Caprettini, Bruno; Voth, Hans-Joachim

Working Paper

Wages and the Great War: Evidence from the largest draft lottery in history

Working Paper, No. 441

Provided in Cooperation with:
Department of Economics, University of Zurich

Suggested Citation: Caprettini, Bruno; Voth, Hans-Joachim (2023) : Wages and the Great War: Evidence from the largest draft lottery in history, Working Paper, No. 441, University of Zurich, Department of Economics, Zurich, https://doi.org/10.5167/uzh-239553

This Version is available at:
https://hdl.handle.net/10419/280777

University of Zurich

Department of Economics

Working Paper Series

ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 441

Wages and the Great War: Evidence from the Largest Draft Lottery in History

Bruno Caprettini and Hans-Joachim Voth

November 2023

WAGES AND THE GREAT WAR: EVIDENCE FROM THE LARGEST DRAFT LOTTERY IN HISTORY*

Bruno Caprettini
University of St. Gallen and CEPR

Hans-Joachim Voth
University of Zurich and CEPR

November 2023

Abstract: Do veterans earn less? During WW I, the US organized "the greatest human lottery in history": a random draft of 24 million men. Ultimately, 2.8 million Americans were selected to join the armed forces. We sample 10% of registrants of the 1917 lottery and match these men with the 1930 and 1940 US Federal Censuses. Low lottery numbers significantly increased the likelihood of serving in World War I. Importantly, military service also had a positive causal effect on earnings and occupational outcomes. Veterans joined professions with higher cognitive skill requirements, including higher intelligence, language, reasoning, and math requirements. Randomly-assigned military service had fundamentally different effects during World War I than in Vietnam. We rationalize this finding by analyzing complier characteristics.

Keywords: veterans' income, lottery, IV, effect of war participation

JEL Classification: N42, J45, I23, J24, N32

* We thank Ewan Rawcliffe, Giacomo Gattorno, Luca Vitale, Matteo Machiorlatti, and Marcel Caesmann for outstanding research assistance, and the SNF for generous funding (SNF 100018_176455).

1 Introduction

Violent conflict is as old as mankind, and continues to affect the lives of millions of people (Blattman 2023; Morris 2014). In addition to civilian and military casualties and the destruction of property, armed conflict also affects the lives of the men and women who fight. One research question in particular has attracted attention in economics: the effect of military service on occupational outcomes. Millions of Americans served in the armed forces during the last century. Following Angrist's (1990) seminal work using the Vietnam era lottery as an instrument, numerous papers have confirmed and extended the key finding that exogenously induced military service tends to lower income and leads to worse economic outcomes.[1] While the majority of empirical work has focused on the Vietnam era, similar, negative effects have been documented for a range of countries and periods – veterans in general do unambiguously worse in the labor market.

Like all instrumental variable strategies, the Vietnam lottery identifies the causal effect on compliers – young men who, at the time, would not otherwise have served in the military. As such, the estimated effect depends on the specific economic, social, and cultural context in which it occurs. Arguably, conditions for veterans returning from the battlefields of South-East Asia in the 1970s were special, and arguably unfavorable: Vietnam was the first war the United States unambiguously lost. The war divided the nation, and numerous reports about atrocities by US forces, such as the infamous My Lai massacre, undermined the public image of soldiers. Veterans returned at a time when the US economy's great post-war boom was coming to an end, with unemployment rising and foreign competition increasing in many sectors. All these factors suggest that the external validity of Angrist's famous Vietnam lottery finding can be called into question.

In this paper, we present evidence of a causal, positive effect of military service on earnings and occupational outcomes. During WWI, as in other conflicts, physical fitness

[1] For example, Angrist and Krueger (1994) find either negative or zero effects for US WW2 veterans. Subsequent research on Vietnam by Angrist et al. (2011) suggests that the negative effect on labor market outcomes for veterans faded in later life.

requirements may have induced positive selection for actual veterans; at the same time, the fact that some skilled laborers and managers were exempt from the draft to maintain economic production may have induced negative selection (Tan 2020). We use randomized treatment arising from the US WWI lottery to identify effects. The Selective Service Act of 1917 introduced a service obligation for all adult men. Those aged 21-31 (later 18-45) were given priority for conscription. Men were selected by what was hailed at the time as the "the greatest human lottery in history". Carried out four times, it registered a total of 24 million men. Ultimately, 2.8 million men were inducted into the US Army. We focus on the first draft lottery, which included 10 million registrants, resulting in 650,000 inductees.

We first show that lower order numbers, as assigned by the lottery, significantly increased the chances of military service – our first stage is strong. Second, we link 170,000 of the 1917 registrants to their census records in 1930, and code the characteristics of their occupations using skill requirements for each profession from the Dictionary of Occupational Titles (DOT). Occupational outcomes are unambiguously positive for draftees – the higher the (lottery-induced) probability of serving in the US Army during WWI, the more they sorted into demanding, better-remunerated, and prestigious occupations. Average occupational scores increased from the equivalent of window dressers (24.96) to postmasters (29). When we focus on *changes* in occupational outcomes, taking into account 1917 starting levels, we find a very similar pattern. With better occupations came higher cognitive, language, mathematics, and reasoning requirements. The effects are quantitatively large: The intelligence requirement of the typical job taken by our controls was equivalent to that of watchmen or roofers; those induced by the lottery to serve ended up, on average, in professions with intelligence requirements of detectives, typesetters, or ship captains. The average math requirement increased by the equivalent of the jump from huckster/peddler to pressman and engraver, and average language ability increased by the equivalent of the distance from cooks to millwrights. Earnings in these occupations were significantly higher than those of registrants not selected by the lottery; those induced by the lottery to serve in the military during WWI earned more than their peers, by 8-13%, by 1930; equivalent to the difference in income between the average tinsmith and the average machinist.

What can explain this positive effect, and the sharp contrast with the post-Vietnam experience? Both the Vietnam and the WWI lottery provide equally credible identification – yet

the sign of the treatment effect is exactly the opposite. We argue this is highly context-dependent – WWI veterans returned to an economic, social and political environment that differed sharply from the one faced by Vietnam-era veterans. Once war had been declared on Germany, the nation was united in its determination to prevail. Demobilized soldiers were hailed as victors, having brought a long and painful conflict to a quick and victorious end. Popular Hollywood films such as "Sergeant York" celebrated the exploits of the 'doughboys'. After a short, sharp recession in 1920-21, the US economy entered a prolonged expansion that only came to an end in 1929-.30. New production techniques, systematic R+D investment, and new management methods raised labor productivity. In contrast, after Vietnam, a generation of disillusioned veterans came back to a deeply polarized society struggling with stagflation.

We can rule out that patriotism is the key mediating variable that explains the positive effect of military service post-WWI. More patriotic areas – those that saw more volunteering, or medal-recipients – are not the ones with the highest returns to military service. If anything, effects may be even negative there.

1.1 Related literature

Our results relate to the literature on the effects of war exposure on labor market outcomes. In his seminal contribution, Angrist (1990) showed that 10 years after fighting in the Vietnam war, white U.S. veterans earned up to 15 percent less than comparable men who did not fight. Angrist is able to identify the causal effect of war on earnings because the Vietnam draft lottery generates random variation in war participation. Several studies have confirmed and expanded these results: Small and Rosenbaum (2008) Angrist, Chen and Song (2011) and Angrist and Chen (2011) look at U.S. veterans; Siminski (2011) at Australian veterans. All of these papers estimate the causal effect of military service. However, all of them focus on the Vietnam war, for which the U.S. and Australian draft lotteries allow clean identification. Estimates of the effect of different wars on the earnings of veterans are presented by Teachman and Tedrow (2004) for World War II U.S. veterans, Lemieux and Card (2001) for Canadian WWII veterans, Costa (2010 and 2014) for Union Army veterans after the U.S. civil war. (2022) show that in the all-volunteer army since 1990, especially black veterans earn significantly more, using an RDD design. With the exception of the all-volunteer study, the estimates of the effect of these other wars do not rely on "quasi-natural experiments" such as the Vietnam draft lottery, and may

suffer from selection bias (because healthier, more skilled men are more likely to volunteer and be selected to fight, on average).

Studies of the effect of wars on earning can be separated into two groups. The first one exploits the Vietnam draft lottery to estimate the causal impact of the Vietnam war, while the second one examines the effect of other wars, in contexts in which a causal interpretation of the estimates is harder to defend. The only exception is Angrist and Krueger (1994) who use the lotteries run in 1943-47 to show that WWII U.S. veterans were likely to earn less than comparable men who did not fight.

We are not the first to examine the case of service during WWI. Tan (2020) looks at US veterans' earnings during the interwar period. He finds no significant effects on income from military service. However, his identification strategy is fundamentally different from ours. Tan uses quarter of birth of potential recruits to identify those at risk from conscription.

2 Historical Background

2.1 Conscription in the U.S. during World War I

The United States declared war on Germany on the 6th of April, 1917. On that day, the U.S. army was 200.000–300.000 men strong, a fraction of the size of European armies (Ayres, 1919; Grotelueschen, 2007). Faced with the need to expand the army quickly, Congress passed the Selective Service Act, authorizing President Wilson to enlist one million men. While expanding the army with volunteers only was a possibility, Wilson's administration opposed this alternative, fearing that it would weaken domestic production.[2] Wilson was also wary of the experience of conscription during the Civil War, when the draft was considered unfair by some. For this reason, the administration decided to draft men through what it considered the fairest system: a random lottery. On the 5th of June 1917, every male resident 21 to 30 years old was

[2] General Crowder explains this worry: "If farms, factories, railroads and industries were not to be left crippled, if not ruined, by the indiscriminate volunteering of key and pivotal men, then, in the face of such an enemy as Germany, the total military effectiveness of the Nation would have been lessened rather than strengthened by the assembly of 1.000.000 volunteers." (Crowder 1918: p. 6; emphasis added).

required to register and make himself available for military service. Two months after registration, a national lottery determined which men would join the army.

Within a year from the lottery, the U.S. army sent more than half a million draftees to France. The lottery was held in Washington D.C. in the summer of 1917, the first conscripts started training in September 1917 and landed in France in the spring of 1918 (Crowder, 1918: p. 469). Between April and September 1918 seventeen National Army divisions with 687.000 draftees arrived in France. Starting from June 1918 three additional lotteries extended the draft to all men 18 to 45 years old: these additional draftees brought the strength of the U.S. army to around 4.7 million men (DeBruyne and Leland 2020), but almost none of them saw combat, because by November 1918, Germany had surrendered and WWI was over.

Draftees were selected in four stages. First, on the 5th of June 1917, every U.S. male resident between 21 and 30 years old had to register for military service in one of 4322 local boards. At registration, men provided personal information and received a lottery number, unique within their local board (see the top-left of a sample registration card on Figure A.1: the number appears in red ink). Second, on the 20th of July 1917, a large lottery in Washington D.C. determined the order in which these draft numbers would be called. During the lottery, high U.S. officials and public clerks drew numbers from a fishbowl and announced them to the press (Figure A.2).

Third, local boards used the lottery order to call men and induct them in the army. Reasons for exemptions included dependents (e.g. children), employment in key sectors (e.g. weapon production) and physical unfitness (e.g. disability). Each board had to provide the army with a given number of soldiers, and called men until it reached this quota. The quotas were a function of board size, minus the volunteers who joined the army before the lottery, and equalized the share of soldiers across the country. Volunteering rates varied substantially by state (Figure 1). In some states and territories, 60 to 75% of the men enlisted in the army were volunteers (Oregon, Vermont, Wyoming, DC); in others they accounted for less than a quarter (Oklahoma, Louisiana, Montana). As a result, where many men had volunteered already, fewer men were drafted. Fourth, selected men were inducted in the army. Conscripts received a few months of training in one of the U.S. camps and then sailed to France, where they usually spent the first weeks in a quiet sector of the front. After that, they were ready for combat, and fought along the whole Western front until the end of the war.

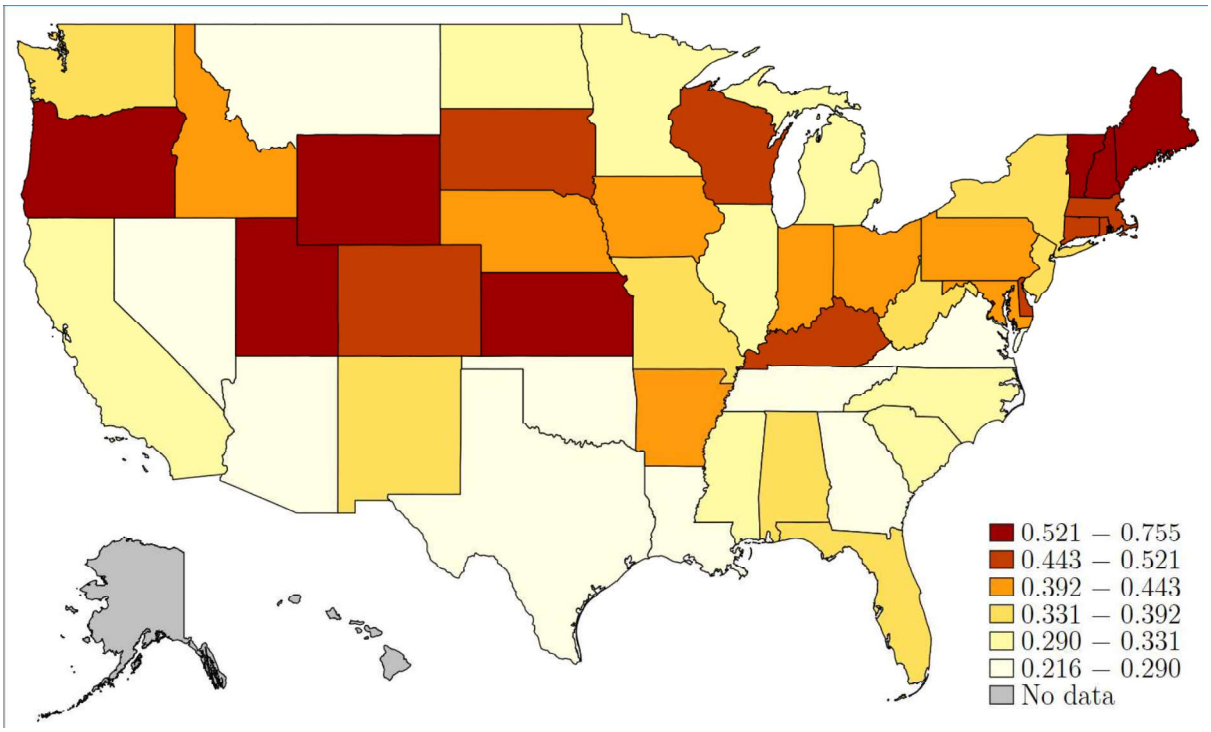


Figure 1: Volunteering rates by state

2.2 The American Expeditionary Force

The high command of the American Expeditionary Force kept conscripts separated from volunteers, and organized drafted men into a "National Army" of 18 divisions (Nishiura, 1989: p. 515). Career officers from the Regular Army (the professional all-volunteer force before 1917) always retained the command of these divisions. However, because the AEF was short of trained personnel, it selected lower rank officers during the first weeks of training among the best draftees. On April the 13th 1918 the first National Army division arrived in France. The other 17 divisions disembarked between April and September of the same year and fought until the end of the war. Despite their brief presence on the Western Front, these soldiers took part in the most important battles of 1918: the battles of Aisne-Marne in June-July; St. Mihiel in September and Meuse-Argonne in September-October. During Germany's final offensives, they blunted the German army's momentum. The rapidly growing strength of the AEF was arguably a key contributing factor to Allied victory.

Fighting WWI required different training, doctrine, and equipment than conflicts that the US had faced before. When President Wilson declared war on Germany, the U.S. Army had just returned from the punitive expedition against Pancho Villa (March 1916-February 1917). This war had reinforced the doctrine developed within the U.S. Army after the Civil War, which discounted the importance of firepower and artillery, and placed emphasis on the courage and initiative of infantry. Infantrymen were expected to take positions through superior maneuvering and with almost no supporting fire. Moreover, the plan of the battles was often only sketched, because the objective of every attack was to break through the defensive lines and penetrate deeply into enemy territory.

These ideas were similar to those held by European commanders at the beginning of the conflict, in August 1914. They caused massive causalities during the first years of the war. However, European generals had had three years to study the new battlefield, and learn the importance of artillery to protect the infantry. In contrast, the AEF had to learn all of it in just a few months. One of the reasons why American soldiers managed to adapt relatively quickly to trench warfare was the high level of delegation on which the AEF relied, another of the tenets of the U.S. warfare doctrine. After a few months of fighting, lower rank officers at the company and battalion level would coordinate with the artillery brigades, and they would decide which of the assigned objectives was feasible and which ones were not. The reliance on pre-war doctrine explains the very high death rates experienced by U.S. divisions during the first months of battle. The ability of lower rank officers to learn from experience helped to reduce the death rate quickly (Appendix Figure A.3).

The high death rate faced by American soldiers may have affected the composition of those who served in WWI, but survived – if killed men were systematically different from survivors. This argument is difficult to examine directly – we do not have granular information on the characteristics of casualties. Since most casualties of WWI fell victim to artillery strikes, selectivity was probably less than during other conflicts. Second, if there was selection, it was probably skewed towards those with more characteristics that are useful on the labor market: First infantry officers had to lead the charges against enemy lines; they were around 20 percent more likely to die in combat than the average soldier (see Appendix Figure A.4, taken from Ayres, 1919). Second, as explained above, around 83 percent of U.S. officers trained after the

declaration of war, and they were selected from the best draftees. As a result, it is likely that the way WWI killed men affected the composition of the men who survived it negatively.

3 Data

The Selective Service Acts of 1917 and 1918 made every 18-45 year old man living in the United States liable for military service. Of these men, those born between the 5th of June 1896 and the 6th of June 1886 make up our population of interest. These cohorts were between 21 and 30 years old on the 5th of June 1917 , and were subject to the first draft lottery. Other cohorts registered to the subsequent 1918 draft lotteries, and had much lower probabilities of serving in the Army. Ten million men registered during the first draft, and 700 thousand of them joined the Army after being selected in the July 1917 lottery.

We collect basic information of everyone who took part in the 1917 lottery from the National Archive Record Administration (NARA). We use their name, date of birth and State of birth to link their records to the U.S. Federal Censuses of 1930 and 1940. We then select a random sample of the men we match to both Censuses: for these individuals, we reconstruct the outcome of the draft lottery and create several labor market outcomes. In this section we describe each of these steps. We leave further details in Appendix A and the description of the record linking procedure in Appendix B.

3.1 Sample

Our main database consists of a random sample of 350,000 men who took part in the 1917 lottery. To construct this sample, we start from the full list of ten million men who took part in the 1917 lottery, digitized from their original registration cards. For every man, we observe basic information including first and last name, as well as year and place of birth. We use this information and the method of Abramitzky, Boustan and Eriksson (2012) to link these men to their records in the 1930 and 1940 U.S. Federal Censuses, where we look for individuals with similar names born in the same state and within a two-year window interval (see Appendix B for details). Out of 10.2 million men observed in 1917, we find 2.2 million in the 1930 Census, 1.9 million in the 1940 Census and 1.2 million in both. We focus on these men which we observe 13

and 23 years after the war: this ensures that we only digitize information of men for whom we have data after the war. We select a random sample of 512'829 men, stratifying by age and state of birth. Table A.1 shows that, as it is common with historical record linking, compared to the full population, matched men in the Censuses are more likely to be white, to be born in the U.S. and to have uncommon names (cols. 4 and 6). The same table shows that, compared to the matched men, our random sample is similar along most observable characteristics (col. 5 and 7). The final sample of 350,000 men consists of observations for which we have valid information for our main 1930 variables (veteran status and occupation).

3.2 Main Variables and Sources

The three key variables for our analysis are the result of the 1917 lottery, veteran status and labor market outcomes in 1930: in this section we describe how we construct each of them. To determine the outcome of the draft lottery, we obtain the image of his registration card for every man in our sample: Figure A.1 reports an example of one such card. Next, we digitize the draft number: these numbers were recorded in red ink on the top-left corner of the cards, and were unique within the recruitment board. Finally, we reconstruct the order in which draft numbers were drawn during the lottery held in Washington DC on the 6th of July 1917. We use the list in three separate newspapers published the day after the lottery, which we cross-check to minimize errors (the three newspapers are: the New York Times, the Altoona Tribune, and the Los Angeles Evening Herald, Appendix Figure A.5 shows two pages from these newspapers as examples). The outcome of this procedure is a (random) number reflecting the order in which each registrant was called by the recruitment office. Called men did not enlist automatically: only men fit for service were drafted. However, men with lower order numbers were called earlier and had a higher chance of being enlisted in the army: we use these numbers as our instrument for veteran status.

We observe veteran status in the US Federal Census of 1930. In that year, the Census collected veteran information from every adult male, including period of service. Around 4 million men were veterans in 1930, and 89.4 percent of them fought in WWI. Our main treatment is a dummy for being a WWI veteran and we report results using a dummy for being a veteran of any war as robustness.

We use reported occupation to construct several labor market outcomes. We observe occupation in both 1917 (from the registration cards) and 1930 (from the Census). We develop novel crosswalks to link these occupations to the list of occupations reported in the 4th edition of the Dictionary of Occupational Titles (National Academy of Sciences 1970, ICPSR 2006 – see Appendix A for details). The procedure assigns to every occupation scores representing the importance of separate skills in each occupation (e.g. intelligence, reasoning, language). We complement these data with the occupational score – a synthetic measure of job value developed by IPUMS. For 1930 we also have information on income, which we reconstruct at the occupation level from the Census Bureau's Statistical Abstracts of the U.S. for the 1930s. While these earnings do not capture variation at the *individual* level, they provide an additional dimension to our occupational scoring. Taken together, these variables provide a comprehensive picture of our men before and after the war.

3.3 Data description and balance

Table 1 shows summary statistics for some of the variables we collected. T Panel A describes characteristics of men in our main sample at the time of registration. Out the 512'829 men we selected in Table A.1, we have complete data for 350'031: this constitutes our main sample. These men are 25.3 years of age on average, mostly white, of medium height and physical build. Half of them were married when the war broke out. In 1917 they were employed in occupations requiring relatively low cognitive skills: 1.9 out of 4 in intelligence, 2.4, 3.1 and 2.5 in mathematical, reasoning and language ability (out of 6). Less than 5 percent are employed in positions of authority. The average occupational score is 21 (out of 100).

Panel B reports statistics from variables observed in the 1930 census. A little less than a third of the men are WWI veterans, and only one percent more are veterans of other wars. Thirteen years after the war these men have moved slightly up in the occupational ladder: their average occupational score is 26 and all cognitive scores have improved slightly. Appendix A describes in detail every variable used in the analysis.

Figure 2 reports beta coefficients from regressions explaining observable characteristics of the 1917 recruits with assigned order number. We find no significant effects for height, physique, marital status, or skills. The one variable for which we find an effect is race: white men have slightly lower order numbers. All beta coefficients are precisely estimated and small

(including the white dummy). Appendix Figure A.6 visually displays the randomness of the order number, by plotting each draft number (x-axis) against the order in which it was drawn (y-axis). As one would expect from the outcome of a fair lottery, no systematic pattern emerges and we interpret the significant coefficient for white as the result of chance. Occupational scores vary widely. At the top physicians, surgeons, dentists, lawyers, judges and justices get 60 or more, at the bottom launderers and servant have between 4 and 6. There is also important geographical variation in occupational characteristics. Appendix Figure A.7 gives an overview of average 1930 occupational scores by state. The North-East, the area around the great lakes, and California/Arizona have relatively high scores; the Mid-West and South tend to be home to jobs with lower occupational scores.

Panel A. 1917 characteristics	Min	Mean	St. dev.	Max	Obs.
Age	21.000	25.324	2.841	30.000	350031
White	0.000	0.907	0.290	1.000	350031
Black	0.000	0.031	0.173	1.000	350031
Race missing	0.000	0.059	0.236	1.000	350031
Height: short	0.000	0.048	0.214	1.000	350031
Height: tall	0.000	0.154	0.361	1.000	350031
Physical build: thin	0.000	0.133	0.340	1.000	350031
Physical build: stout	0.000	0.048	0.213	1.000	350031
Eye color: Blue	0.000	0.212	0.409	1.000	350031
Eye color: Brown	0.000	0.162	0.369	1.000	350031
Hair color: Blond	0.000	0.071	0.256	1.000	350031
Hair color: Brown	0.000	0.272	0.445	1.000	350031
Hair color: Black	0.000	0.087	0.282	1.000	350031
Hair color: Red	0.000	0.011	0.106	1.000	350031
Registration: Northeast	0.000	0.221	0.415	1.000	350031
Registration: South	0.000	0.228	0.420	1.000	350031
Registration: Midwest	0.000	0.440	0.496	1.000	350031
Registration: West	0.000	0.111	0.314	1.000	350031
Married in 1917	0.000	0.513	0.500	1.000	350031
Intelligence	1.012	1.900	0.504	3.916	147383
Mathematical ability	1.057	2.380	0.843	5.531	147383
Reasoning ability	1.552	3.155	0.773	5.900	147383
Language proficiency	1.085	2.548	0.794	5.873	147383
Authority or position	0.000	0.047	0.143	0.882	147383
1st principal component	-4.023	-0.002	2.508	10.208	147383
Occupational score	4.009	21.441	11.778	79.998	147383

Panel B. 1930 characteristics	Min	Mean	St. dev.	Max	Obs.
WWI veteran	0.000	0.303	0.460	1.000	350031
Veteran (any war)	0.000	0.313	0.464	1.000	350031
Intelligence	1.012	2.060	0.571	3.916	350031
Mathematical ability	1.057	2.632	0.914	5.531	350031
Reasoning ability	1.552	3.396	0.836	5.900	350031
Language proficiency	1.085	2.793	0.906	5.873	350031
Authority or position	0.000	0.087	0.190	0.882	350031
1st principal component	-4.371	0.015	2.488	8.687	350031
Occupational score	4.009	25.574	12.606	79.998	350031

Table 1. Summary statistics for the main sample of analysis

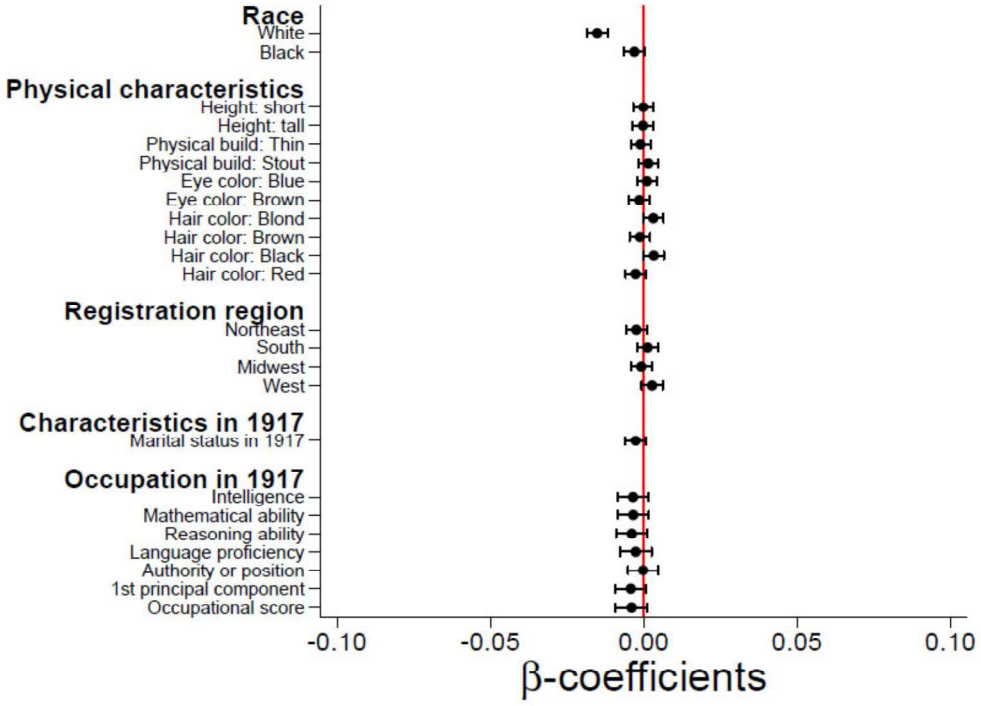


Figure 2: Balance of the treatment

Note: The figure reports the beta coefficient of the assigned order number in separate bivariate regressions. The outcomes of these separate regressions are reported on the left. 95 confidence intervals are calculated with robust standard errors.

We also use the earnings by profession reported in the 1930 census as an indicator of labor market outcomes. These are derived from the US Census Bureau's Statistical Abstracts. While they do not capture earnings variation at the *individual* level, they add an additional dimension to our occupational scoring. Appendix Figure A.8 shows a binscatter of log income against the occupational score. While clearly highly correlated, there is ample additional variation - a simple bivariate regression only captures 25% of the variation in log earnings.

4 Empirical analysis

We are interested in estimating the causal effect of fighting World War I on the earnings of veterans. We first show the strength of our first stage from random assignment. Next, we

examine the impact of exogenously-induced military service on occupational outcomes, including earnings.

4.1 Data exploration and OLS estimates

We first analyze basic patterns in our data. Then, to establish a baseline, we estimate OLS regressions.

As is commonly observed, veterans in general do better on a range of outcomes. Figure 3 gives an overview:

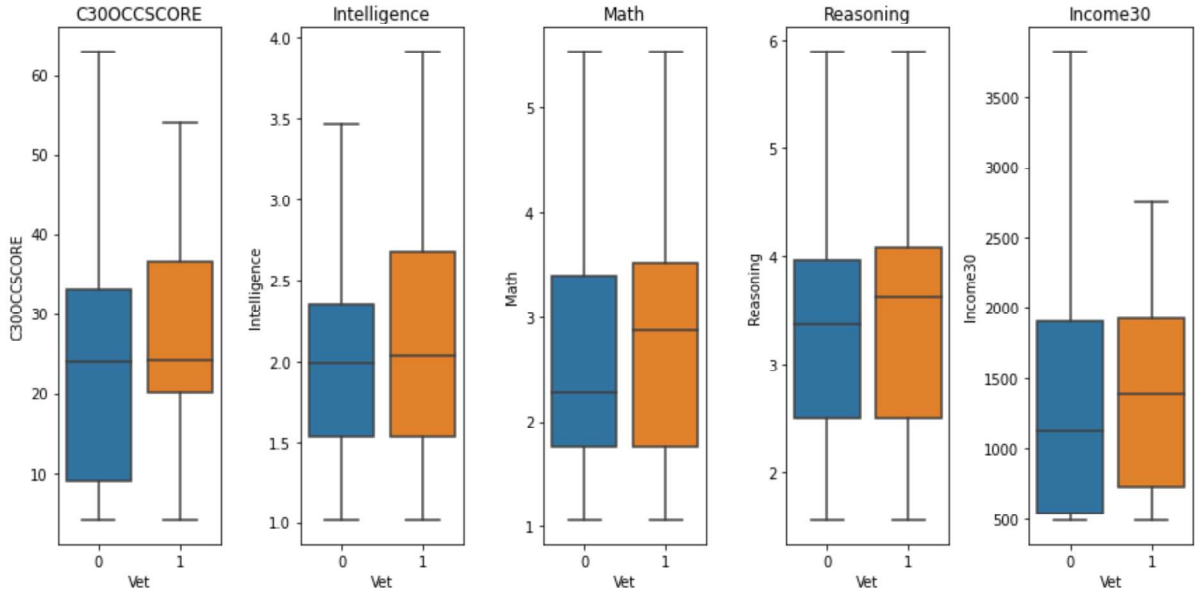


Figure 3: Boxplots of five occupational outcomes, veteran status (yes-no)

Positions filled by veterans have relatively similar median occupational scores, but the mean and upper end are higher. They work in occupations requiring more math, more reasoning, and somewhat higher intelligence. They also earn more by 1930, by 21% (an increase of $279 compared to an average of $1,356). We find very similar results for both younger and older cohorts (Figure A.9). Appendix Figure A.10 shows, based on quantile regressions, that the effect of war service is particularly strong for people in the lower part of the income distribution, where the conditional effect at the 20th to 40th percentile ranges from 20-40%.

4.2 Instrumenting War Participation

The simple OLS estimates are likely to be biased for a number of well-known reasons (Angrist 1990). We use order numbers as drawn in the draft lottery to instrument war participation. On average, higher order numbers spelled a lower likelihood of war participation. Many factors were considered after registration, such as physical fitness and marital status; all of them affected the likelihood of military service. Nonetheless, we can use the lottery number to predict actual war participation. Figure 4 gives an overview for the entire sample. Those with the lowest order numbers had a probability of serving of about 33%. For those with order numbers above 8,000, this fell to about 28%. Note that the overall averages in Figure 4 are pushed up by volunteering, which occurred independent of order number. Having an order number below 4,500 increased the risk of military service by 3.5 percentage points, or almost 12% of the baseline risk.

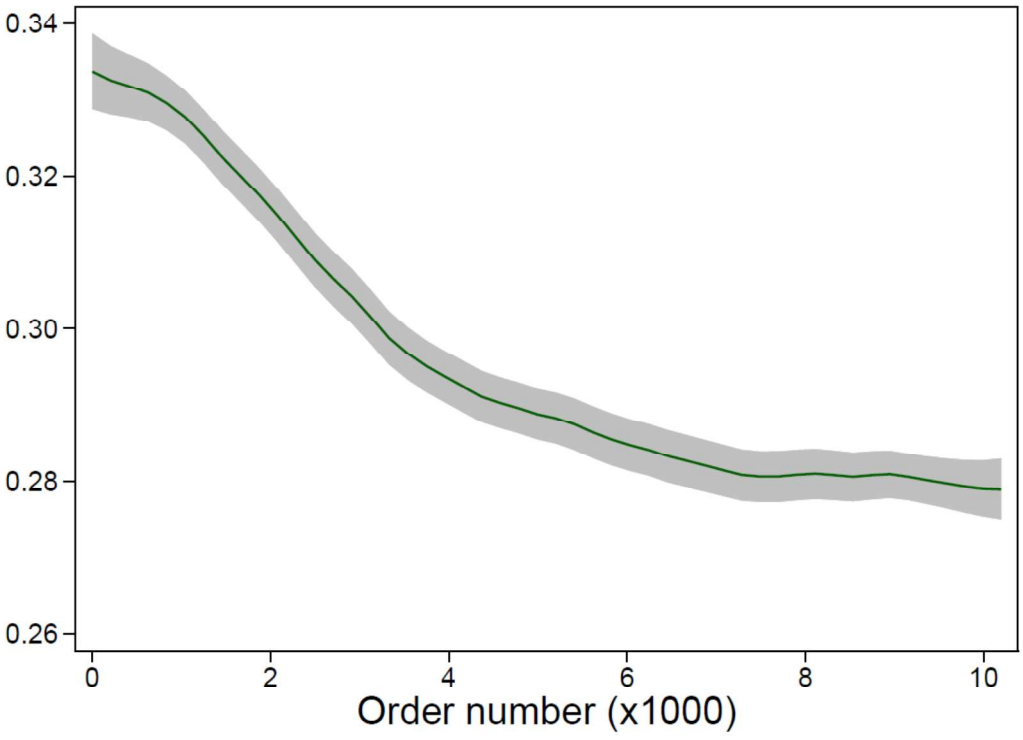


Figure 4: First stage

Note: The figure reports a non-parametric estimation of the relationship between order number in 1917 (x-axis) and probability of being a WWI veteran in 1930 (y-axis). Local polynomial density is estimated with a Epanechnikov kernel and 95 percent confidence intervals are plotted in gray.

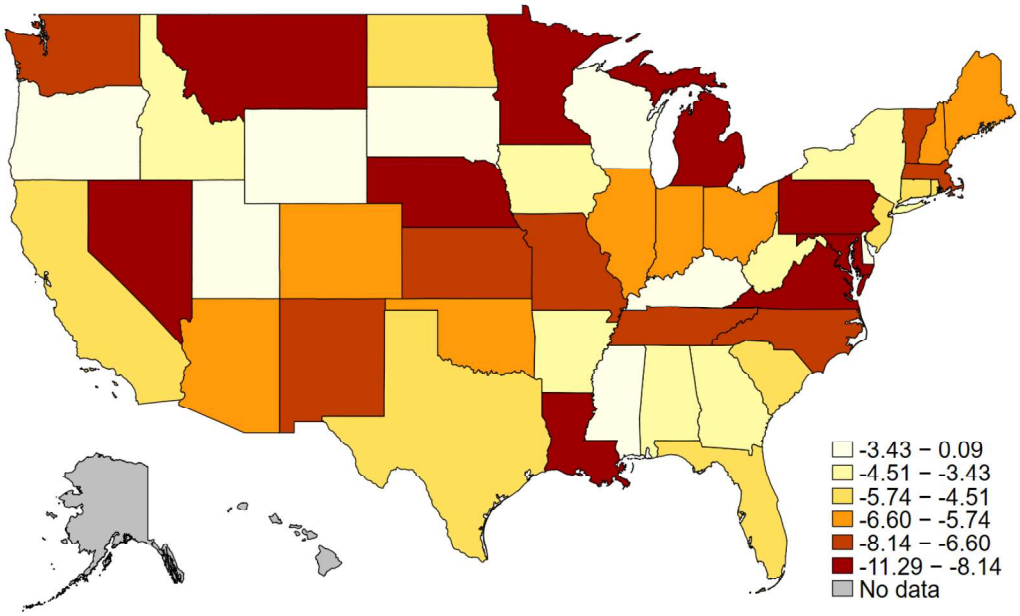


Figure 5: First stage across space

Note: The map reports the point estimates of first stage regressions with all controls but run separately for each of the 48 contiguous states.

Our instrument is strong. We obtain F-statistics of 514 without controls and up to 687 when we control for all 1917 characteristics. The effect of the lottery was not uniform across space. As Figure 5 shows, it has the highest predictive power in New Mexico and the Dakotas as well as Washington; in Florida, Utah, Delaware and Arizona, it was relatively weak. In 7 out of 49 states, the mean estimated effect is negative. The states with low estimated coefficients often had high volunteering rates, making it harder to pin down the effect of the lottery.

Table 2 examines the effect of order numbers on war participation in more detail. Controlling for age barely changes the coefficient. Being married reduced the risk of service substantially, by almost half. Again, it does not affect the magnitude of the lottery number's effect. In fact, Figure A.11 shows that order number significantly predict WWI participation both among unmarried and married men. Adding controls for race leaves the coefficient unaffected, suggesting that the implementation of the 1917 lottery led to clean randomization.

	WWI veteran (0/1)				
	(1)	(2)	(3)	(4)	(5)
Order number	-0.006***	-0.006***	-0.006***	-0.006***	-0.006***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Married before 1917			-0.412***		-0.381***
			[0.002]		[0.002]
Black				-0.020***	-0.015***
				[0.005]	[0.004]
Other race				-0.103***	-0.122***
				[0.016]	[0.015]
Race missing				0.044***	-0.065***
				[0.003]	[0.004]
Age FEs	No	Yes	No	No	Yes
R^2	0.002	0.059	0.202	0.002	0.214
Mean dep. var.	2.632	2.632	2.632	2.632	2.632
F-test of excluded instrument	513.9	601.5	674.9	526.1	686.8
Observations	357152	357152	357152	357152	357152

Table 2: First stage

Note: The table reports coefficients of a regression of WWI veteran status against order number. Standard errors clustered at registration district in parentheses. P-values based on standard errors clustered at registration district level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

4.3 Reduced form

In Figure 6, we explore the reduced form patterns graphically. Higher order numbers meant that a man was less likely to be drafted, all else equal. We see that the lines in terms of occupational outcomes slope downwards for each of the six indicators shown – lower risk of being conscripted was systematically associated with a lower occupational score, jobs that required less intelligence, less math, less reasoning and language ability, and paid less.

How do order numbers correlate with earnings in 1930? Effects may be heterogenous for younger and older war participants. In Figure 7, we show a contour plot of expected income in 1930, as a function of the year of birth and the order number from the lottery. The share of high earnings, indicated by red and yellow, declines with the order number; that of the lowest earnings (blue and magenta) increases. In other words, the greater the risk of war participation, the higher the income measured in 1930, some 11 years after the troops came home. This is

particularly true for the youngest servicemen, whereas for the oldest ones, the increase is more gradual.

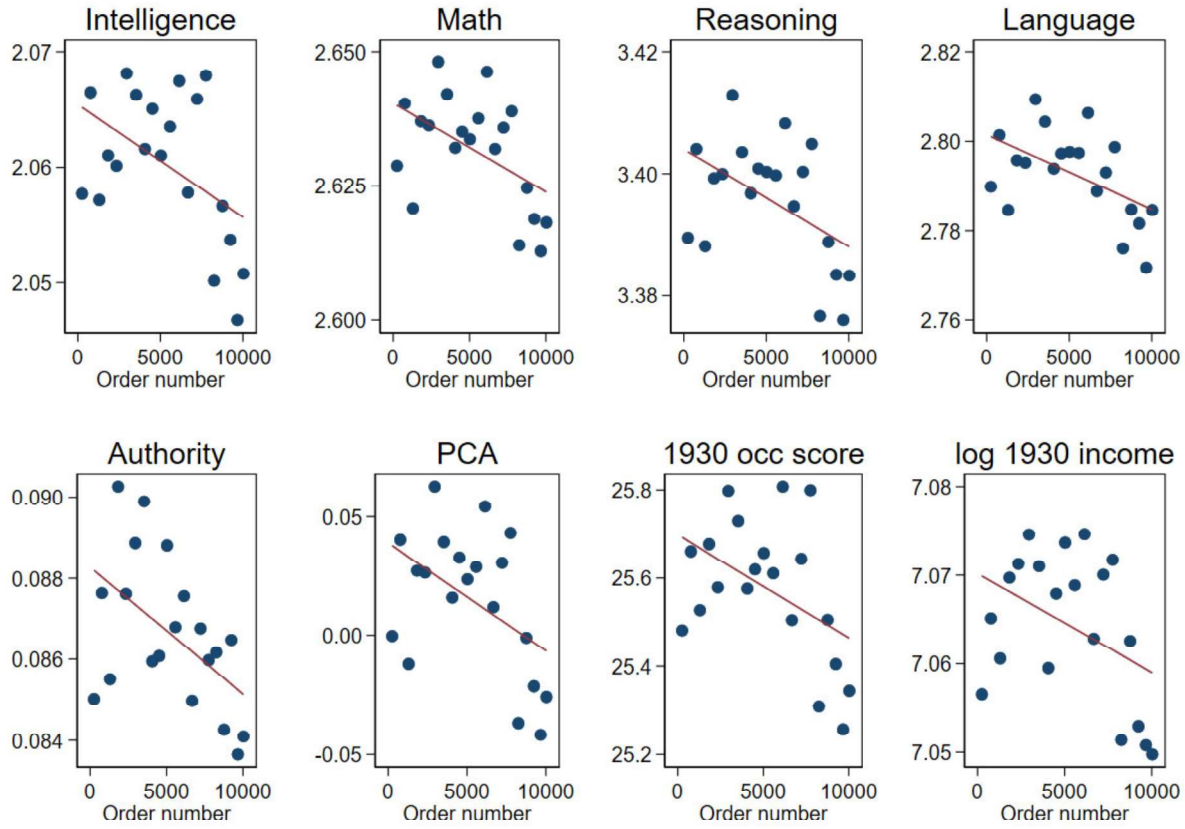


Figure 6: Lottery numbers and occupational outcomes in 1930

In Table 3, we go beyond the visual evidence and estimate reduced form regressions including controls. We examine 8 outcomes in total – the occupational score of men in the 1930 census, their income, and the occupational skill requirements of their job, divided into five categories (intelligence, mathematical ability, reasoning, language skill, and authority).

We find positive, significant effects throughout. The higher the order number, the lower the risk of military service – and the lower the skill requirements of the jobs that a man would by 1930, across all categories except for authority. Overall occupational scores were higher for those at greater risk of being drafted, and their income was higher compared to their peers. These results hold when using age fixed effects, to control for the fact that older men were less likely to serve. When we combine the different occupational characteristics through principal component analysis, we again find a strong and significant effect (col. 6). Similarly, occupational score and

1930 income are significantly lower for men with higher order numbers (cols. 7-8). Finally, Appendix Figure A.12 shows the effect of the lottery number on earnings across different states: while there's considerable variation, in most states men who received smaller order numbers earned more in 1930.

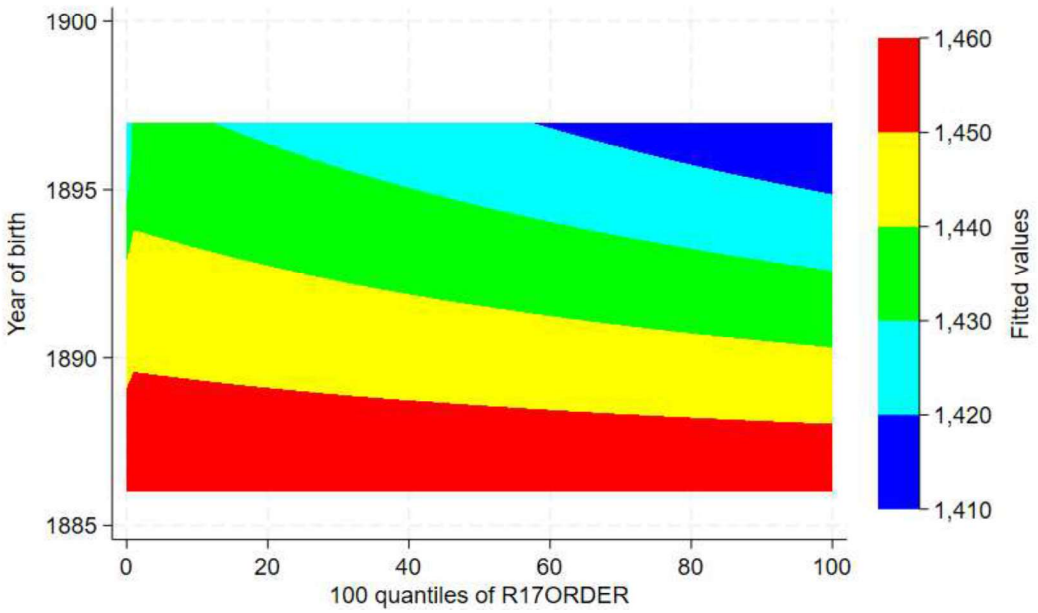


Figure 7: Contourplot, expected earnings as a function of order number and birth cohort

	Intelligence	Math	Reasoning	Language	Authority	PCA	Occ Score	log income
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Order number	-0.005***	-0.005***	-0.006***	-0.005***	-0.005***	-0.005***	-0.005***	-0.005***
	[0.004]	[0.002]	[0.001]	[0.002]	[0.004]	[0.002]	[0.001]	[0.002]
Married before 1917	-0.037***	-0.030***	-0.039***	-0.036***	-0.026***	-0.040***	-0.045***	-0.047***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Black	-0.094***	-0.098***	-0.103***	-0.097***	-0.055***	-0.101***	-0.073***	-0.091***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Other race	-0.023***	-0.022***	-0.024***	-0.024***	-0.012***	-0.023***	-0.015***	-0.020***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Race missing	-0.012***	-0.010***	-0.011***	-0.012***	-0.010***	-0.011***	-0.006***	-0.010***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.002]	[0.000]
Age FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.011	0.011	0.013	0.012	0.004	0.013	0.008	0.011
Mean dep. var.	2.060	2.632	3.396	2.793	0.087	0.015	25.574	7.064
Observations	350031	350031	350031	350031	350031	350031	350031	349145

Table 3: Reduced Form

Note: The table reports beta coefficients from regressions of 1930 occupational outcomes on the order number determined by the 1917 lottery. P-values based on standard errors clustered at registration district level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

4.4 Two-stage least squares

Next, we estimate the causal effect of war participation on occupational outcomes and earnings:

$$W = C + \gamma N_i + \epsilon_0 \quad (1)$$

$$Y = C + \beta W_i + \epsilon_1 \quad (2)$$

where W is a dummy variable for veteran status, C is a constant, N is the lottery number of each registrant i, and Y is the outcome of interest.

Results are summarized in Table 4. In panel A, we find strong and significant effects throughout, except for authority. Beta coefficients are in the range of 10-17% of a standard deviation, with the largest effect on occupational scores. The 2SLS-estimates are similar in magnitude to the OLS ones, with an approximately 25-50% gain in the size of the beta coefficients. The fact that both OLS and 2SLS estimates are positive strongly suggests that positive selection of volunteers was not a dominant feature of the recruiting process.

Panel B shows results with state fixed effects. Coefficients are slightly smaller throughout, and near-identical to OLS coefficients. The occupational characteristics results including individual components remain significant; the coefficient for log income falls below standard levels of statistical significance.

Panel A. Baseline results.

	Intelligence	Math	Reasoning	Language	Authority	PCA	Occ Score	log income
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
WWI veteran	0.124***	0.132***	0.140***	0.134***	0.125***	0.133***	0.138***	0.133***
	[0.004]	[0.002]	[0.001]	[0.002]	[0.004]	[0.002]	[0.001]	[0.002]
Married before 1917	0.015	0.026	0.020	0.020	0.026	0.016	0.013	0.009
	[0.412]	[0.156]	[0.279]	[0.272]	[0.144]	[0.372]	[0.455]	[0.614]
Black	-0.093***	-0.097***	-0.102***	-0.096***	-0.055***	-0.100***	-0.072***	-0.091***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Other race	-0.021***	-0.021***	-0.022***	-0.022***	-0.010***	-0.021***	-0.013***	-0.018***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Race missing	-0.007***	-0.006**	-0.006**	-0.007***	-0.005**	-0.006**	-0.002	-0.005**
	[0.003]	[0.014]	[0.012]	[0.002]	[0.021]	[0.012]	[0.541]	[0.038]
Age FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.023	0.020	0.025	0.023	0.006	0.025	0.020	0.025
Mean dep. var.	2.060	2.632	3.396	2.793	0.087	0.015	25.574	7.064
Rubin-Anderson test (p-value)	0.004	0.002	0.001	0.002	0.004	0.002	0.001	0.002
Observations	350031	350031	350031	350031	350031	350031	350031	349145

Panel B. State fixed effects.

	Intelligence	Math	Reasoning	Language	Authority	PCA	Occ Score	log income
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
WWI veteran	0.085**	0.089**	0.093**	0.098**	0.110**	0.086**	0.087**	0.082**
	[0.042]	[0.032]	[0.023]	[0.018]	[0.010]	[0.036]	[0.033]	[0.042]
Married before 1917	0.004	0.012	0.005	0.009	0.022	0.002	-0.003	-0.007
	[0.837]	[0.486]	[0.765]	[0.596]	[0.218]	[0.925]	[0.873]	[0.688]
Black	-0.083***	-0.084***	-0.089***	-0.088***	-0.052***	-0.086***	-0.055***	-0.074***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Other race	-0.020***	-0.020***	-0.021***	-0.022***	-0.010***	-0.020***	-0.012***	-0.016***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Race missing	-0.008***	-0.007***	-0.008***	-0.008***	-0.006**	-0.008***	-0.003	-0.007***
	[0.000]	[0.002]	[0.001]	[0.000]	[0.016]	[0.001]	[0.174]	[0.004]
Age FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.053	0.056	0.067	0.048	0.012	0.066	0.067	0.072
Mean dep. var.	2.060	2.632	3.396	2.793	0.087	0.015	25.574	7.064
Rubin-Anderson test (p-value)	0.043	0.033	0.024	0.019	0.010	0.037	0.034	0.043
Observations	350031	350031	350031	350031	350031	350031	350031	349145

Table 4. Two stages least squares

Note: The table reports beta coefficients from regressions of 1930 occupational outcomes on WWI veteran status instrumented with the order number determined by the 1917 lottery. Panel A reports baseline results and Panel B reports results with state fixed effects. P-values based on standard errors clustered at registration district level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

4.5 Coarsened exact matching results

Establishing the effect of military service assumes that the treated and untreated are comparable. While this is a compelling assumption for the 2SLS strategy, the OLS results may arguably be affected by differential selection into treatment. To compare like with like, we use coarsened exact matching (CEM, Iacus et al., 2012)) – forming comparison groups of highly similar individuals (based on 1917 observables), with military service the only difference. To ensure that comparison groups are sufficiently large, the estimator forms broader groups.

In table 5, we use exact matching on marital status, race, age class, 9 geographic areas based on the Army's Service Commands, as well as quartiles for all 1917 occupational variables (intelligence, math, reasoning, language, authority and occupational score). Panel A shows that the simple OLS coefficients are essentially identical to the baseline when we reduce the sample with CEM. We lose more than half of the observations because there is no match for treated individuals. The fact that results are not affected when we force the sample to only contain treated and untreated individuals that are similar suggests that imbalances in the two populations are unlikely to drive results.

For completeness, we also report reduced-form and two stages least squares results estimated on CEM samples. For these exercises we define the treatment as having an order number below 4500 (an almost fool-proof indicator that a man would at least be called for Army examination). Panel B reports reduced form results and Panel C two stages least squares. As for OLS, results are unaffected although the samples are drastically reduced.

In the appendix (Table A.2), we show that an even more demanding specification of our CEM approach in which we only keep treated and untreated individuals that share also the exact age and state yields larger coefficients; the more comparable the samples of treated and untreated, the bigger the effects.

Panel A: OLS regressions

	Intelligence	Math	Reasoning	Language	Authority	PCA	Occ Score	log income
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
WWI veteran	0.123***	0.109***	0.125***	0.118***	0.070***	0.124***	0.129***	0.134***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Married before 1917	0.027***	0.026***	0.023***	0.028***	0.011***	0.023***	0.014***	0.014***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.001]	[0.000]	[0.000]	[0.000]
Black	-0.098***	-0.100***	-0.108***	-0.101***	-0.056***	-0.105***	-0.077***	-0.096***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Other race	-0.020***	-0.019***	-0.021***	-0.019***	-0.008***	-0.020***	-0.016***	-0.019***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Race missing	-0.010***	-0.009***	-0.009***	-0.009***	-0.004	-0.009***	-0.003	-0.006**
	[0.000]	[0.001]	[0.001]	[0.001]	[0.139]	[0.001]	[0.206]	[0.039]
Age FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.024	0.021	0.026	0.023	0.008	0.025	0.021	0.026
Mean dep. var.	2.029	2.590	3.348	2.747	0.083	-0.121	24.761	7.025
Observations	141713	141713	141713	141713	141713	141713	141713	141444

Panel B: Reduced form

	Intelligence	Math	Reasoning	Language	Authority	PCA	Occ Score	log income
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Order number	-0.007***	-0.007***	-0.007**	-0.006**	-0.004	-0.007***	-0.008***	-0.004
	[0.005]	[0.010]	[0.011]	[0.014]	[0.146]	[0.010]	[0.003]	[0.130]
Married before 1917	-0.022***	-0.016***	-0.025***	-0.018***	-0.019***	-0.024***	-0.034***	-0.036***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Black	-0.099***	-0.100***	-0.109***	-0.102***	-0.057***	-0.106***	-0.079***	-0.098***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Other race	-0.021***	-0.021***	-0.023***	-0.022***	-0.009***	-0.023***	-0.016***	-0.021***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Race missing	-0.011***	-0.010***	-0.010***	-0.010***	-0.004	-0.010***	-0.004	-0.007**
	[0.000]	[0.000]	[0.000]	[0.000]	[0.110]	[0.000]	[0.110]	[0.011]
Age FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.012	0.012	0.014	0.013	0.004	0.014	0.008	0.012
Mean dep. var.	2.031	2.593	3.352	2.749	0.083	-0.109	24.838	7.029
Observations	145226	145226	145226	145226	145226	145226	145226	144942

Panel C: 2 SLS

	Intelligence	Math	Reasoning	Language	Authority	PCA	Occ Score	log income
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
WWI veteran	0.183***	0.170***	0.166**	0.160**	0.095	0.169***	0.194***	0.098
	[0.005]	[0.010]	[0.011]	[0.014]	[0.147]	[0.009]	[0.003]	[0.129]
Married before 1917	0.057**	0.056**	0.046*	0.050*	0.022	0.048*	0.049*	0.006
	[0.044]	[0.046]	[0.097]	[0.076]	[0.436]	[0.088]	[0.084]	[0.826]
Black	-0.098***	-0.099***	-0.108***	-0.101***	-0.056***	-0.105***	-0.078***	-0.097***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Other race	-0.019***	-0.019***	-0.021***	-0.020***	-0.008***	-0.021***	-0.014***	-0.020***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Race missing	-0.009***	-0.008***	-0.009***	-0.009***	-0.004	-0.009***	-0.003	-0.006**
	[0.001]	[0.002]	[0.002]	[0.001]	[0.188]	[0.002]	[0.291]	[0.025]
Age FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.021	0.018	0.024	0.022	0.007	0.024	0.017	0.024
Mean dep. var.	2.031	2.593	3.352	2.749	0.083	-0.109	24.838	7.029
Observations	145226	145226	145226	145226	145226	145226	145226	144942

Table 5. Coarsened Exact Match samples

Note: The table reports beta coefficients from regressions run on reduced sample comprising only treated and control men who have the same observable characteristics. Panel A: OLS regressions: we define treatment as being

WWI veteran. Panel B: reduced form regressions: we define treatment as having an order number below 4500. Panel C: two stages least squares regressions: we define treatment as having an order number below 4500.

4.5 Feature importance

Many factors influenced earnings and occupational standing in 1930. One simple way to assess the importance of military service is to use machine learning methods to predict income in 1930, using the universe of observable variables in 1917 plus an individual's draft number as explanatory variables. Machine learning methods like Random Forest Estimation will then combine subsets of variables in often non-linear ways to generate the best prediction. While hard to interpret in its own right, ML predictions have the advantage of establishing a "frontier" for predictive performance.

Here, we look at variable importance, equivalent to the reduction in predictive performance (RMSE) if a variable is not used. Figure 8 plots the importance of each variable available, including the occupational characteristics of the professions from which recruits came in 1917. Unsurprisingly, factors like age and reasoning requirements are important in predicting earnings in 1930. The single most influential variable, however, is an individual's order number – it has more predictive power than the next two variables combined. The vast majority of job characteristics in 1917, in contrast, have little or no predictive power for 1930 earnings.

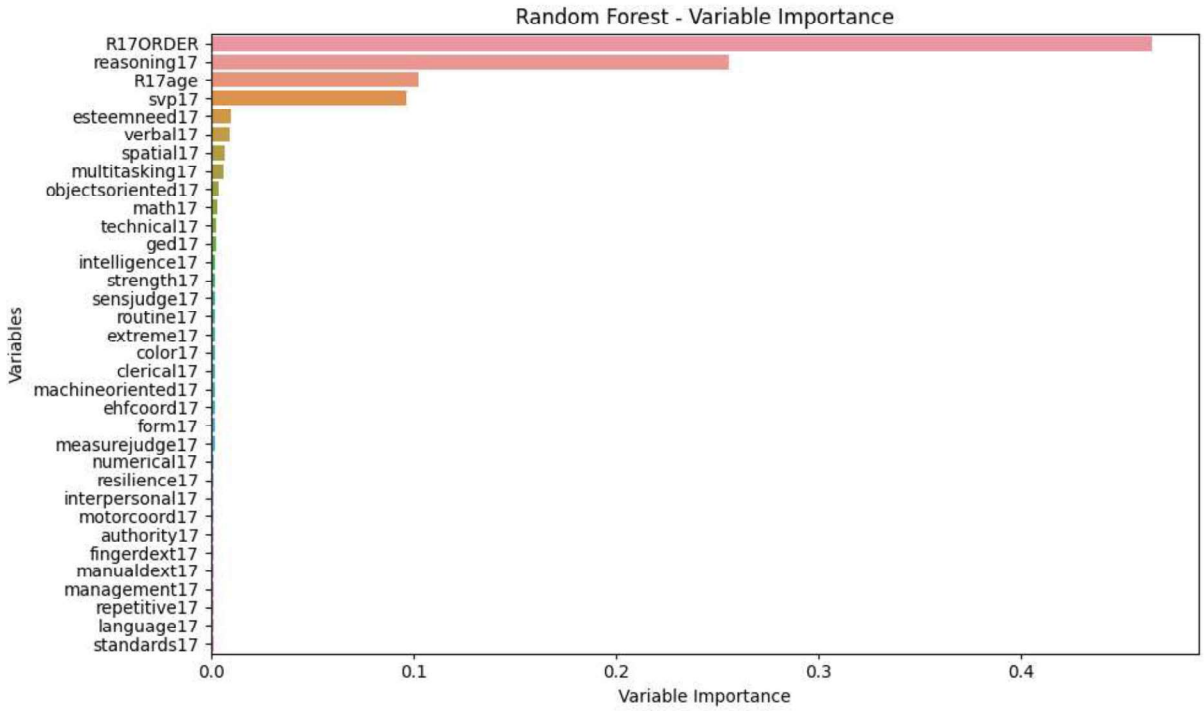


Figure 8: Variable Importance - Random Forest Prediction Exercise

Note: We predict 1930 income using a large set of observable characteristics of potential draftees in 1917. Importance is measured by the reduction in the root mean square error of out of sample predictions when a variable is included in a model. Occupational characteristics such as reasoning requirements and 1917 age have predictive power for 1930 earnings – but none has as high an influence as the (randomly assigned) 1917 order number.

5 Robustness

Our results are robust to several exercises.

First, spatial autocorrelation across registration districts is unlikely to bias standard errors downwards. Our main source of variation are order numbers which by construction have no spatial pattern (in fact, each order number may appear in each registration district). We confirm the robustness of our results to spatial inference in Appendix Table A.3, where we report standard errors computed with the method of Conley (1999), which allows disturbances of places to be correlated until a given threshold. We experiment with several threshold and find that results remain significant throughout.

Second, outliers and influential observations are not affecting our results. All our variables are bounded and have relatively little skew: WWI veteran status is a binary indicator,

order number is distributed uniformly between 1 and 10,000, and occupational outcomes have few extreme values. Thus outliers and influential observations are unlikely to play a major role in our results. Appendix Table A.4 confirms this intuition: it reports robust regression results that downweight influential observations (Li 1985); it shows that neither coefficients nor significance are affected by the weighting.

Third, our results are not driven by specific geographic areas. Figure 7 showed that the strength of the first stage varies across US states, which may suggest that our conclusions depend on a few individual states. Appendix Figures A.13–A.15 investigate the strength of our results to re-estimating our models after dropping one state at a time. The figures plot beta coefficients for the OLS (Figure A.13), the reduced form (Figure A.14) and two stages least squares (Figure A.15): each panel reports results from regressions with a different dependent variable. Results are remarkably stable throughout.

We use a treatment variable aimed at capturing precisely exposure to WWI in the 1930 Census. We classify men as WWI veterans if two conditions apply: first, they answer "yes" to Question 30 "whether a veteran of US military or naval forces;" and second they answer "WWI" to Question 31 "What war or expedition?" The variable likely minimizes false positives. However, errors in transcriptions or digitization mean that we may be missing some WWI veterans that gave non-standard answers to the second question (or whose answer was recorded in non-standard ways). Appendix Table A.6 reproduces the IV results with a more liberal measure of veteran status, that considers everyone who declared to be a veteran, regardless of its war or expedition. Results hold with this different measure of veteran status.

6 Channels

Did all conscripts benefit equally from military service during WWI? Here, we examine the effect of military service due to the lottery for subgroups.

We divide our conscripts into terciles by age, and examine their occupational scores in 1930 as a function of order numbers. Figure 9, Panel A, shows a progressive steepening of the relationship between occupation scores and order number by age. Older men (aged 28-30) most likely to be drafted saw markedly better outcomes than those with high numbers. There is still a downward slope for men of the middle tercile (aged 25-27), but for the youngest in our sample (aged 21-24), the curve is flat. A similar pattern is visible for earnings, with a steeper gradient for

the oldest draftees, a flattening of the curve for the mid-twenties recruits, and a flat or positive slope for the youngest (indicating that military service diminished their earnings). Statistical results from regressions including all controls are in Table A.6 in the Appendix.

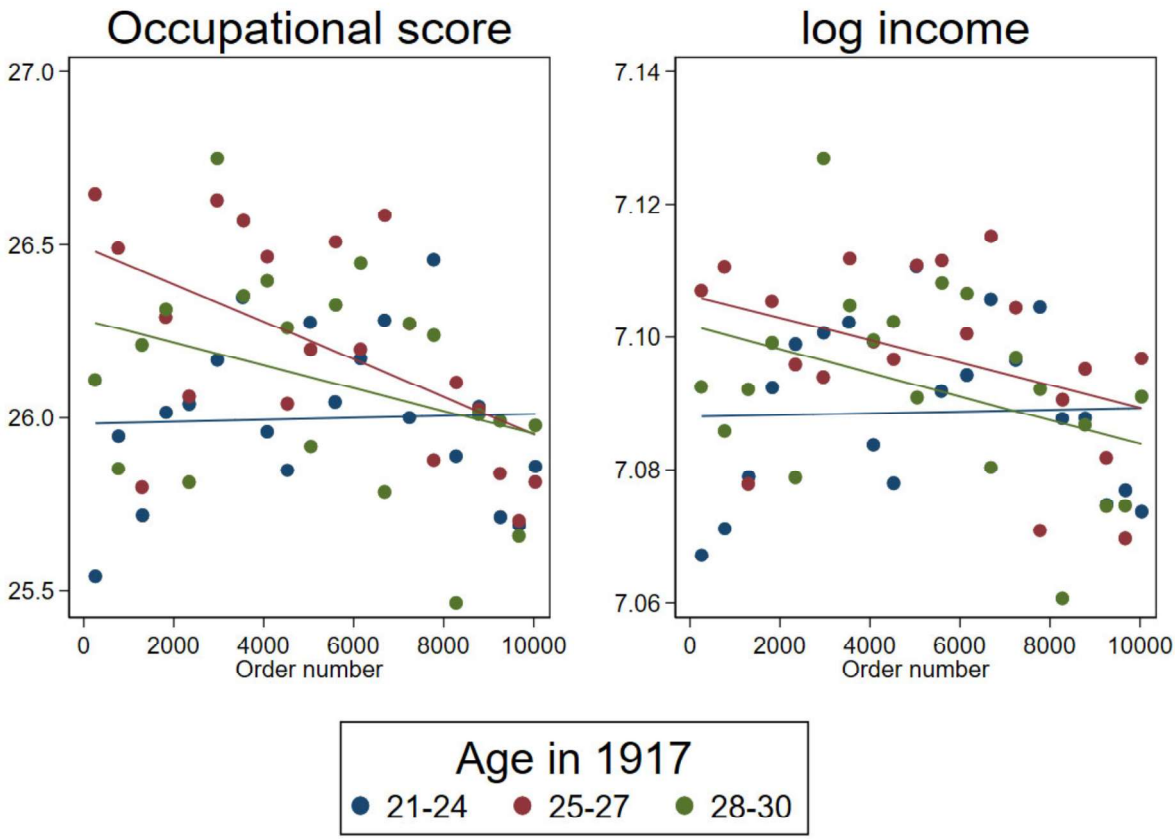


Figure 9: Occupational Scores, Log 1930 Earnings, and Lottery Numbers by age tercile

The US army in 1917 was short of officers, NCOs, and men. The army quickly realized that the more experienced and older men could potentially act as leaders. Accordingly, their chances of being trained to lead men into battle, or to receive additional training for headquarter tasks, were probably higher. While we do not have the service records of the inductees, ample anecdotal evidence suggests that conscription and subsequent military service opened a window of opportunity for some – and especially those with slightly more experience.

7 Conclusions

The "credibility revolution" in empirical economics has made important contributions to the understanding of key questions ranging from the effects of immigration to the impact of global trade and beyond. As a demonstration of a technique's potential, few studies have been more influential than Angrist's hallmark study of veteran status as a determinant of income, pioneering the use of instrumental variables. Criticisms of instrumental variable methods often emphasize the limited nature of plausible variation that is pinned down by IV.

In this study, we re-visit the effect of veteran status on earnings, using another draft lottery in the US – the one conducted during World War I. While very similar in nature and motivation, and affecting millions of draftees, it occurred under very different circumstances – WWI was popular at home and there was no opposition, nor were returning veterans spat at and insulted as "baby killers". The war was also victorious, and returning soldiers stepped into a labor market that was buoyant.

Our results are the exact opposite of Angrist's. Where he found declining earnings of returning veterans, we show that those drafted during WWI ended up earning *more,* by up to 13%. They also worked in more attractive jobs, requiring greater skills. This is true for a range of job characteristics, from language ability to authority and intellectual requirements as well as mathematical skills.

The radically different conclusion of our exercise compared with Angrist's may reflect several factors. Returning as heroes from a victorious war may have facilitated veterans' careers after 1918; army training in the 1910s may have been more useful in civilian work than it was in the 1970s. Our results establish suggest that clean identification alone is insufficient to establish the sign of the effect of one particular treatment on outcomes, beyond the individual case in which it was derived. In a different setting, the exact opposite pattern may emerge, with equally good identification (Deaton 2010). This suggests that "historicist" views of the progress of knowledge in economics have a good deal of veracity on their side – as scholars have discovered in related fields like development as well.[3]

[3]Duflo's (2017) famous metaphor of the economist as "plumber" is equally cognizant of the fact that even well-identified results may not scale or apply to very similar contexts elsewhere.

References

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. "Europe's tired, poor, huddled masses: Self-selection and economic outcomes in the age of mass migration." *American Economic Review* 102, no. 5 (2012): 1832-1856.

Abramitzky, Ran, Leah Boustan, Katherine Eriksson, James Feigenbaum, and Santiago Pérez. "Automated linking of historical data." *Journal of Economic Literature* 59, no. 3 (2021): 865-918.

Altoona Tribune. 21st July 1917. Draft Numbers That Will Call Ten Million Men to Build New American Armies for Big War. Altoona, PA: 2 and 6.

American Battle and Monuments Commission. 1938. American Armies and Battlefields in Europe. Government Printing Office. Washington, DC.

Ancestry.com. 2005. U.S., World War I Draft Registration Cards, 1917-1918 [database on-line]. Provo, UT, USA: Ancestry.com Operations Inc.

Angrist, Joshua D. 1990. "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records." American Economic Review, 80(3): 313-336.

Ayres, Leonard P. 1919. The War with Germany. A Statistical Summary. Government Printing Office. Washington, DC.

Crowder, Enoch H. 1918. Report of the Marshal General to the Secretary of War on the First Draft under the Selective Service Act, 1917. Government Printing Office. Washington, DC.

Crowder, Enoch H. 1919. Second Report of the Marshal General to the Secretary of War on the Operations of the Selective Service System to December 20, 1918. Government Printing Office. Washington, DC.

Crowder, Enoch H. 1920. Final Report of the Marshal General to the Secretary of War on the Operations of the Selective Service System to July 15, 1919. Government Printing Office. Washington, DC.

Deaton, A. 1997. The analysis of household surveys. A microeconometric approach to development policy. World Bank Publications. Washington, DC.

Deaton, A. 2010. "Instruments, Randomization, and Learning about Development." Journal of Economic Literature 48, pp. 424-455.

DeBruyne, Nese F., and Anne Leland. 2020. "American war and military operations casualties: Lists and statistics." Library of Congress, Congressional Research Service.

Duflo, E., 2017. “The economist as plumber.” American Economic Review, 107(5), pp.1-26.

Feigenbaum, J. 2015. “Automated Census Record Linking.” Working paper. University of Harvard.

Grotelueschen, Mark E. 2007. The AEF Way of War. The American Army and Combat in World War I. Cambridge University Press. New York, NY.

New York Times. 21st July 1917a. Draft Drawing Plan is Changed Owing to Error. New York, NY.

New York Times. 21st July 1917b. How to Find Immediately Your Place in the Draft. New York, NY: 6-7 and 9.

Li, G. 1985. Robust regression. In *Exploring Data Tables, Trends, and Shapes*, ed. D. C. Hoaglin, C. F. Mosteller,and J. W. Tukey, 281–340. New York: Wiley.

Los Angeles Evening Herald. 21st July 1917. Here are complete draft numbers in order of drawing. Los Angeles, CA: 21.

Selective Service Act. 18 May 1917.

Tan, H. R. (2020). “Did military service during World War I affect the economic status of American veterans?” *Explorations in Economic History*, 75.

Appendix

A Variable Description

Registrants: number of men who registered for the draft lottery on the 6th of June 1917. The primary source for this data is the "1st Report of the Provost Marshal to the Secretary of War" (Crowder, 1918).

Volunteers: number of men who voluntarily enlisted in the Army between the declaration of war (4th of April 1917) and the date of the lottery (6th of June 1917). The information is sourced from the "1st Report of the Provost Marshal to the Secretary of War" (Crowder, 1918), where it is recorded for each registration board as “registration credit” (registration boards were assigned credits based on how many volunteers had come forward before the lottery; credits were used to determine the number of men to induct through the lottery).

Region: location where men registered during the draft lottery. The categories are North-East, South, Mid-West, and West.

Service Command: nine geographic units assigned to separate Army organization units.

Order Number: we construct this variable with information from NARA registration cards and the outcome of the lottery as reported by three newspapers the day after the lottery (Aaltona Tribune, Los Angeles Herald, and New York Times). We first digitize the lottery number from the original draft card taken during registration (Appendix Figure A.1). Next we determine the order in which these draft numbers were extracted during the lottery using the newspapers (Appendix Figure A.5). Hence, the order number determine the order in which each registrant is called at the registration board.

Full Name: Extracted from the NARA registration cards filled out during the lottery registration, this variable provides the complete name of the registrant. We observe the same variable in the 1930 Federal Census and use it for linking records (see Appendix B).

State of Birth: This variable records the place of birth: either a US state or a foreign state. The primary source is the NARA registration cards. We observe the same variable in the 1930 Federal Census and use it for linking records (see Appendix B).

Age: Derived from the NARA registration cards, this variable shows the age of the registrants at the the time of the draft lottery. We observe the same variable in the 1930 Federal Census and use it for linking records (see Appendix B).

Race: The information comes from the NARA registration cards collected during the lottery registration. It classifies men into white, Black, American Indian, Chinese, Japanese, and Mexican or Latino. White and black make up most of the linked records.

Height: This variable categorizes the registrants based on their height: short, medium, or tall. It was asked at the time of the lottery registration and recorded on the NARA registration cards.

Physical Build: This variable categorizes the registrants based on their height: thin, medium, or stout. It was asked at the time of the lottery registration and recorded on the NARA registration cards.

Eye Color: This variable categorizes the registrants based on their eye color: blue, brown, gray or green (0.21% was green-eyed). It was asked at the time of the lottery registration and recorded on the NARA registration cards.

Hair Color: This variable categorizes the registrants based on their hair color: blond, grey, white, brown, black, red. It was asked at the time of the lottery registration and recorded on the NARA registration cards.

Married in 1917: This binary variable indicates whether the registrant was married in 1917. The data originates from the NARA registration cards.

WWI Veteran Status: This variable identifies whether the registrant was a WWI veteran, as recorded in the 1930 Federal Census.

1930 Occupation: The occupational field of the registrant in the year 1930 is sourced from the Federal Census of the same year.

Skill requirement: intelligence. Skill variables are constructed based on the 1930 occupation, and they specify the skills demanded by a specific job role. The 1930 occupations were matched with those in the National Academy of Sciences 1970 and ICPSR 2006 to extract the occupational scores of the Dictionary of Occupational Titles. Intelligence, measured as "aptitude" required by a specific job, has a scale from 1 to 4, with higher being better.

Skill requirement: math. Skill variables are constructed based on the 1930 occupation, and they specify the skills demanded by a specific job role. The 1930 occupations were matched with those in the National Academy of Sciences 1970 and ICPSR 2006 to extract the occupational scores of the Dictionary of Occupational Titles. Math is measured using the GED scores, with a scale ranging from 1 to 6, where higher scores denote better performance.

Skill requirement: reasoning. Skill variables are constructed based on the 1930 occupation, and they specify the skills demanded by a specific job role. The 1930 occupations were matched with those in the National Academy of Sciences 1970 and ICPSR 2006 to extract the occupational scores of the Dictionary of Occupational Titles. Reasoning is measured using the GED scores, with a scale ranging from 1 to 6, where higher scores denote better performance.

Skill requirement: language. Skill variables are constructed based on the 1930 occupation, and they specify the skills demanded by a specific job role. The 1930 occupations were matched with those in the National Academy of Sciences 1970 and ICPSR 2006 to extract the occupational scores of the Dictionary of Occupational Titles. Language is measured using the GED scores, with a scale ranging from 1 to 6, where higher scores denote better performance.

Skill requirement: authority. Skill variables are constructed based on the 1930 occupation, and they specify the skills demanded by a specific job role. The 1930 occupations were matched with those in the National Academy of Sciences 1970 and ICPSR 2006 to extract the occupational scores of the Dictionary of Occupational Titles. Authority, a "temperament",

measures the importance that influencing people has on the job. It is measured on a scale from 0 to 1, where higher indicates higher importance.

Skill requirement: principal component. This is the first principal component of the skill measures.

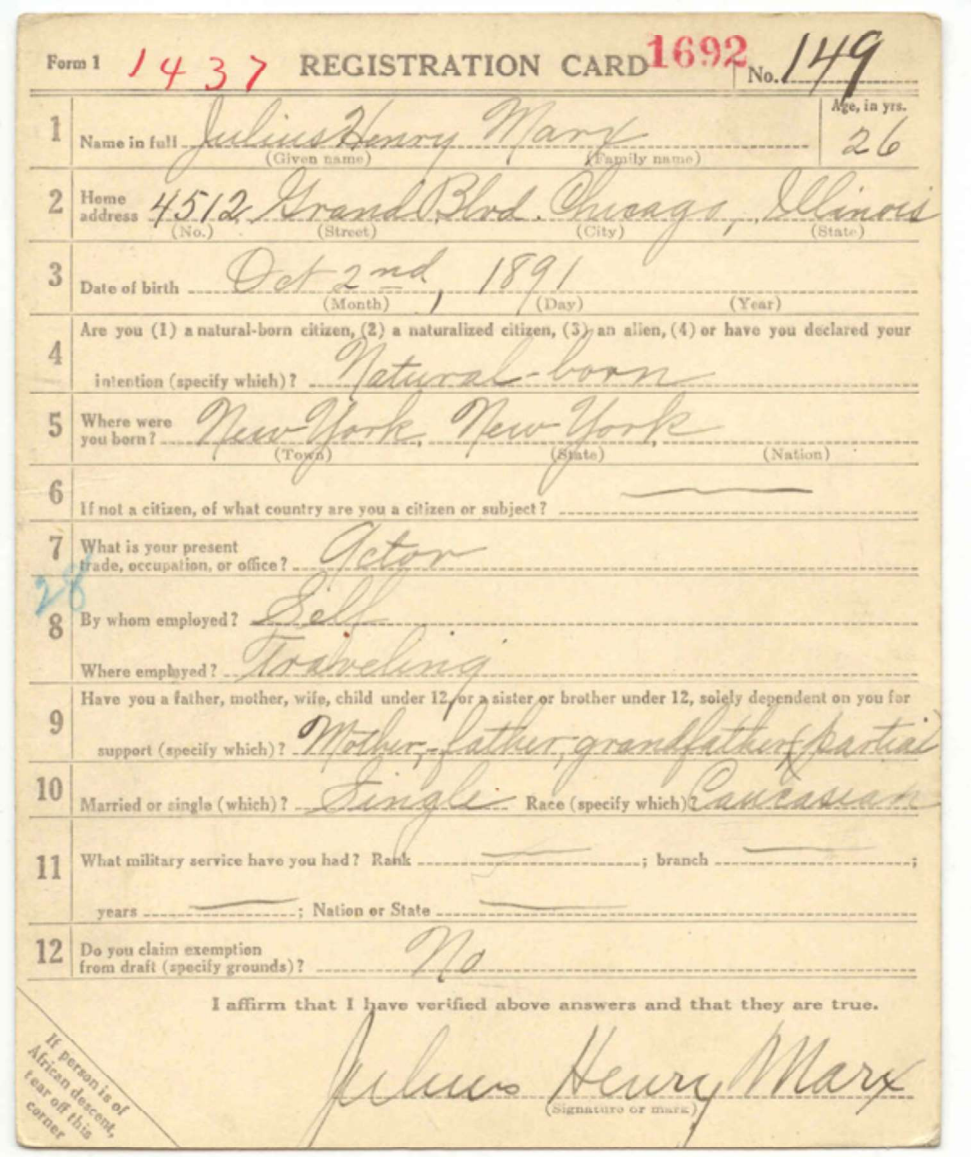
Occupational Score: Created by IPUMS, this score (ranging from 0 to 100) provides information on the socio-economic level of a particular occupation. The variable is constructed based on the 1930 occupation data.

Income 1930: Income information for the year 1930 has been derived based on the occupation of the registrant. This data varies at the occupation level (not at the individual level). The primary source for this information is the Census Bureau's Statistical Abstracts of the U.S. for the 1930s.

B Record Linkage

We reconstruct occupational outcomes of men who took part in the 1917 draft lottery by matching them to their selves in the 1930 US Population Census. We link records with Abramitzky, Boustan and Eriksson (2012) algorithm. More specifically, we focus on the males in the 1930 Census (as all lottery registrants are male). We extract from both sources the following variables: first and last name, year of birth and place of birth (US state or foreign country). Next, we standardize names in both cards and Census, first converting common male nicknames into their corresponding name (e.g. "Ben" for "Benjamin") and then considering the NYSIIS phonetic equivalent of each name (so that names spelled differently but pronounced similarly are considered the same name – e.g. "Voth" and "Vogt"). These standardizations are an effective way to deal with errors in reporting, transcription and digitization (Abramitzky*et al.* 2021). We then search exact matches of the standardized first and last names in the two databases, forcing matches to be born in the same US state or foreign country, and by allowing matches to be born in within 5 years from each other. The result is a many-to-many matches (with many individual left unmatched in both databases). In order to minimize false matches and reduce noise, we add two conservative conditions to identify valid matches. First, we drop anyone who within either database has someone else carrying the same first and last name and is born within a 5-year window. Second, we drop anyone who has more than one match. While these conditions reduce the size of the matched sample considerably, they allow us to concentrate on high quality matches, with high probability of being correct.

C Appendix Figures and Tables



Form 1 1437 REGISTRATION CARD 1692 No. 149

1 Name in full Julius Henry Marx (Given name) (Family name) Age, in yrs. 26

2 Home address 4512 Grand Blvd. (No.) (Street) Chicago (City) Illinois (State)

3 Date of birth Oct 2nd 1891 (Month) (Day) (Year)

4 Are you (1) a natural-born citizen, (2) a naturalized citizen, (3) an alien, (4) or have you declared your intention (specify which)? Natural-born

5 Where were you born? New York (Town) New York (State) (Nation)

6 If not a citizen, of what country are you a citizen or subject? —

7 What is your present trade, occupation, or office? Actor

8 By whom employed? Self

Where employed? Footlights

9 Have you a father, mother, wife, child under 12, or a sister or brother under 12, solely dependent on you for support (specify which)? Mother, father, grandfather, parents

10 Married or single (which)? Single Race (specify which) Caucasian

11 What military service have you had? Rank —; branch —; years —; Nation or State —

12 Do you claim exemption from draft (specify grounds)? No

I affirm that I have verified above answers and that they are true.

Julius Henry Marx (Signature or mark)

If person is of African descent, tear off this corner

Figure A.1: Example of registration card

Note: The card records information of Julius Henry ("Groucho") Marx. The lottery number is recorded in red on the top-left ("1437"). We digitize lottery numbers and demographic information from a random sample of men who registered during the 1917 lottery.

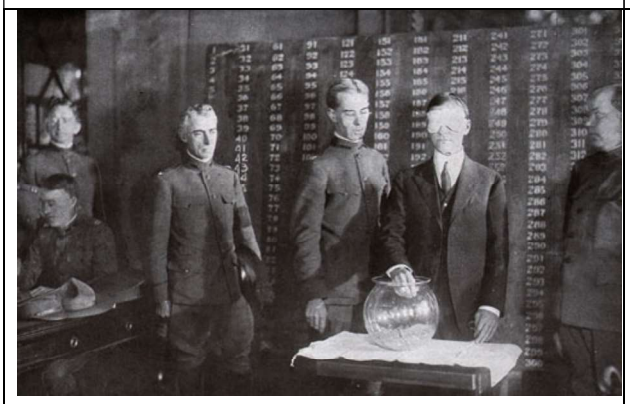
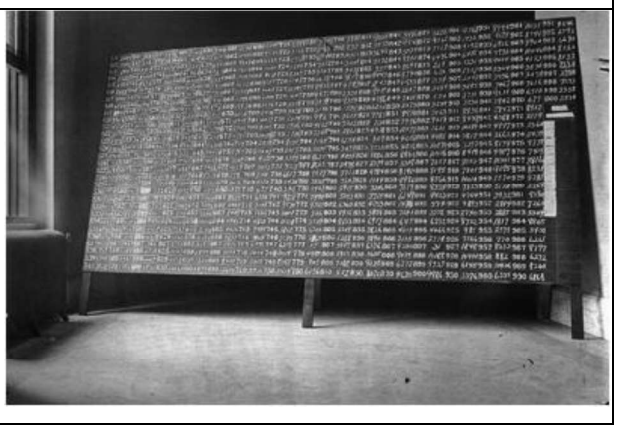
Panel A	Panel B
	

Figure A.2: The draft

Note: Panel A: U.S. Secretary of War Newton D. Baker, blindfolded draws the first lottery number in Washington D.C. on the 20th of July 1917. Panel B: the blackboard where the order of the 10'000 lottery number drawn was displayed.

Figure A.3. Number of battle deaths as percent of men engaged

Note: Own calculations based on Ayres (1919: p. 105 and 121).

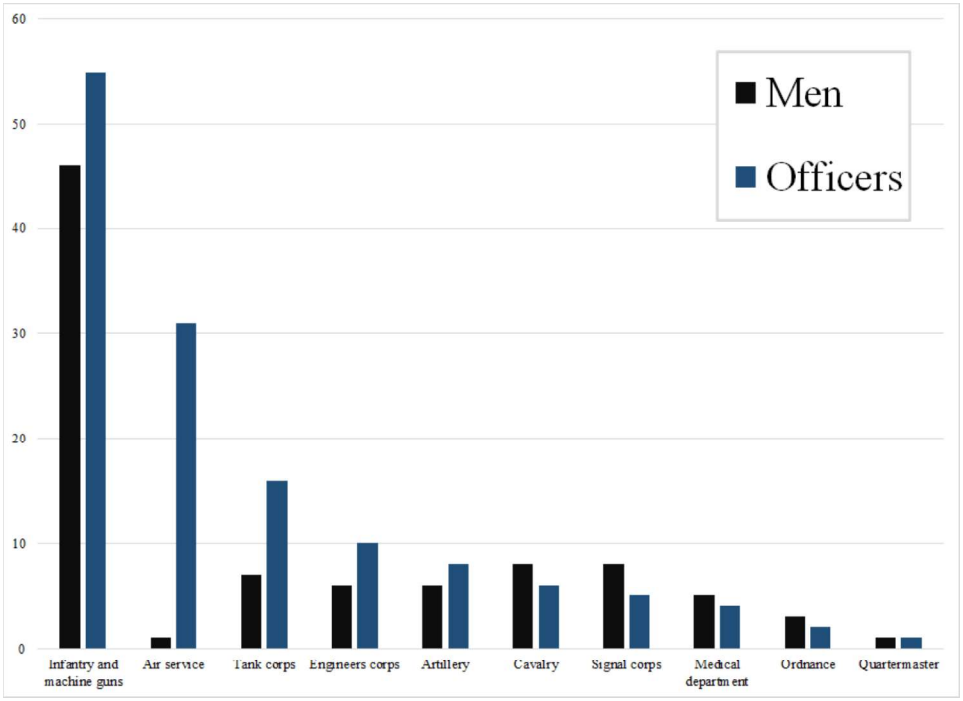


Figure A.4: Number of battle deaths per thousand men

Note: The figure reports deaths per thousand men by branch of service. Source: Ayres (1919: p. 121).

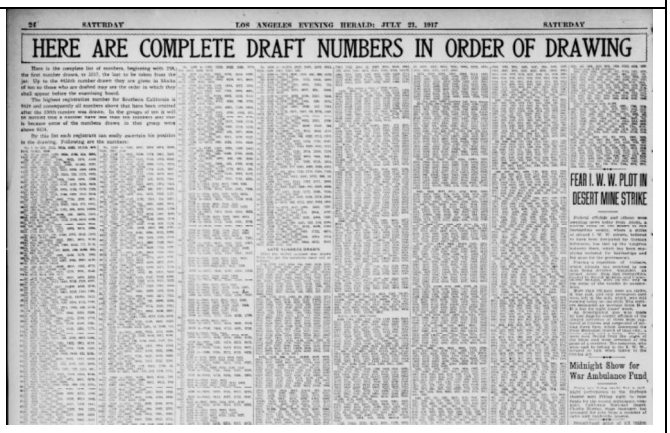
Panel A	Panel B
	

Figure A.5: Order of draft numbers

Note: Pages from the 21st of July 1917 New York Times (Panel A) and Los Angeles Herald (Panel B) showing the list of draft numbers drawn in D.C. the day before. We digitize both lists and manually correct inconsistencies.

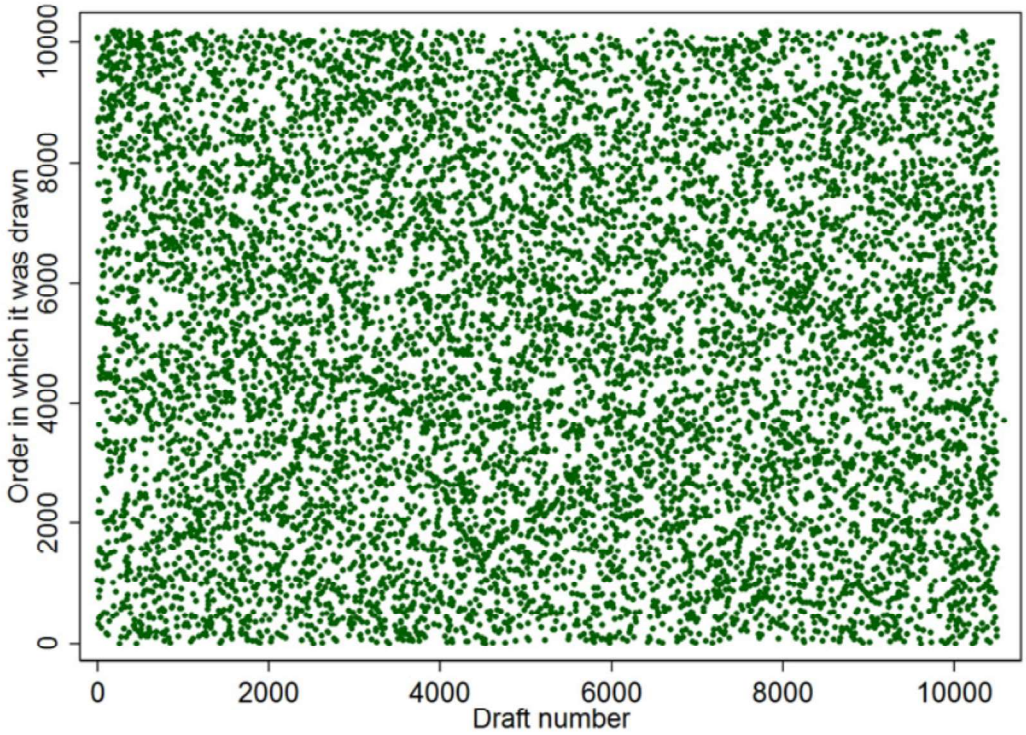


Figure A.6: Randomness of the lottery

Note: Horizontal axis: draft numbers as assigned to lottery registrants. Vertical axis: order in which each of these numbers was drawn during the July 1917 lottery in Washington DC.

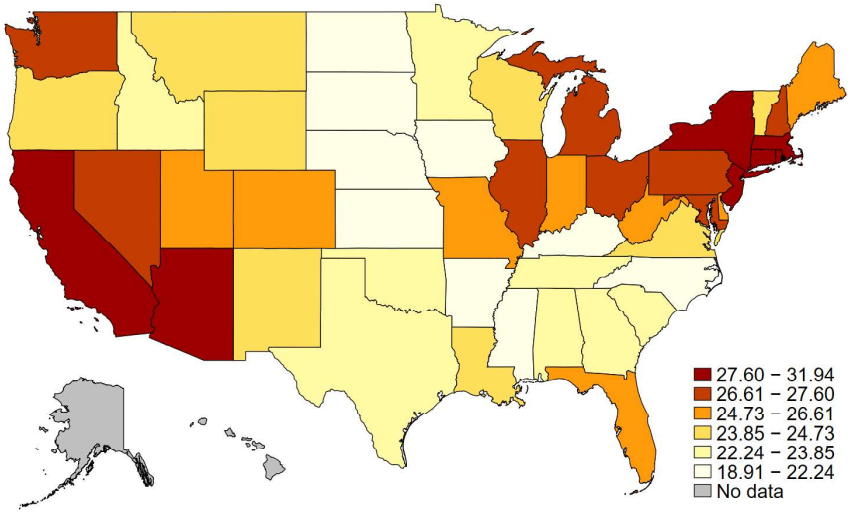


Figure A.7: Occupation scores in 1930, by state

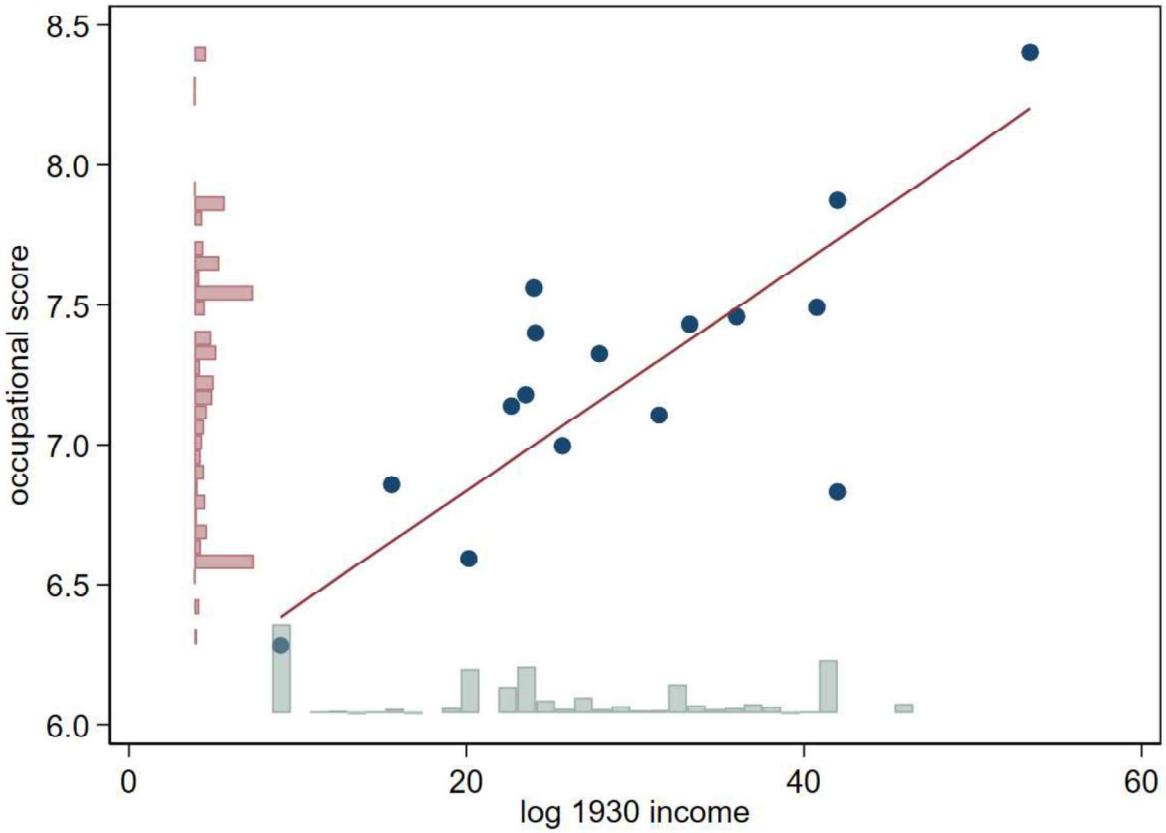


Figure A.8: Binscatter of occupational scores and earnings

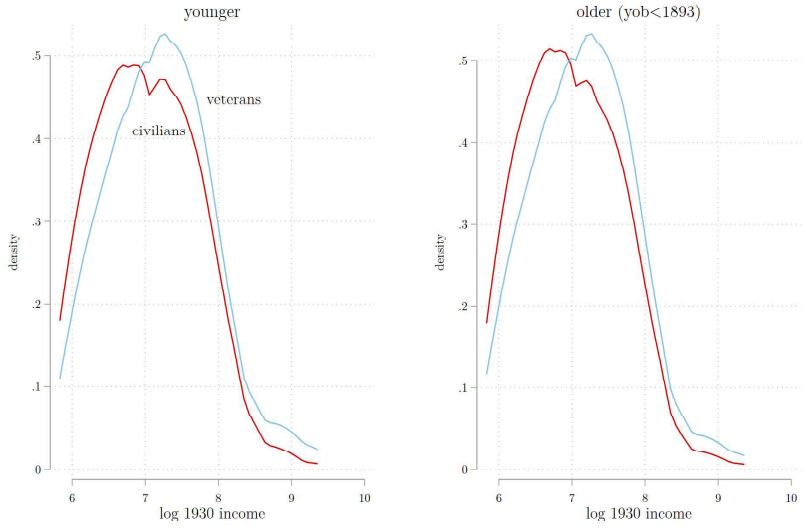


Figure A.9: Distribution of 1930 log earnings, by age and veteran status

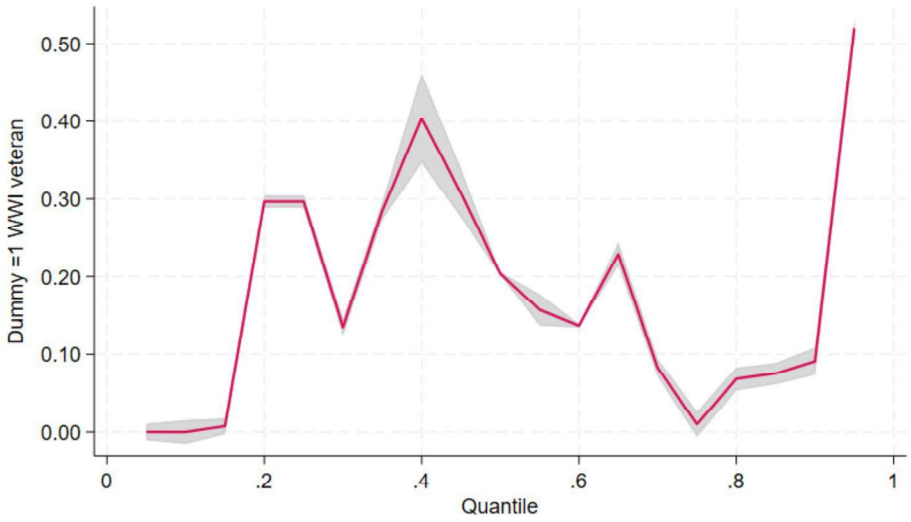


Figure A.10: Quantile regressions, conditional effects

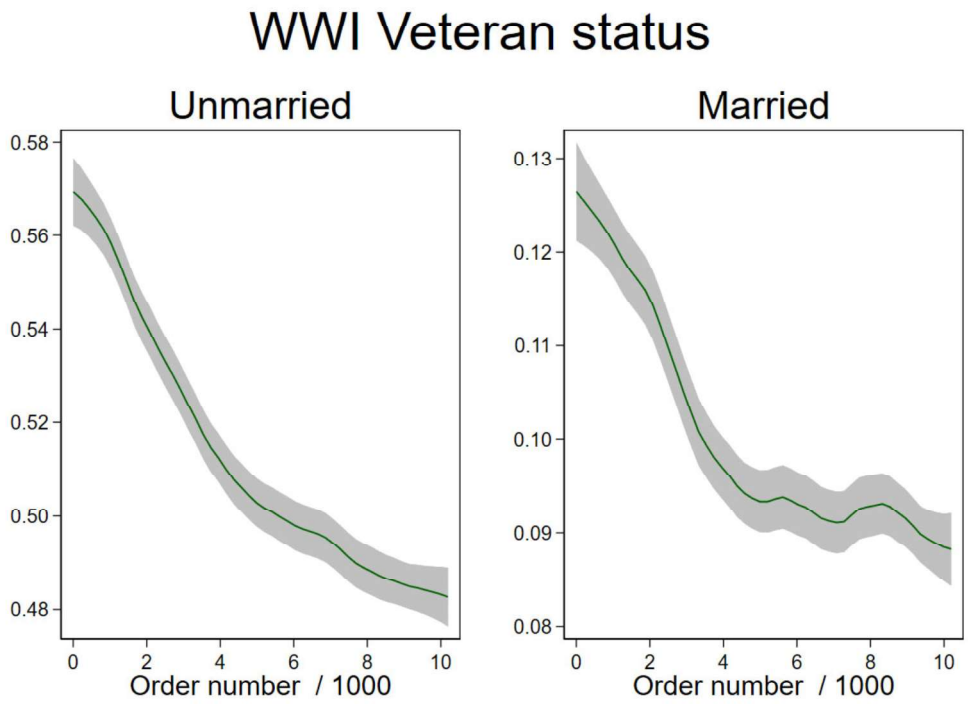


Figure A.11: Probability of military service and order number, married vs unmarried men

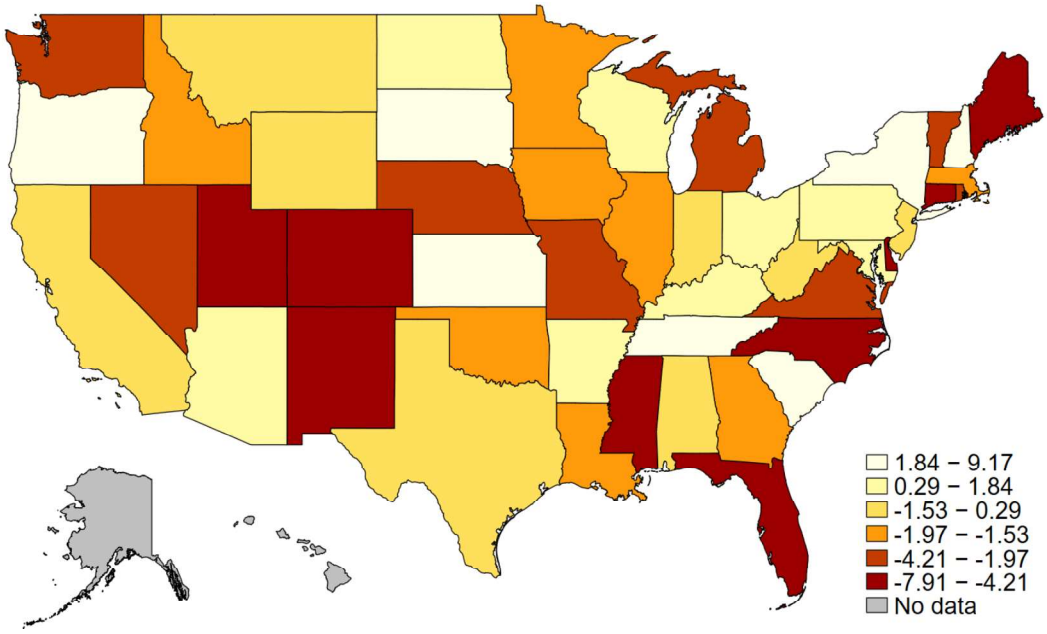


Figure A.12: Effect of Order Number on Earnings, by State

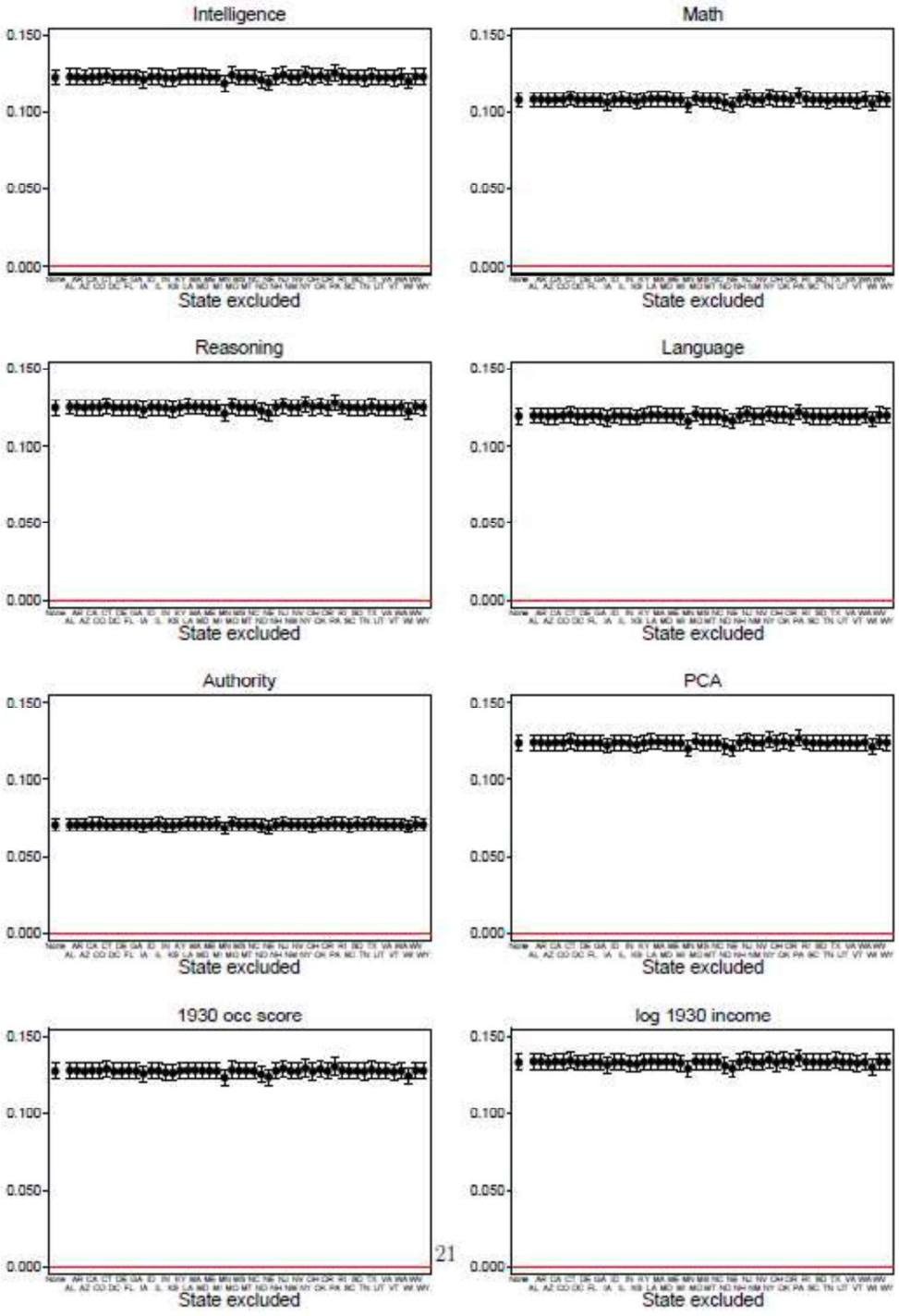


Figure A.13: OLS Results - Robustness to dropping one state at a time

Note: Coefficients represent the standardized effect of WWI veteran status on each separate dependent variable. The leftmost coefficient in each panel reproduces the baseline results obtained keeping all observations.

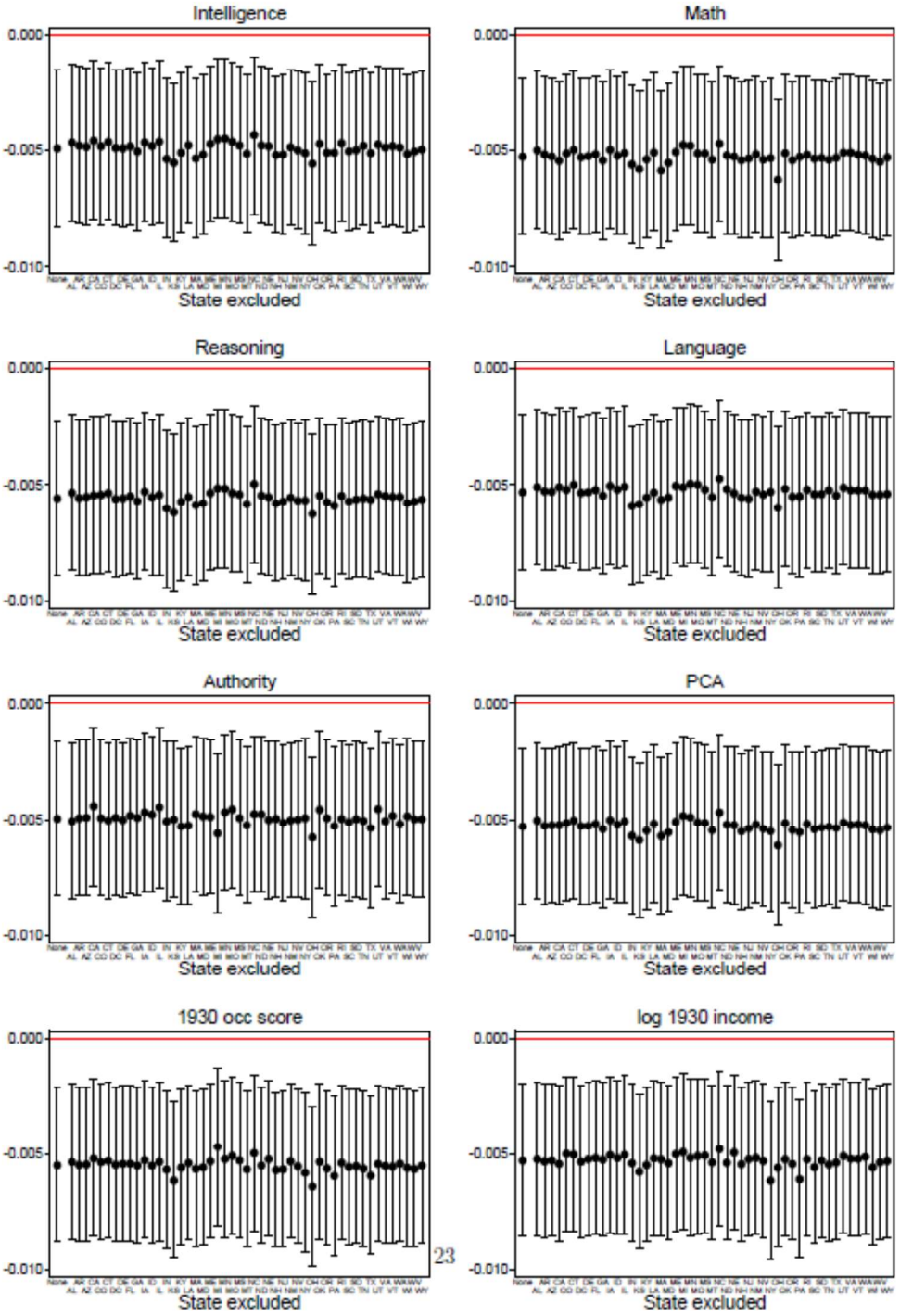


Figure A.14: Robustness of reduced form to dropping one state at a time

Note: Coefficients represent the standardized effect of the order number on each separate dependent variable. The leftmost coefficient in each panel reproduces the baseline results obtained keeping all observations.

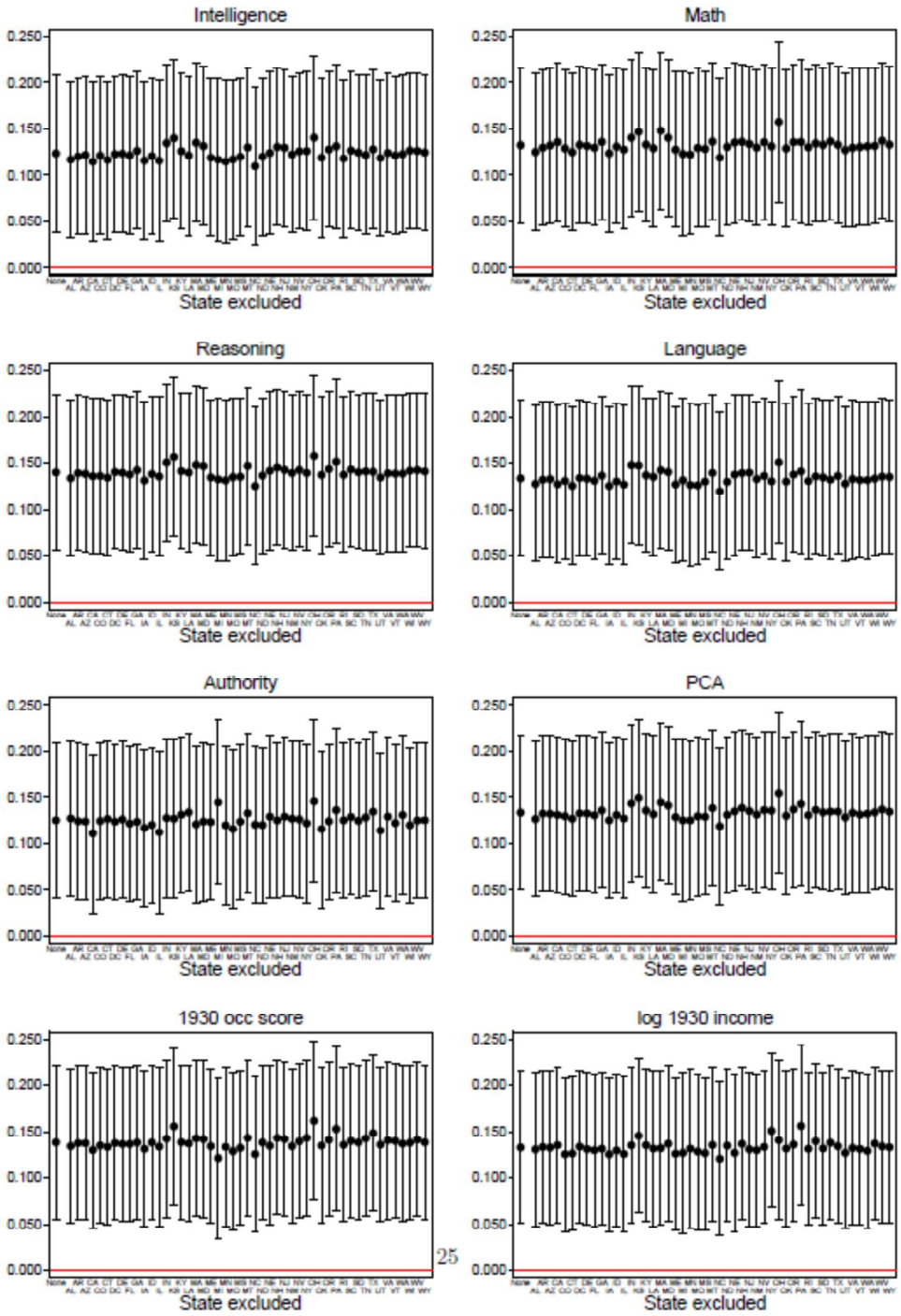


Figure A.15: Robustness of IV to dropping one state at a time

Note: Coefficients represent the standardized effect of WWI veterans status instrumented with the lottery on each separate dependent variable. The leftmost coefficient in each panel reproduces the baseline results obtained keeping all observations.

	Population	Matched	Sample	Diff (Pop-Matched)	Diff (Matched-Sample)	t-stat (Pop-Sample)	t-stat (Matched-Sample)
Birth year 1886	0.046	0.042	0.042	0.004	0.000	12.0	-0.009
Birth year 1887	0.083	0.076	0.076	0.006	0.000	15.0	-0.003
Birth year 1888	0.093	0.086	0.086	0.007	0.000	17.0	0.018
Birth year 1889	0.090	0.084	0.084	0.006	0.000	15.0	0.011
Birth year 1890	0.089	0.085	0.085	0.004	0.000	9.6	-0.004
Birth year 1891	0.087	0.085	0.085	0.002	0.000	4.6	0.009
Birth year 1892	0.096	0.095	0.095	0.001	0.000	3.0	0.009
Birth year 1893	0.092	0.093	0.093	-0.001	0.000	-1.0	-0.037
Birth year 1894	0.094	0.097	0.097	-0.003	0.000	-7.0	0.004
Birth year 1895	0.095	0.099	0.099	-0.004	0.000	-10.0	0.004
Birth year 1896	0.089	0.101	0.101	-0.012	0.000	-28.0	-0.008
White	0.783	0.850	0.850	-0.067	0.000	-114.0	0.022
Black	0.094	0.030	0.030	0.064	0.000	155.0	0.267
First Name Freq.	201,372	110,937	111,226	90,145	-289	212.0	-0.791
Last Name Freq.	9,836	4,580	4,546	5,289	33	211.0	1.781
Born Abroad	0.251	0.081	0.081	0.170	-0.001	277.0	-1.000
Born Midwest	0.264	0.423	0.422	-0.158	0.001	-248.0	1.230
Born Northeast	0.175	0.205	0.204	-0.029	0.000	-53.0	0.504
Born South	0.279	0.227	0.228	0.051	-0.001	79.0	-1.000
Born West	0.032	0.064	0.065	-0.033	0.000	-128.0	-0.457
Observations	10'588'563	1'234'547	512'829				

Table A.1: Population of interest, matched sample and sample of analysis

Note: The first three columns report average characteristics for the population of men who signed up for the 1917 lottery (col. 1), the sub-set of these men who could be matched to both 1930 and 1940 samples (col. 3) and the men randomly selected from the matched to be part of the analysis. Cols. 4 and 5 show differences between col. 1 and cols. 2 and 3. Cols. 6 and 7 report corresponding *t*-stats for the null that these differences are 0.

Panel A: OLS

	Intelligence	Math	Reasoning	Language	Authority	PCA	Occ Score	log income
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
WWI veteran	0.129***	0.119***	0.135***	0.124***	0.072***	0.133***	0.138***	0.144***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Married before 1917	0.014***	0.013***	0.007*	0.018***	0.009**	0.007*	-0.007*	-0.006*
	[0.000]	[0.000]	[0.076]	[0.000]	[0.014]	[0.051]	[0.067]	[0.097]
Black	-0.091***	-0.091***	-0.098***	-0.093***	-0.051***	-0.096***	-0.072***	-0.087***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Other race	-0.009***	-0.008***	-0.010***	-0.009***	-0.003	-0.009***	-0.009***	-0.007**
	[0.001]	[0.003]	[0.001]	[0.001]	[0.199]	[0.001]	[0.001]	[0.012]
Race missing	-0.008***	-0.007**	-0.007**	-0.007**	-0.004	-0.007**	-0.001	-0.003
	[0.005]	[0.028]	[0.027]	[0.017]	[0.175]	[0.031]	[0.700]	[0.255]
Age FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.024	0.022	0.027	0.023	0.007	0.026	0.025	0.029
Mean dep. var.	2.002	2.549	3.301	2.708	0.080	-0.256	23.984	6.983
Observations	112704	112704	112704	112704	112704	112704	112704	112507

Panel B: reduced form

	Intelligence	Math	Reasoning	Language	Authority	PCA	Occ Score	log income
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Order number	-0.011***	-0.010***	-0.011***	-0.010***	-0.004	-0.010***	-0.011***	-0.008***
	[0.000]	[0.001]	[0.000]	[0.001]	[0.107]	[0.000]	[0.000]	[0.004]
Married before 1917	-0.021***	-0.016***	-0.024***	-0.017***	-0.018***	-0.024***	-0.036***	-0.038***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Black	-0.101***	-0.101***	-0.110***	-0.102***	-0.054***	-0.108***	-0.084***	-0.099***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Other race	-0.012***	-0.012***	-0.014***	-0.013***	-0.006***	-0.013***	-0.011***	-0.012***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.002]	[0.000]	[0.000]	[0.000]
Race missing	-0.009***	-0.008***	-0.008***	-0.008***	-0.005	-0.008***	-0.001	-0.004
	[0.003]	[0.010]	[0.007]	[0.004]	[0.105]	[0.010]	[0.748]	[0.190]
Age FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.012	0.012	0.014	0.012	0.003	0.014	0.009	0.012
Mean dep. var.	2.010	2.564	3.318	2.719	0.080	-0.209	24.252	6.998
Observations	127267	127267	127267	127267	127267	127267	127267	127038

Panel C: two stages least squares

	Intelligence	Math	Reasoning	Language	Authority	PCA	Occ Score	log income
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
WWI veteran	0.256***	0.222***	0.245***	0.223***	0.103	0.239***	0.255***	0.184***
	[0.000]	[0.001]	[0.000]	[0.001]	[0.106]	[0.000]	[0.000]	[0.004]
Married before 1917	0.089***	0.079***	0.081***	0.078***	0.026	0.079***	0.073***	0.041
	[0.001]	[0.005]	[0.004]	[0.005]	[0.346]	[0.005]	[0.010]	[0.132]
Black	-0.100***	-0.100***	-0.109***	-0.101***	-0.054***	-0.107***	-0.082***	-0.098***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Other race	-0.011***	-0.011***	-0.012***	-0.012***	-0.005***	-0.012***	-0.009***	-0.011***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.007]	[0.000]	[0.001]	[0.000]
Race missing	-0.007**	-0.006*	-0.006*	-0.007**	-0.004	-0.006*	0.001	-0.002
	[0.027]	[0.056]	[0.052]	[0.028]	[0.202]	[0.063]	[0.676]	[0.449]
Age FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.010	0.012	0.015	0.015	0.006	0.015	0.009	0.023
Mean dep. var.	2.010	2.564	3.318	2.719	0.080	-0.209	24.252	6.998
Observations	127267	127267	127267	127267	127267	127267	127267	127038

Table A.2: Coarsened Exact Match samples with stricter condition

Note: The table reports beta coefficients from regressions run on reduced sample comprising only treated and control men who have the same observable characteristics. Panel A: OLS regressions: we define treatment as being WWI veteran. Panel B and C: reduced form and two stages least squares regressions: we define treatment as having an order number below 4500. Relative to Table 5, we force matches within age and state. P-values based on standard errors clustered at registration district level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

	Intelligence	Math	Reasoning	Language	Authority	PCA	Occ Score	log income
WWI veteran	0.123	0.108	0.125	0.119	0.071	0.124	0.127	0.133
Cluster: registration board	[0.000]***	[0.000]***	[0.000]***	[0.000]***	[0.000]***	[0.000]***	[0.000]***	[0.000]***
Conley s.e. (p-value): 10 Km cutoff	[0.000]***	[0.000]***	[0.000]***	[0.000]***	[0.000]***	[0.000]***	[0.000]***	[0.000]***
Conley s.e. (p-value): 25 Km cutoff	[0.000]***	[0.000]***	[0.000]***	[0.000]***	[0.000]***	[0.000]***	[0.000]***	[0.000]***
Conley s.e. (p-value): 50 Km cutoff	[0.000]***	[0.000]***	[0.000]***	[0.000]***	[0.000]***	[0.000]***	[0.000]***	[0.000]***

Table A.3: Robustness to spatial inference

Note: The table reports beta-coefficients for all main regressions. P-values based on standard errors allowing for spatial correlation are in parentheses. We use the method of Conley (1999) and calculate s.e. assuming that spatial correlation disappears at different distances. *** p<0.01, ** p<0.05, * p<0.1.

Panel A: Robustness regression for OLS

	Intelligence	Math	Reasoning	Language	Authority	PCA	Occ Score	log income
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
WWI veteran	0.119***	0.116***	0.126***	0.114***	0.002***	0.123***	0.114***	0.133***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Married before 1917	0.022***	0.021***	0.019***	0.026***	-0.001***	0.017***	0.016***	0.014***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Black	-0.095***	-0.102***	-0.107***	-0.098***	-0.000***	-0.104***	-0.074***	-0.094***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Other race	-0.022***	-0.023***	-0.024***	-0.023***	-0.000	-0.023***	-0.014***	-0.019***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.602]	[0.000]	[0.000]	[0.000]
Race missing	-0.008***	-0.008***	-0.007***	-0.009***	0.000	-0.007***	-0.002	-0.006***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.753]	[0.000]	[0.226]	[0.004]
Age FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.022	0.020	0.024	0.022	0.007	0.024	0.017	0.023
Mean dep. var.	2.060	2.632	3.396	2.793	0.087	0.015	25.574	7.064
Observations	350031	350031	350031	350031	350031	350031	350031	349145

Panel B: Robust Regressions for reduced form

	Intelligence	Math	Reasoning	Language	Authority	PCA	Occ Score	log income
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Order number	-0.006***	-0.006***	-0.006***	-0.006***	-0.000	-0.006***	-0.006***	-0.006***
	[0.001]	[0.001]	[0.000]	[0.000]	[0.409]	[0.000]	[0.000]	[0.001]
Married before 1917	-0.029***	-0.027***	-0.034***	-0.024***	-0.002***	-0.034***	-0.032***	-0.042***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Black	-0.096***	-0.102***	-0.108***	-0.098***	-0.000***	-0.104***	-0.074***	-0.094***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Other race	-0.023***	-0.024***	-0.025***	-0.024***	-0.000	-0.024***	-0.015***	-0.020***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.307]	[0.000]	[0.000]	[0.000]
Race missing	-0.012***	-0.012***	-0.012***	-0.012***	-0.000	-0.011***	-0.006***	-0.010***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.408]	[0.000]	[0.002]	[0.000]
Age FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.011	0.011	0.013	0.011	0.003	0.013	0.007	0.010
Mean dep. var.	2.060	2.632	3.396	2.793	0.087	0.015	25.574	7.064
Observations	350031	350031	350031	350031	350031	350031	350031	349145

Table A.4: Robustness to outliers and influential observations

Note: The table reports beta-coefficients from regressions where outliers and influential observations are downweighted, as per the method in Li (1985). Panel A: OLS regressions. Panel B: reduced form. P-values based on standard errors clustered at registration district level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

	Intelligence	Math	Reasoning	Language	Authority	PCA	Occ Score	log income
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Veteran (any war)	0.122***	0.131***	0.138***	0.132***	0.123***	0.132***	0.136***	0.131***
	[0.004]	[0.002]	[0.001]	[0.002]	[0.004]	[0.002]	[0.001]	[0.002]
Married before 1917	0.015	0.026	0.019	0.020	0.026	0.016	0.013	0.009
	[0.414]	[0.157]	[0.280]	[0.273]	[0.144]	[0.373]	[0.457]	[0.615]
Black	-0.093***	-0.097***	-0.102***	-0.096***	-0.055***	-0.100***	-0.073***	-0.091***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Other race	-0.021***	-0.021***	-0.022***	-0.022***	-0.010***	-0.021***	-0.013***	-0.018***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Race missing	-0.007***	-0.006**	-0.006**	-0.007***	-0.005**	-0.006**	-0.002	-0.005**
	[0.003]	[0.014]	[0.011]	[0.002]	[0.021]	[0.012]	[0.538]	[0.038]
Age FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.023	0.020	0.025	0.022	0.005	0.024	0.020	0.024
Mean dep. var.	2.060	2.632	3.396	2.793	0.087	0.015	25.574	7.064
Rubin-Anderson test (p-value)	0.004	0.002	0.001	0.002	0.004	0.002	0.001	0.002
Observations	350031	350031	350031	350031	350031	350031	350031	349145

Table A.5: Robustness to definition of veteran status

Note: The table reports beta-coefficients from 2SLS regressions where the endogenous variable is a dummy equal to one if the man reported being a veteran (regardless of the war he declared). The instrument is the order number determined by the lottery. P-values based on standard errors clustered at registration district level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

	1930 occ score			log 1930 income		
	(1)	(2)	(3)	(4)	(5)	(6)
	21-24	25-27	28-30	21-24	25-27	28-30
Order number	0.001	-0.013**	-0.008*	0.001	-0.008	-0.008*
	[0.812]	[0.017]	[0.072]	[0.832]	[0.139]	[0.052]
Black	-0.060***	-0.053***	-0.054***	-0.080***	-0.075***	-0.071***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Other race	-0.016***	-0.013***	-0.012***	-0.021***	-0.017***	-0.015***
	[0.000]	[0.005]	[0.001]	[0.000]	[0.000]	[0.000]
Race missing	-0.008*	-0.008*	-0.003	-0.015***	-0.010**	-0.001
	[0.065]	[0.081]	[0.521]	[0.000]	[0.040]	[0.843]
Age FEs	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.005	0.003	0.003	0.007	0.006	0.006
Observations	77966	37690	54906	77751	37557	54719

Table A.6: Reduced form by age cohort

Note: The table reports beta coefficients of the effect of order number on 1930 occupational scores (cols 1–3) and log 1930 earnings (cols 4–6), by age tercile. P-values based on standard errors clustered at registration district level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.