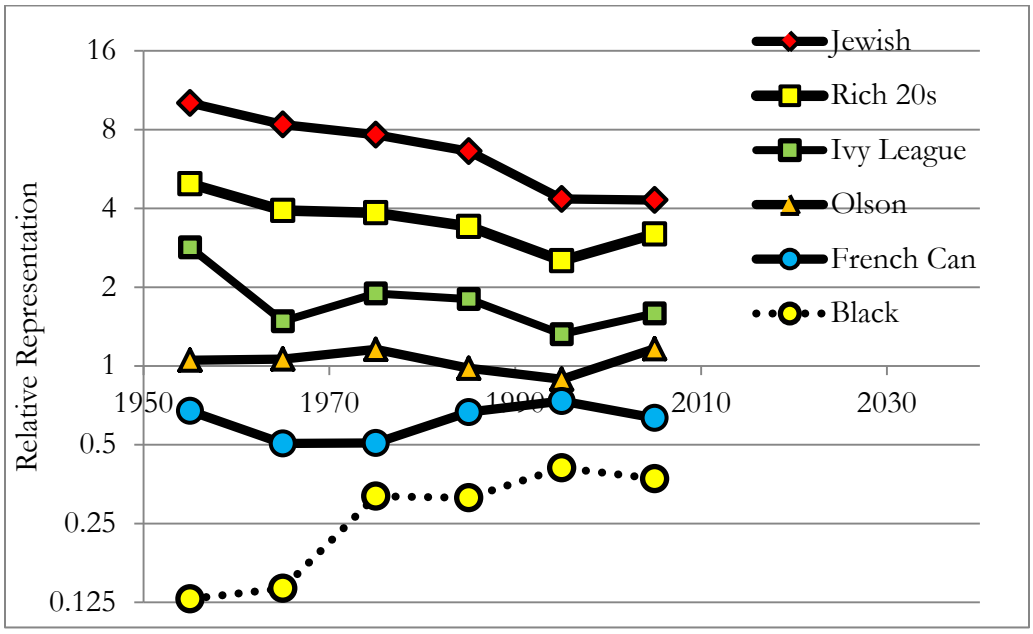


Figure 10: Relative Representation by Surname Type, by Decade, Doctors



This measure of mobility looks at what happens on average for people entering medicine between 1950-79 and 1980-2009. We can examine even more recent mobility by considering surname representation by decade, as in figure 10. This figure immediately suggests that the relatively high black mobility rates across the generations 1950-79 to 1980-2009 was likely partly a product of the dramatic institutional changes of the Civil Rights era of the 1960s, and has not been sustained in more recent generations. As a measure of the most recent mobility rates we compare relative representation of surnames among doctors in 1970-9 to that thirty years later, one generation, in 2000-9. The implied bs from this comparison are shown in table 5, under two different assumptions about the share of foreign doctors in the US in 1970-9, and using the SSDI to measure relative cohort sizes.

Table 5 suggests rates of social mobility remain very low in even the most recent data. The average b for these groups is 0.76-0.77, based on the census cohort corrections. This average is largely independent of the assumption made about the proportion of foreign doctors in 1970-9. Assuming that was 15% as opposed to 20% raises the implied b for the elite surnames, but lowers it for the disadvantaged surnames, making little difference to the average.

Although there is much less data for the earlier generations, 1920-49 to 1950-79, we can also estimate implied mobility rates here. The bs estimated for this earlier generation are even lower than those of tables 4 and 5. Correcting populations using the census, and assuming that throughout 15% of the doctor population was of foreign origin, the implied bs for each surname group 1920-49 to 1950-79 are: Jewish, 0.83, 20s Rich, 0.85, "Ivy League" 0.70, New France, 0.97, Black, 0.99. But as noted, for some of these groups the numbers of doctors observed in the earliest generation, 1920-49 is small. So other than noting that the implied overall mobility rates are again very low, we should not make much of these individual results.

Table 5: Implied bs, 1970-9 to 2000-9

Group	Pop. Share adjusted from censuses Foreign Doctors 20% in 1970-79	Pop. Share from censuses Foreign Doctors 15% in 1970-79	Pop. Share from SSDI Foreign Doctors 20% in 1970-79
Jewish	0.69	0.71	0.89
1920s Rich	0.85	0.89	0.82
Ivy League	0.72	0.79	-
French Canadian	0.68	0.63	0.75
Black	0.87	0.83	-
Average	0.76	0.77	0.82

Attorneys

Though we observe marked status differences, and slow intergenerational social mobility using doctors as a marker, the pattern observed here is indicative of a general one that will be found across all high and low status occupations. This we confirm by carrying out the same procedure with attorneys. This is more difficult because attorneys are licensed at the state level, so that the records of surname distribution among attorneys are contained in 50 different places. To make this check feasible we have thus checked surname frequencies in only a sample of states, we use a smaller set of surnames for each group, and we take the surnames Olson/Olsen as indicating the average frequency of the domestic population among attorneys by decade. Also state bar associations and court systems employ different practices with respect to recording attorneys who are now inactive. Some, such as Illinois, record even attorneys first licensed in the 19th century. Others, such as Michigan, give details only of currently active attorneys. Because surname types are distributed differently across states this introduces potential error into the process. Further, attorneys can be licensed in multiple states, and we make no attempt to eliminate multiple listings.

Using the records of just 25 states we are able, based on the distribution of doctors with these surnames, to observe 88% of the expected attorney stock of *Katz*, 86% of the 20s Rich, 71% of *Olson/Olsen*, 82% of New France surnames, and 72% of *Washington*.[8] The lower representation of Olson/Olsen comes from the fact that it is more evenly distributed across states unlike names such as Katz which are heavily concentrated in a few states.

However, with these limitations, we can confirm the patterns found among doctors. Table 6 shows the numbers of occurrences of each surname type - Katz (Jewish), the 40 most common surnames of the 1923-4 rich, Olson/Olsen, the 20 most common New France surnames, and Washington (Black) by decade in the 25 state sample. The attorney data actually goes back a bit further than that for doctors, with reasonable numbers of observations even in the 1920s. The relative representation is measured relative to Olson/Olsen. As figure 11 shows the implied

[8] These states are Alabama, Arizona, California, Colorado, Connecticut, Florida, Georgia, Illinois, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, New York, North Carolina, Ohio, Oregon, Pennsylvania, Texas, Utah, Vermont, Washington, and Wisconsin.

Table 6: Attorneys in 25 States with each surname, by Licensing Date

Period	*Katz* N	20s Rich N	*Olson/Olsen* N	New France N	*Washington* N
1920-9	13	4	15	7	3
1930-9	41	9	22	12	1
1940-9	17	9	27	10	2
1950-9	68	14	72	24	9
1960-9	108	23	91	39	6
1970-9	239	78	242	142	41
1980-9	340	70	314	256	49
1990-9	320	85	355	351	107
2000-12	306	105	448	462	116
All	1,452	398	1,586	1,303	334
N 2000	30,615	13,502	223,402	285,075	163,036

relative representation of these groups among attorneys closely echoes their relative representation among doctors for these same surnames. However, since attorneys are a less high status segment of the population what we should actually expect based on the rational of the section on measuring mobility rates would be less extreme over and underrepresentation of these surname types.

We again see a pattern of slow but persistent regression to the mean for all groups. Figure 12 shows relative representation over three generations of admissions to the bar, 1920-1949, 1950-1979, and 1980-2012 for the four surname groups used. All four surnames regress to the mean rate of representation, assumed to be that of Olson/Olsen, but at slow rates in most cases. Table 7 shows the b implied for each surname type and period by figure 12, assuming that attorneys

Figure 11: Relative Representation of Surname Types, Attorneys versus Doctors

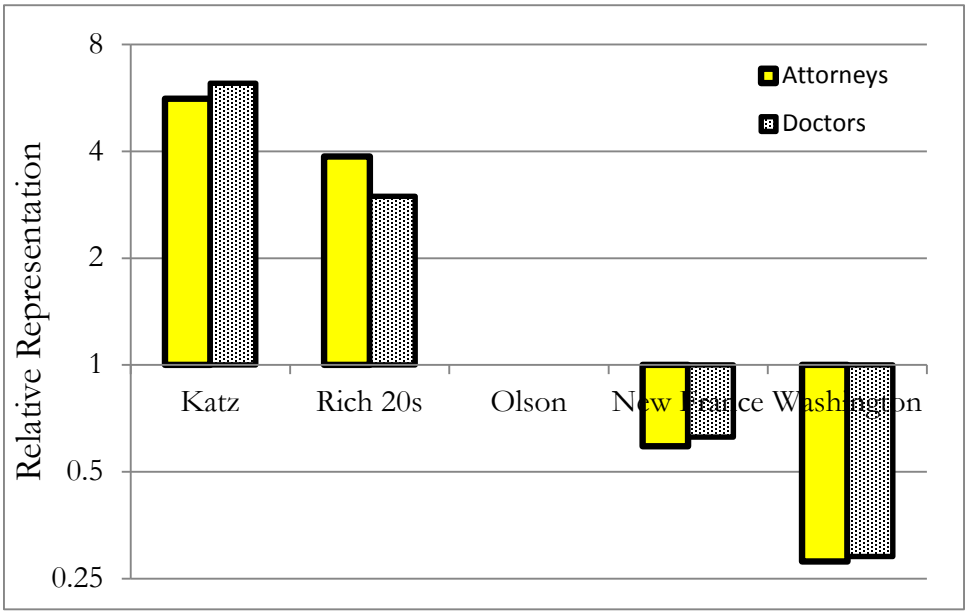


Figure 12: Relative Representation by Generation among Attorneys

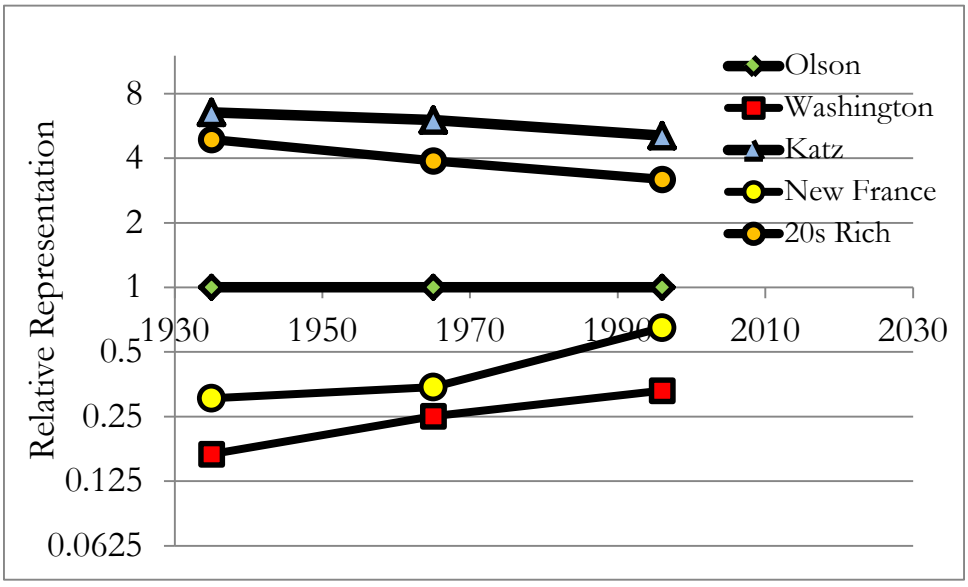


Table 7: Implied bs for Surname Groups Implied by Attorneys

Period	Katz	1923-4 Rich	New France	Washington	Average
1920-49 to 1950-79	0.95	0.84	0.91	0.79	0.87
1950-79 to 1980-2012	0.89	0.84	0.42	0.81	0.74
1970-79 to 2000-2012	0.77	0.76	0.40	1.01	0.73

represent the top 1% of the occupational status distribution (whereas doctors were assumed to represent the top 0.5% of the distribution). For 1950-79 to 1980-2012 the average implied b is very similar to that for the doctors, and again is very high at 0.74. Moving to the most recent measurement, which compares the cohort 2000-12 to 1970-9, we see no sign of any change in mobility rates. Looking at the earlier generation, 1920-49 to 1950-79 we observe even slower mobility rates, but as before these estimates are subject to substantial margins of error. The only curiosity here is that for more recent years the New France surnames have regressed much more rapidly towards the mean. But since these surnames are concentrated in a unique set of states, this may be in part just the problem of differential retention by states of older surnames on their attorney rolls.

Interpretation

Why do the results presented here differ so much from those of conventional mobility studies, even when these studies have corrected for measurement error in earnings or education?

Current one generation studies suffer a key limitation. Suppose in particular we assume that the various aspects of social status in generation t, y_t – income, wealth, education, occupation – are all linked to some fundamental social competence or status of families, x_t, such that $y_t = \theta x_t + e_t$, where e_t is some random component. The random component exists for two reasons. First there is an element of luck in the status attained by individuals given their underlying aptitudes. People happen to choose a successful field to work in, or firm to work for. They just succeed in being admitted to Harvard, as opposed to just failing. But second people trade off income and other aspects of status. They choose to be philosophy professors as opposed to finance executives.

The one generation studies, as long as y is correctly measured, will indeed report what the b is across one generation, for any particular aspect of status. However the regression to the mean exhibited by each partial measure of underlying status, y, will overestimate the regression to the mean of the underlying status x. For the $\hat{b}$ estimated from the partial measures will be related to the b for the underlying status through

$$\hat{b} = b\frac{1}{1+\left(\frac{\sigma_e^2}{\theta^2\sigma_x^2}\right)}$$

where σ_x^2 is the variance of the underlying social status, and σ_e^2 is the variance of the random components linking the underlying status to the measured aspect.

But it is this underlying b that governs long run social mobility, and that also governs mobility on more comprehensive measures of status. Suppose the bs across generations for income, education, occupational status, and wealth were all 0.3. It does not follow that the regression to the mean across two generations will be b^2, so that initial differences in social status quickly disappear. It also does not follow that the b for a more general measure of status that averages income, wealth, education and occupational status would be 0.3, or even anywhere close to 0.3. When we classify people by religion, race, or ethnic or national origin, the b that applies to such groupings will also not be 0.3.

The conventional studies have been often misinterpreted as speaking more generally about the mobility of society than they can, as we see in figure 1 above. When we classify families as high or low status based on partial measures such as income, wealth, education or occupation there will appear to be substantial regression to the mean. But if we took a more aggregate measure of status, which averaged the various partial y measures the regression will be substantially lower. These partial measures are correlated, so that with such an aggregate measure the variance of the error term will decline relative the variance of x. [9] So the measured intergenerational correlation will be greater.

With conventional measures, the intergenerational correlation of status, even for broad measures of status, will not predict long run social mobility across many generations. And its relationship to long run mobility will depend on the relative importance of the error components in the first generation. This is because when we classify families as high or low status originally based on their status in an original generation, the measures incorporate random errors of measurement and luck. The regression to the mean observed in the first generation incorporates this error correction, and that component will not occur across subsequent generations. So even the regression of an aggregate status for a first generation t to t+1 will be greater than that for the same families from generation t+1 to t+2, and so on.

The greater are the random components in determining measures of status such as income, relative to the systematic elements stemming from underlying status, the greater will be the degree of mismatch between such partial one generation estimates of regression to the mean and the underlying regression of fundamental social status. The USA, for example, has much greater inequality in earnings than does Sweden. Figure 13 shows, for example, the salaries in $2010 for some comparable high and low status occupations in Sweden and the USA. A US doctor earns 6 times the wage of a bus driver, while in Sweden the ratio is only 2.3 times. A US professor earns 60% more than a bus driver, in Sweden it is only 40% more.

This can be interpreted as meaning that the θ in the expression

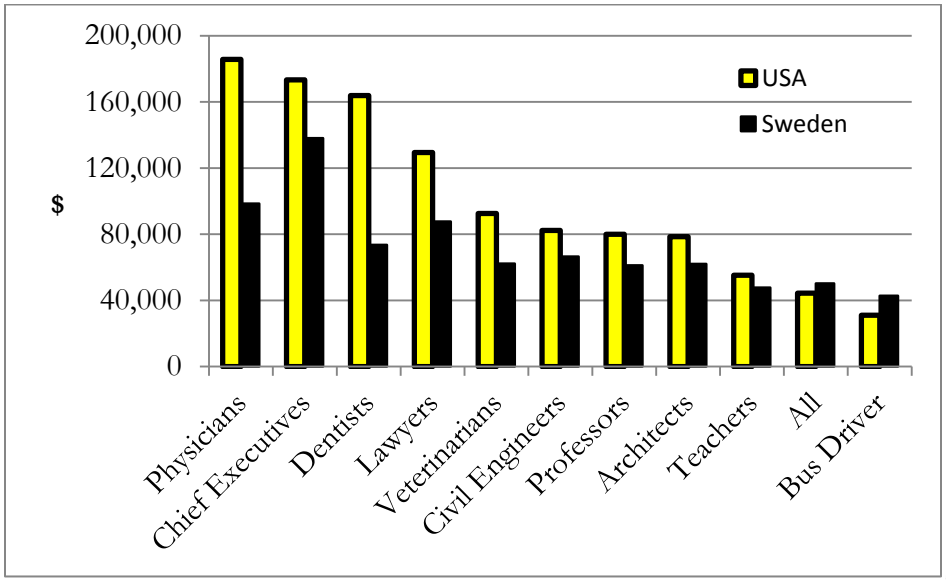
$$y_t = \theta x_t + e_t$$

linking social status to earnings is higher in the US than in Sweden. That in turn implies that the measured b for earnings will be lower in Sweden than in the US even were the underlying rate of regression to the mean of status the same, because the share of earnings variation contributed by random elements is greater in Sweden.

---

[9] Indeed, if people are trading off aspects of status the individual error elements will be negatively correlated, so reducing even further the aggregate error.

Figure 13: Average Earnings by Occupation, Sweden and the USA, 2008



Sources: USA – Bureau of Labor Statistics, National Occupational Employment and Wage Estimates, May 2010. Sweden – Statistics Sweden, Wage and salary structures, private sector (SLP), 2011.

Table 8: Estimates of b from Surnames

Country	Measure	Period	b
England	Attorneys, Doctors	1950-2012	0.69-1.00
England	Wealth	1950-2012	0.70
England	Education	1950-2012	0.77
Chile	Occupations	1940-2010	0.74
China	Education	1905-2011	0.71
India	Doctors	1900-2010	0.80
Japan	Education	1940-2012	0.84
Sweden	Doctors	1950-2012	0.71
Sweden	Attorneys	1950-2012	0.88
Sweden	Education	1950-2012	0.66-0.84

Sources: England, Clark and Cummins, 2012**,** China, Hua and Clark, 2012, India, Clark and Landes, 2012, Japan, Clark and Tatsuya, 2012, Chile communication from Daniel Diaz, Sweden, Clark, 2012.

Studies of surname distributions in a variety of countries, using the methods employed above, find the results in table 8, for recent implied persistence of various measures of status. There we see that in general the implied b for the last few generations is very similar to the high numbers estimated for the USA above. Sweden, in particular, no longer looks like a society with high rates of social mobility compared to the USA. Scandinavia is no longer an exception to the rule, but is instead an exemplar of the existence of high rates of persistence in underlying status in all societies.

Thus there is no support here for three propositions that have gained currency in recent discussions of mobility. The first is that the USA has low rates of social mobility compared to other high income societies (Corak, 2006, Jäntti, 2006). Generalized mobility rates are no lower in the US than elsewhere. The second is that rates of social mobility have declined recently in the USA.[10] And the third is that there is a link between income and wealth inequality and social mobility rates (Corak, 2012). Instead the rate of generalized social mobility seems to be closer to a universal constant across societies, changing little across social systems and epochs.

Even though we measure social mobility here using surname frequencies among social elites, the slowness of mobility cannot be explained as being a peculiar property just of the upper end of the status distribution. For then we would see a mismatch in status for groups between the upper percentiles of the distribution and the lower percentiles. Yet we see above in figures 5 and 6 that the high and low representation of the Jewish and Black subgroups among doctors and attorneys is coupled with a reverse under and overrepresentation at the lower end of the educational attainment distribution.

The fact that mobility rates for social groups, such as Jews or Blacks will be measured by the underlying b, rather than the conventionally measured b, means that if indicators for such groups are included in conventional intergenerational mobility estimates then it will appear that these groups are not regressing to the social mean. This is exactly what Hertz (2005) finds with reference to both the Black and Jewish sub-groups in the NYLS when he measures intergeneration income mobility, in this case using the log of family income. Table 9 shows his estimated regression

[10] Though this notion has gained popular currency (see, for example, Foroohar, 2011) there seem to be no academic studies suggesting it.

Table 9: Regression to the mean controlling for race and religion, USA

Independent Variable	No controls	Only Race	All Observable Parental Characteristics
Ln Family Income of Parents	0.52**	0.43**	0.20**
Black	-	-0.33**	-0.28**
Latino	-	-0.27**	-0.15
Jewish	-	-	0.33**

Notes: ** = significant at the 1 percent level. Only 3 percent of the sample was Latino.

Source: Hertz, 2005, table 6.

coefficients, with and without dummies for race and religion, for a sample of 3,568 parental incomes in 1967-71, and the income of adult children in 1994-2000.

The regression estimates imply that, even when we control for all other measured attributes of parents in 1967-71 such as education, occupation, and household cleanliness, we can predict that Black, Latino and Jewish families are all regressing more slowly to the mean than is found for the population as a whole.[11] The Hertz interpretation is that this is because of special characteristics of these groups. Our interpretation, however, is that if we included a dummy for membership in any high or low income group, such as the descendants of the 1923-4 rich, then it would have a significant coefficient also. This is because the underlying rate of regression to the mean for all families is much lower than the conventional regression estimates imply. Thus once we can identify families as collectively belonging to groups of on average high or low incomes, we can predict much better the expected income in the next generation.

[11] Hertz, 2005.

This same effect of group background was found by George Borjas in his study of immigrants where he regressed

$$y_{ijt+1} = b_0 y_{ijt} + b_1 \bar{y}_{jt}$$

where y was log wage or years of education, i indexed families, j the country of origin of fathers, and t the generation (Borjas, 1995). $\bar{y}_{jt}$ was the average log wage or years of education of all men from that country, estimated from the 1980 census reports of education and occupation. In both the case of education and earnings the average status of people from the country of origin was predictive of the outcome for sons (b_0+b_1 equalled 0.44 for education and 0.70 for earnings) (Borjas, 1995, table 8).

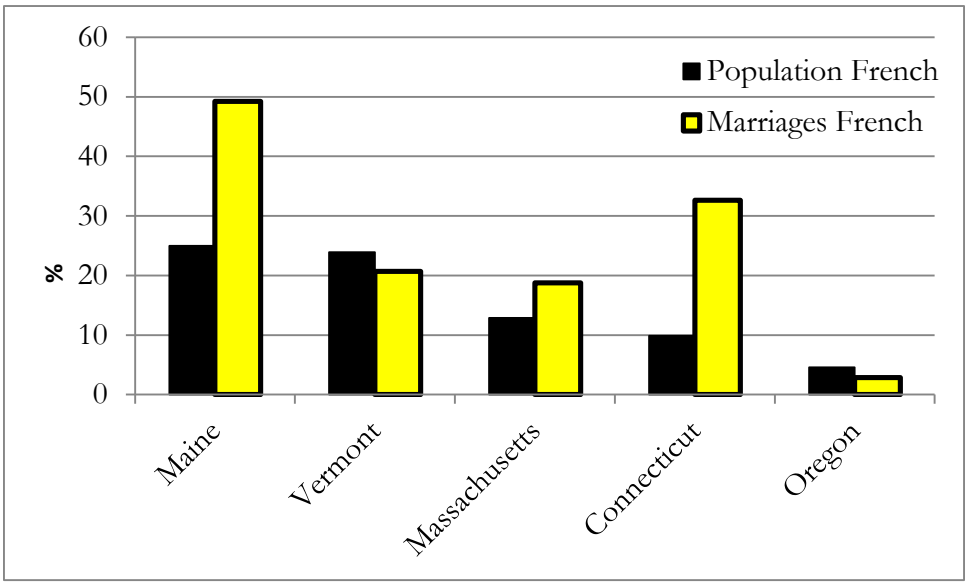
Borjas interprets this as the result of "ethnic capital" externalities. Sons from ethnic groups with high average education levels do better than would be predicted from the education of the father alone, because of spillovers from the education of others in the community. But again our interpretation would be that there is likely little or no externality. It is just that information on the country of origin allows a better prediction of the likely "true" underlying status of families, and so a better prediction of the son's outcomes. That is why the same effect appears for the wealthy of 1923-4, who span many ethnic communities.

The Case of New France

The low representation of surnames whose origin lies in the French settlers of New France in the colonial era among doctors and attorneys is a surprise. This group has not previously been identified as an underprivileged minority in the US. By design the surnames selected in this group had less than 5% of holders in the 2000 census declaring themselves Black. Thus these surnames largely exclude the common surnames of the Cajun population of Louisiana, such as *Landry* where 12% of holders in 2000 declared themselves Black. They instead tend to be concentrated in the New England, either as a result of the takeover of parts of Acadia in the eighteenth century by the American colonies, or as a result of immigration 1865-1920 of French Canadians from Quebec and New Brunswick. So their low representation in the elite cannot be attributed either to their being concentrated in poor areas of the US. As a largely invisible minority their low representation among the medical and legal elites also cannot stem from acts of discrimination.

What then explains the low social status of these surnames? One possible explanation, that Borjas has emphasized in his work, is the "cultural capital" of those of New French descent. Could they as a community have a cultural legacy that impeded and impedes upward mobility? However, interestingly, even if we go back to the 1950s and to states with many people of New French descent, rates of

Figure 14: Marital Endogamy among the New France Descendants, 1950s



Source: Marriage rates from Ancestry.com.

intermarriage between those with New France surnames, and those of surnames of other heritages have been very substantial. Figure 14 thus shows the percentage of those in four New England states, and in Oregon, of French American heritage, according to the 2000 census. Also shown is the fraction of those women with a selection of New France surnames (Gagnon, Belanger,…) whose groom at marriage 1950-9 also had a New France surname. By the 1950s a large majority of New France descendants appear to have been marrying outside that community, even in Maine and Vermont where they constitute a quarter of the population.

An alternative explanation of the low socio-economic status of French Canadian surnames in the USA is that the founders of New France represented a draw from the lower ends of the French ability distribution. The modern population of 6-8 m. people of New French descent derives from a rather small stock of immigrants to the French possessions in the seventeenth and early eighteenth centuries. It is estimated, for example, that the French population of Canada in the late nineteenth century derived from fewer than 8,500 original French settlers (Scriven, 2001, 76). But even within this settler population, some people contributed much more to the modern genetic stock than others. Quebec, for example, has a high rate of incidence of Mendelian (single gene) diseases. This has been argued to stem from a disproportionate contribution of a small share of the founder population to the modern genetic stock in French Canada. "As few as 15% of the founders could account for 90% of the total genetic contribution from the founders" (Scriven, 2001,

78). One of the distinctive features also of the demographic regime in Quebec in the seventeenth to nineteenth centuries was that the most reproductively successful group in the population was the lower socio-economic group.

Conclusions

This paper establishes through analysis of surname distributions that the underlying social mobility rates in the US in the period 1920-2012 are much lower than conventional estimates of social mobility would suggest. The conventional estimates correctly predict how well particular aspects of social status will be inherited. However, if we are concerned at what rates of social mobility will be for more comprehensive measures of social status, these conventional estimates will suggest much faster mobility than actually applies. The surname estimates will prove the applicable ones in these cases. Also if we ask what mobility rates will be for ethnic or religious groups high or low in the status distribution, again the surname estimates will be the ones that apply. If we ask what mobility rates will be across multiple generations, again the surname estimates are the relevant ones.

References

American Medical Association. 2006. *International Medical Graduates in the U.S. Workforce: A Discussion Paper.*

American Medical Association. 2010. *Directory of Physicians in the United States.*

Black, Sandra E. and Paul J. Devereux. 2010. "Recent Developments in Intergenerational Mobility." *Handbook of Labor Economics.*

Borjas, George J. 1995. "Ethnicity, Neighborhoods, and Human-Capital Externalities." *American Economic Review*, Vol. 85, No. 3 (Jun.): 365-390.

Corak, Miles. 2006. "Do Poor Children Become Poor Adults? Lessons from a Cross Country Comparison of Generational Earnings Mobility." IZA Discussion Paper No. 1993.

Corak, Miles. 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, forthcoming.

Hertz, Thomas. 2005. "Rags, Riches and Race: The Intergenerational Mobility of Black and White Families in the United States" in Samuel Bowles, Herbert Gintis and Melissa Osborne (eds.), *Unequal Chances: Family Background and Economic Success*, 165-191. New York: Russell Sage and Princeton University Press.

Hertz, Thomas, Tamara Jayasundera, Patrizio Piraino, Sibel Selcuk, Nicole Smith and AlinaVerashchagina. 2007. "The inheritance of educational inequality: International comparisons and fifty-year trends." *B.E. Journal of Economic Analysis & Policy 7*, Article 10.

Hobbs, Frank and Nicole Snoops. 2002. *Demographic Trends in the 20th Century.* Washington, D.C.: U.S. Census Bureau.

Foroohar, Rana. 2011. "What Ever Happened to Upward Mobility?" *Time*, Nov. 14.

Jäntti, Markus, Knut Røed, Robin Naylor, Anders Björklund, Bernt Bratsberg, Oddbjørn Raaum, Eva Österbacka, and Tod Eriksson. 2006. "American Exceptionalism in a New Light: A Comparison of Intergenerational Earnings Mobility in the Nordic Countries, the United Kingdom, and the United States." IZA, Discussion Paper #1938.

New York Times. 1924. *Taxpayer Listings.* 10/24 to 11/2, 11/4-11/8, 11/11-11/14.

New York Times. 1925. *Taxpayer Listings.* 9/2 -9/6, 9/8-9/11, 9/13.

OECD/WHO, 2010. Policy Brief on the International Migration of Health Workforce. Feb. www.oecd.org/health/workforce

Scriver, Charles R. 2001. "Human Genetics: Lessons from Quebec Populations." *Annual Review of Genomics and Human Genetics, 2*: 69-101.

Ward, David L., Charles D. Coleman, Robert Nunziata, and Robert Kominski. No Date. *Demographic As[ects of Surnames from Census 2000.* Washington, D.C.: US Bureau of the Census. http://www.census.gov/genealogy/www/data/2000surnames/

"Ivy League" Graduates

Brown University.

Columbia University.

Dartmouth College.

Harvard University**.** 1915. *Quinquennial Catalogue of the Officers and Graduates of Harvard University, 1636-1915*. Cambridge, Mass.: Harvard University.

University of Pennsylvania.

Rutgers University.

College of William and Mary. 1941. *A Provisional List of Alumni, Grammar School Students, Members of the Faculty, and Members of the Board of Visitors of the College of William and Mary in Virginia, from 1693 to 1888: Issued as an Appeal for Additional Information.* Richmond, VA: College of William and Mary.

State Attorney Directories

Alabama State Bar
Alaska Bar Association
State Bar of Arizona
Arkansas Judiciary
State Bar of California
Colorado Supreme Court
Connecticut Statewide Grievance Committee
D.C. Bar
Florida Bar
State Bar of Georgia
Hawaii State Bar Association
Idaho State Bar
Supreme Court of Illinois
Indiana Supreme Court
Iowa Judicial Branch
Kentucky Bar Association
Louisiana State Bar Association
Maine, Board of Overseers of the Bar
Client Protection Fund of the Bar of Maryland
The Massachusetts Board of Bar Overseers
State Bar of Michigan
Minnesota Judicial Branch, Lawyer Registration Office
Mississippi Supreme Court
Missouri Bar
State Bar of Montana
Nebraska State Bar Association
State Bar of Nevada
New Jersey Courts
State Bar of New Mexico
New York State Unified Court System
North Carolina State Bar
North Dakota Supreme Court
Supreme Court of Ohio
Oregon Bar Association
Disciplinary Board of the Supreme Court of Pennsylvania
Rhode Island Judiciary
South Carolina Bar
Board of Professional Responsibility of the Supreme Court of Tennessee
State Bar of Texas

Utah State Bar
Vermont Judiciary, Attorney Licensing
Virginia State Bar, Membership Department
Washington State Bar Association
West Virginia State Bar
State Bar of Wisconsin
Wyoming State Bar

Some states do not have any publicly available attorney directories.