Human-capital shocks and innovation: Evidence from Britain's Lost Generation

Luca Repetto ®[1]

Uppsala University

Davide Cipullo ®[2]

Università Cattolica del Sacro Cuore

Edward Pinchbeck ®[3]

University of Birmingham

Jan Bietenbeck ®[4]

Lund University

Abstract

This paper studies how World War I mortality shocks to British communities affected long-run innovation. Linking parish-level military deaths to the universe of patent records (1895–1979) and inventor data, we compare areas exposed to different levels of mortality. Moving a parish from the 25th to the 75th percentile of WWI mortality reduces the probability that it produces any patent by about 9 percent of the mean, with much larger effects for breakthrough patents; these declines persist for decades. Mortality reduces both the emergence of new inventors and subsequent patenting by established ones, particularly in frontier and technologically complex fields. The decline in new inventors is smaller in areas with stronger pre-existing innovation ecosystems.

Keywords: World War; Innovation; Human Capital; Patents; Lost Generation
JEL classification: D74, O15, O31

First version: Sept. 2024. This version: September 2026

[1]Department of Economics, Uppsala University. Email: luca.repetto@nek.uu.se.
[2]Department of Economics, Università Cattolica del Sacro Cuore. Email: davide.cipullo@unicatt.it
[3]Department of Economics, University of Birmingham. Email: e.w.pinchbeck@bham.ac.uk.
[4]Department of Economics, Lund University. Email: jan.bietenbeck@nek.lu.se

Authors' order was randomised. We would like to thank Jingyuan Zeng, Gustaf Holmberg and Rebecca Wiechers for excellent research assistance, as well as Felipe Carozzi, Adrian Gortzak, Elias Papaioannou, Vincenzo Scrutinio, Marco Tabellini and participants in seminars held at ASREC Europe Conference 2023 and 2025, University of Birmingham, Uppsala University, Università Cattolica in Milan, UIL/HEUU/Urban lab workshop in Sigtuna, ASWEDE 2024, SIEP 2025, Università di Bologna, CEPR Applied Micro History Workshop (St. Gallen) for comments and suggestions. This work was supported by BA/Leverhulme Small Research Grant SRG2021-210936, Jan Wallanders och Tom Hedelius stiftelse samt Tore Browaldhs stiftelse grants P22-0060 and P24-0048, and the Centre for Economic Performance. This work is based on data provided through www.VisionofBritain.org.uk and uses statistical material which is copyright of the Great Britain Historical GIS Project, Humphrey Southall and the University of Portsmouth.

1. Introduction

Technological innovation is central to long-run economic growth (e.g. Romer, 1990; Aghion and Howitt, 1992; Glaeser et al., 1992), and human capital is one of its key determinants. Theories of directed technical change predict that labour scarcity increases incentives for firms to develop labour-saving technologies (Hicks, 1932; Acemoglu, 2002), and evidence from historical population shocks is broadly consistent with this channel (Habakkuk, 1962; Hanlon, 2015; Bergeaud, Chaniot and Malgouyres, 2025). At the same time, innovation depends on local knowledge transmission and on the supply of scarce expertise embodied in inventors, engineers, and other complementary workers (Bell et al., 2019*b*; Akcigit et al., 2018), who may be adversely affected by such shocks. The two channels can therefore work in opposite directions: labour scarcity may stimulate labour-saving innovation, while human-capital losses may depress innovative activity when mortality falls disproportionately on the people who produce and transmit ideas. Existing research has convincingly documented how innovation responds to labour scarcity, but we know comparatively less about whether human-capital losses inhibit innovation by disrupting the local supply of ideas.

In this paper, we focus on this channel and study how local human-capital losses caused by mortality in World War I (WWI) affected innovation in Britain in the long run. We exploit substantial variation in WWI soldier mortality across parishes. In aggregate, military deaths were a small share of the pre-war population, but the losses were particularly severe among highly educated groups, with death rates of around 14 percent among officers and close to 20 percent among Oxford and Cambridge graduates (Winter, 1977). The war therefore imposed a disproportionately large shock to communities' human capital, contributing to the notion of a "lost generation" of young men (Winter, 1977). This combination of a modest aggregate loss of labour but a concentrated loss of human capital makes WWI Britain particularly informative about the supply-of-ideas channel. Because fighting took place abroad, Britain was spared the large-scale destruction of domestic physical capital that often accompanies war, allowing us to isolate the consequences of these human-capital losses from the direct effects of physical destruction.

Our main result is that WWI mortality persistently reduced local innovation, with substantially larger effects on the most innovative patents. The decline is not driven by the disappearance of inventors who were killed in the war. Instead, higher mortality reduces both subsequent patenting among surviving pre-war inventors and the emergence of successful new inventors. Both margins point to a human-capital supply-side channel, with effects concentrated where innovation relies most on skilled and complementary workers and attenuated where local ecosystems can substitute for lost human capital.

To document these effects, we build a new dataset linking WWI mortality records to patent data for England and Wales, 1895–1979. Mortality and military participation come from Commonwealth War Graves Commission death records and military service records geolocated to 1911 parishes, building on Carozzi, Pinchbeck and Repetto (forthcoming). We measure innovation by combining PatentCity geocoded patent filings, which we map to 1911 parish boundaries, with PATSTAT inventor identifiers and patent abstracts. The linked data

cover the universe of georeferenced WWI deaths and patents filed in England and Wales over the period, allowing us to analyse responses at both the parish and inventor levels.

The first part of our analysis examines how WWI fatalities affected parish-level patenting. Patenting is rare at this level – only one in ten parishes registers a patent in a given year – so we aggregate outcomes to 5-year periods and use as the baseline outcome an indicator equal to one if a parish registers at least one patent. Building on Kelly et al. (2021), we also construct text-based measures of patent importance.[1] We estimate difference-in-differences and event-study specifications relating parish outcomes to the (log) number of local WWI deaths, with parish and period-by-county fixed effects. Identification therefore comes from comparing parishes within the same county, while county-wide shocks – including broader demand-side responses to the war – are absorbed.

The identifying assumption is that parishes with different mortality would have followed parallel patenting trends in the absence of the war (Callaway, Goodman-Bacon and Sant'Anna, 2024). A concern is that WWI mortality partly reflects differences in mobilisation and military assignment that may themselves be related to communities' pre-war economic characteristics and subsequent trajectories. Volunteering was uneven across communities, and reserved-occupation exemptions protected some industrial workers but not others, so high-mortality parishes may have differed systematically from low-mortality ones. We address this concern in three ways. First, event-study estimates show no differential patenting trends across parishes in the three 5-year periods before the war, followed by a sharp decline after 1915. Second, our estimates are stable when we control for pre-war mobilisation and for 1911 population, income, white-collar employment, and unemployment, each interacted with time. Third, in a complementary design we instrument parish deaths with a shift-share predictor that combines battalion-level death rates with each parish's distribution of servicemen across battalions, following Carozzi, Pinchbeck and Repetto (forthcoming).[2]

We find that WWI mortality had a persistent negative effect on innovation. In our preferred specification, moving a parish from the 25th to the 75th percentile of the WWI mortality distribution lowers the probability that it produces any patent in a 5-year period by about 2.2 percentage points, close to 9 percent of baseline patenting; equivalently, a 10 percent increase in WWI deaths lowers this probability by about 0.12 percentage points, or roughly 0.5 percent relative to the baseline mean. The magnitude varies over the post-war period: the effect appears immediately in the 1915–19 period, is somewhat smaller and statistically insignificant in the early 1920s, but increases again from the mid-1920s and is statistically significant by the early 1930s, before the onset of WWII. Coefficients remain negative thereafter through the end of our sample period. Importantly, our conclusions do not rely on the

[1] We focus on "breakthrough patents", those in the right tail of the importance distribution. This helps us address the well-known issue that patents are an imperfect measure of innovation, as not all innovations are patented and not all patents represent true innovation (e.g. Griliches, 1990; Nagaoka, Motohashi and Goto, 2010).

[2] The identifying idea is that, conditional on parish characteristics, battalion-level mortality reflects deployment and the fortunes of war rather than local determinants of innovation. The instrument is uncorrelated with pre-war parish observables, the first stage is strong, and the IV estimates are close to, and slightly larger than, the DiD estimates.

post-WWII decades: estimates using only pre-WWII data are similar, and the inventor-level results described below show no additional decline around WWII.

Additional results further characterise the negative effect on patenting. The effect on the probability that a parish produces any breakthrough patent is roughly three times as large in absolute terms, and larger still relative to its mean. Because the number of deaths scales with parish size, we verify that the estimates remain negative and precisely estimated when we control flexibly for 1911 parish population, when we measure mortality as a death rate or an above-median indicator rather than log deaths, when we retain zero-death parishes using $\log(1+\text{deaths})$, and when we exclude London or all large cities from the sample. Taken together, these results indicate that WWI altered both the geography and the composition of British invention for decades.

Having documented the negative effect of WWI mortality on innovation, we turn to the channels underlying this decline. We organise the analysis around two margins: the subsequent innovative activity of inventors who were already active before WWI, and the emergence of successful new inventors ("Lost Einsteins," Bell et al. 2019*b*).

To study the first margin, we construct an inventor-level panel of roughly 54,000 patentees who registered at least one patent in the pre-war period and track them through 1979. We assign each inventor WWI mortality exposure based on their pre-war location and follow their subsequent patenting irrespective of where they later reside. Inventors initially located in higher-mortality parishes patent less after the war, and the effect remains strongly negative when we exclude inventors who were themselves killed in WWI. Thus, the decline among established inventors reflects a reduction in the subsequent innovative activity of surviving inventors rather than simply the direct loss of patentees in the war.

The inventor-level analysis uncovers several dimensions of heterogeneity. The effect is not uniform across technological fields, but is larger among inventors patenting in sectors such as electricity, mechanics, transportation, physics, and chemistry, and modest and not statistically significant in human necessities and construction. In addition, inventors with engineering occupations and those whose first patents are more complex are more heavily affected.[3] Overall, the declines are concentrated among inventors operating in more skill-intensive and frontier domains (Smil, 2005), consistent with innovation in these sectors relying on specialised teams, complementary technicians and engineers, and supporting infrastructure (Neffke, 2019; Jaravel, Petkova and Bell, 2018; Hanlon, 2025).

Subsequently, we examine whether mortality changed the composition of patenting of established inventors along three dimensions: location, co-authorship, and field. First, we find that patenting in the inventor's original parish falls while patenting in a different parish rises, consistent with relocation as a margin of adjustment to the mortality shock (Bell et al., 2019*b*). Second, the decline is concentrated in single-authored patents, while the estimate for co-authored patents is close to zero, although imprecise. This is consistent with collaboration insulating inventors from the loss of local complementary workers (Jaravel, Petkova

[3]To ensure comparability across specifications, we assign to each inventor a field and measures of patent complexity and importance based on their first pre-war patent.

and Bell, 2018), though it may partly reflect a change in the composition of the patents that are filed. By contrast, patenting falls both within and outside the inventor's original field, suggesting that switching fields does not offset the decline in patenting.

Turning to the second margin, we test the "lost generation" hypothesis: whether the loss of highly educated and technically skilled individuals reduced communities' ability to generate successful new inventors. Conditional on overall WWI mortality, parishes that lost at least one alumnus of Oxford, Cambridge, or an elite school experienced an additional decline in patenting by new inventors. Using detailed soldier records, we find a particularly large additional decline when a parish experienced at least one engineering-related death (Royal Engineers service or an engineering-related rank among non-privates), whereas other non-private deaths have no detectable effect.[4] The engineering-related estimate is unchanged when we control for the pre-war shares of engineers and white-collar workers interacted with time, suggesting that it does not merely reflect the fact that technical men came from industrial places. When elite and engineering-related deaths are entered jointly, both coefficients remain negative, with the engineer effect being more robust. These results indicate that the composition of wartime losses mattered: deaths of highly educated and technically skilled servicemen had particularly persistent consequences for the emergence of new inventors.[5]

We next examine heterogeneity by local innovation ecosystems, which prior work has shown to matter (see, e.g., Akcigit, Grigsby and Nicholas 2017*b*; Bell et al. 2019*b*; Andrews 2023). The effect on new inventors is attenuated in counties with above-median access to engineering-institute members and above-median proximity to universities, but not in counties with better access to rail stations or post offices. Because baseline patenting by new inventors is about twice as high in the high-access counties, the attenuation is larger in relative terms: the estimated effect amounts to roughly a fifth of the mean in low-access counties but less than a tenth in high-access ones. These patterns suggest that stronger local ecosystems of technical expertise and higher education can partly buffer the effects of human-capital losses on the emergence of new inventors.

Our paper contributes to the literature on the relationship between labour scarcity and technological change. Previous theoretical and empirical work has examined when economies respond to worker shortages by adopting or developing labour-saving technologies (Acemoglu, 2007; Voth, Caprettini and Trew, forthcoming). While our findings are not directly informative on whether directed technological change led to shifts in relative prices or in the direction of innovation, they provide evidence on a distinct local supply-side margin: the loss of human capital, especially at the upper tail, is associated with a persistent decline in innovative capacity. Importantly, our setting differs from earlier studies in that much of the prior

[4]This latter finding aligns with recent evidence that engineers are pivotal to the production of innovative output (Hanlon, 2025) and underscores the role of technically oriented individuals in driving technology adoption and economic growth (Maloney and Valencia Caicedo, 2022).

[5]These results on patenting by first-time inventors can reflect two channels that we cannot separate: the direct loss of men who would have gone on to invent, and the loss of role models and mentors for those who did not serve (Bell et al., 2019*b*).

evidence concerns low-skilled labour (e.g. Hanlon 2015; Andersson, Karadja and Prawitz 2022), whose tasks may have been more readily substituted by machines, and economies in the transition out of agriculture, whereas our evidence comes from an industrialised economy. The closest study is Bergeaud, Chaniot and Malgouyres (2025), who find that WWI mortality increased patenting across French départements, particularly in more educated ones and in labour-saving technologies, consistent with a demand-side response. Our findings are complementary: at a much finer spatial scale, we find that local human-capital losses are associated with lower subsequent innovation.[6] Our contribution is to show that this local supply-side loss extends beyond the inventors who died to the subsequent patenting of surviving inventors and to the emergence of new inventors.

The two sets of findings may be reconciled by the level of analysis. Innovation is highly localised: in our data, 82 percent of the variation in patenting across parishes, and 93 percent of the variation in WWI mortality, lies within rather than between counties. Bergeaud, Chaniot and Malgouyres (2025)'s analysis is at the département level (of which there are about 90), a scale more suited to capturing broader labour-market responses to scarcity. We observe both mortality and invention across roughly 10,000 parishes and can therefore hold county-wide shocks fixed while studying highly local responses in innovative activity. Indeed, once parish mortality is accounted for, county-level mortality has no detectable long-run association with parish patenting. This level of granularity is potentially important because demand-side adjustments to labour scarcity may operate at broader market levels, whereas the productivity of inventors and their collaborators may respond to much more localised losses of human capital. Analyses at broader geographic levels may therefore attenuate or miss these local responses. In this sense, our evidence complements prior work on demand-side effects by highlighting a local human-capital supply margin.

Our paper also contributes to the large body of research on the determinants of innovation and its distribution in space and over time. Recent research has highlighted the importance of individual background as well as external factors, such as exposure to an innovative environment (Bell et al., 2019*b*; Akcigit, Grigsby and Nicholas, 2017*a*; Andrews, 2023). The spatial distribution of innovation is remarkably persistent over time (Andrews and Whalley, 2022) and often concentrated in highly productive clusters (e.g., Moretti 2021; Audretsch and Feldman 1996). However, history offers striking examples of the rapid decline of innovation hubs that once played pivotal roles, such as Florence during the Renaissance, Germany's Ruhr Valley, and Detroit's automobile industry. Our results provide new evidence that exposure to major external shocks such as wars may persistently harm the long-run innovative potential of communities, both in terms of quantity and quality.[7] We also bridge the gap between studies analysing community-level outcomes and those focusing on individual

[6] Consistent with this distinction, Bergeaud, Chaniot and Malgouyres (2025) also find that the deaths of inventors themselves reduced patenting in the affected technology classes.

[7] A notable exception to the general lack of long-term focus is the work on the impact of the Nazi regime on science. Waldinger (2016) uses WWII destruction as a shock to physical and human capital, while Waldinger (2010) shows that the dismissal of scientists in Nazi Germany had long-lasting negative effects on academic careers and the capacity of departments to train new talent.

inventors by providing evidence on both in the same setting and over the long run.

We also contribute to the literature on the long-run economic consequences of war.[8] Prior work finds heterogeneous effects across settings and conflict types: bombings can have persistent negative impacts (Lin, 2022; Riaño and Valencia Caicedo, 2024), yet population and activity often recover (Davis and Weinstein, 2002; Brakman, Garretsen and Schramm, 2004). Chupilkin and Koczan (2022) show that long-run GDP effects vary with conflict characteristics, with on-territory wars yielding larger persistent losses than off-territory wars. Our setting isolates military human-capital losses without physical capital destruction and at a granular spatial scale, allowing us to show that these losses depress innovation at the community level for decades. Besides this finding, we also advance this literature by tracing the evolution of outcomes before, during, and after the war. This allows us to validate our results by inspecting pre-trends and the dynamics of the treatment effect, mitigating concerns about time-varying shocks and the compression of history between treatment and a later outcome in persistence studies (Voth, 2021).

2. Background

2.1. The evolving patent landscape in Britain

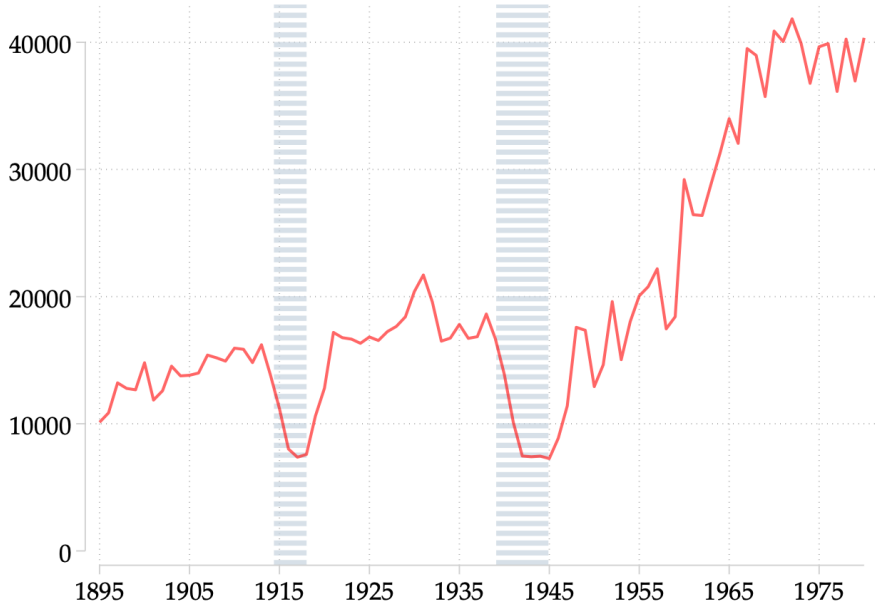
The British Patent Office opened in 1852 and by the start of the period we study (1895–1979), the British patent system was well established. In 1883, an Act significantly reduced patent fees, and around the same time, Britain began to provide full patent rights to signatory countries under an international convention. From 1902, patent applications became subject to an examination of novelty prior to grant. The legislative framework was modernised by a 1977 Act, which underpins the current system.

By the turn of the 20^{th} century, the global distribution of international patents was heavily concentrated in a small group of industrialised nations. Britain was no longer the leading or fastest-growing industrial power, but retained substantial inventive capacity in areas such as mechanical engineering, shipbuilding, textiles, and emerging electrical industries. British inventors were active both domestically and in filing patents in major markets such as the United States and parts of continental Europe. Britain's share of global patenting declined over our analysis period, but the number of patents registered in Britain increased substantially, interrupted by sharp declines during both World Wars (Figure 1). There was also a marked shift towards corporate invention: until the 1920s, the vast majority of patents were registered to individual inventors, whereas by 1950 companies were registering as many patents as individuals (Nicholas, 2011).

An active market for patent rights existed within Britain in the first half of the 20^{th} century and was largely structured around specialist intermediaries — patent agents —

[8]Many papers also consider how war-related losses affect outcomes in the short run, including marriage patterns, female labour force participation and gender norms (Abramitzky, Delavande and Vasconcelos, 2011; Boehnke and Gay, 2020; Brainerd, 2017), while Carozzi, Pinchbeck and Repetto (forthcoming) find that WWI mortality led to the creation of civic capital in Britain. A broader body of related work also examines other mortality shocks, for example due to infectious diseases, see e.g. Franck (2024) for an example that also looks at innovation.

FIGURE 1
PATENTS REGISTERED IN ENGLAND AND WALES



Notes: Total number of patents registered at the Great Britain Patent Office, all patentees, yearly totals for all periods from 1895 to 1979. Source: authors' elaboration from PatentCity.

rather than direct trade between inventors and manufacturers (Nicholas, 2011). The role of agents was twofold: they helped translate inventors' ideas into codified technical knowledge while also acting as commercial brokers between inventors and the market. The Chartered Institute of Patent Agents, incorporated by Royal Charter in 1891, professionalised this intermediation, and the Patents Act of 1902 required agents to register (Pretel, 2018). The number of registered agents rose from 245 in 1894 to 342 in 1930 (Nicholas, 2011). Agents were concentrated in cities, especially London, but written instructions could be conveyed cheaply under flat-rate penny postage. Patent markets for sale, assignment, and licensing were therefore national and increasingly international during the period we study (Khan, 2013).

There is no single dataset that provides comprehensive information about the age, occupation, education, and income of inventors in England and Wales in the first half of the 20^{th} century. Few patentees described their occupation as "inventor" in Census records, and patent databases do not contain comprehensive information on inventor characteristics. Data from Nicholas (2011), however, suggest that between 35 and 45% of independent inventors registering patents between 1910 and 1930 were engineers. British data on inventors' educational backgrounds are more limited. For context, Akcigit, Grigsby and Nicholas (2017*b*) show that around 40% of US inventors in 1940 had a university degree, compared with 10% of non-inventors.

2.2. The British Armed Forces during WWI

Over 4.5 million men from England and Wales served in the British Army during WWI, and around 800,000 more served in the Royal Navy and what became the Royal Air Force (Winter, 1977). The Army expanded rapidly after heavy early casualties, first through volunteering and then through conscription from 1916. Around 330,000 of those who served were commissioned officers (Winter, 1977). At the outbreak of the war, the roughly 20,000 officers in the serving and reserve armed forces were drawn primarily from Britain's upper-middle and upper classes, with recruitment closely tied to prestigious public schools and universities. As demand for officers grew, however, increasing numbers of men from outside these traditional backgrounds were promoted from the ranks. Although exemptions existed for some "reserved occupations," scientific and technical workers were not systematically shielded from frontline service, particularly in the early years of the war. Military service therefore exposed not only a large share of young men, but also substantial numbers of highly educated and technically trained workers to wartime mortality. This composition of service is important for understanding the human-capital losses generated by the war.

2.3. Mortality in WWI and the Lost Generation

Because the war was fought overseas, war-related deaths in Britain were overwhelmingly military: over 750,000 armed-forces deaths compared with fewer than 17,000 civilian deaths from enemy action. Mortality was also strongly age-specific, with particularly high losses among young men and cohorts under 30 (Winter, 1977). This concentration of deaths among young adults underpins the notion of a British "Lost Generation".

Mortality was also unevenly distributed across social groups. Officers, who faced particularly high frontline exposure, had substantially higher death rates than enlisted soldiers, especially in 1914–15 (about 14% versus 6%) (Winter, 1977). Because officer recruitment in the early war years remained closely tied to elite educational institutions, university-educated men were disproportionately affected. Among Oxford and Cambridge graduates, wartime death rates were around 19% and 18%, respectively, and exceeded 25% among students who matriculated between 1910 and 1914.

Evidence for non-elite groups is less clear. Bailey, Hatton and Inwood (2023) show that, among non-officer servicemen, wartime mortality was strongly related to branch and timing of enlistment, with weaker evidence of disadvantage among men from white-collar backgrounds. At the same time, demobilisation statistics indicate that engineering workers represented a substantial share of non-officer servicemen (BWO, 1922). This is relevant because invention during this period was heavily concentrated among engineers. Taken together, these patterns motivate our focus on how localised wartime losses of highly educated and technically trained men affected subsequent innovation.

3. Data and descriptives

We construct a parish-level panel for England and Wales covering 5-year intervals from 1895 to 1979, together with an inventor-level panel spanning the same period.[9] Parishes were the primary units of local government in early twentieth-century England and Wales and closely approximate local communities. We harmonise all spatial data to 1911 parish boundaries, the last Census boundaries observed before WWI. In what follows, we describe the main data sources used and the procedures for constructing our sample.

3.1. Data on WWI deaths

Our main data on WWI soldier deaths come from the casualty database of the Commonwealth War Graves Commission (CWGC) (Commonwealth War Graves Commission, 2023), complemented by Forces War Records (FWR) (Forces War Records, 2023). These sources provide soldier-level information on name, age, date of death, rank, regiment, service unit, and place of birth or residence.[10] We geolocate soldiers to their parishes of origin using these reported locations, following the approach of Carozzi, Pinchbeck and Repetto (forthcoming); Appendix B describes construction and validation in detail.

To construct measures of loss of elite individuals, we match CWGC-FWR records to the Oxford University Roll of Service (Craig, 1920) and the Cambridge War List (Carey, 1921), and we use rank information to identify officers. In addition, we exploit descriptive text fields in the CWGC records to identify further alumni of Oxford, Cambridge, and elite schools in the UK.[11] For mechanism analysis, we also construct an indicator for engineering-related deaths. After dropping privates, we classify as engineering-related those deaths with Royal Engineers service or an engineering-related rank (e.g. Sapper, Pioneer, or ranks containing Engineer, Mechanic, Electrician, Wireless, Signal, or Fitter); remaining non-private deaths are grouped into other non-private categories used to measure non-engineering officer-type losses.

We then aggregate the soldier-level records to the parish level to construct our main treatment variable, the total number of WWI deaths, as well as indicators for whether a parish lost at least one elite-school graduate, at least one officer death, and at least one engineering-related death.

[9] We exclude Scotland from our analysis as Census data and digital maps for Scotland are unavailable for this period.

[10] Our baseline mortality measure prioritises residence and uses birthplace when residence is unavailable or cannot be matched to a parish. To assess whether this choice affects the results, Appendix Table A.5 re-estimates the baseline specifications separately using deaths assigned by parish of residence and deaths assigned by parish of birth. The estimates are very similar across the two measures.

[11] Elite schools are defined as the nine leading schools identified by the Clarendon Commission in 1861: Eton, Charterhouse, Harrow, Rugby, Shrewsbury, Westminster, Winchester, St Paul's, and Merchant Taylors (Shrosbree, 1988).

3.2. *Data on patents*

Our primary source of patent data is PatentCity (Bergeaud and Verluise, 2024), which we restrict to patents registered at the Great Britain Patent Office between 1895 and 1979.[12] The raw data contain about 1.7 million patents and 2.8 million inventor-patent pairs. An important feature of PatentCity is geocoded inventor locations. We map inventor-patent records to 1911 parish boundaries and exclude foreign residents and records without sufficiently precise geolocation. The resulting sample contains about 855,000 inventor-patent pairs, corresponding to around 600,000 patents. We merge these patents to PATSTAT (European Patent Office, 2023) using the common patent identifier. The near-complete match adds persistent inventor identifiers, patent abstracts, and co-authorship information, allowing us to track inventors over time.[13]

Using PATSTAT abstracts, we construct a text-based patent-importance measure following Kelly et al. (2021). We use Cooperative Patent Classification (CPC) codes to identify technological fields and, where necessary, predict fields from patent titles using NLP methods. Appendix C and Appendix D provide details.[14] Using the same patent abstracts, we also construct a measure of patent breadth intended to capture the technological scope of an invention: patents are broader when the language used in their abstract is associated relatively evenly with several technological fields, rather than being concentrated in the vocabulary of a single field. Appendix E provides details.

3.3. *Further data sources*

We supplement these data with several additional sources. We use I-CeM Census microdata for England and Wales in 1911 from IPUMS International (Minnesota Population Center, 2019) to measure parish population and pre-war characteristics, including the share of households with no servants, the share of male white-collar workers, and a proxy for unemployment. We also use British Army Service Records (1914–1918) from FamilySearch (FamilySearch, 2023) to construct parish-level measures of mobilisation.[15] We combine these records with CWGC deaths to construct the shift-share IV sample, applying the restrictions described in Appendix B.

To measure population over time, we construct 5-year parish populations using Census data from "A Vision of Britain through Time" (VoB), spatially re-weighting to 1911 parish polygons and linearly interpolating between Census years. Finally, we construct proxies for pre-war regional innovation ecosystems from four sources: membership lists of engineering trade bodies, a list of pre-WWI universities, rail stations in 1910, and post offices in

[12] We exclude earlier and later years because of incomplete early coverage and a structural break in the data after 1979. Appendix B.5 discusses the construction and terminology of the UK patent records.

[13] Appendix B.5 provides details on geolocation, parish assignment, and the PATSTAT match; Appendix Figure A.1 documents the unmatched records.

[14] Prior to 1978, providing an abstract was optional; nevertheless, fewer than 9 percent of patents in our PATSTAT-based sample lack an abstract.

[15] These records are incomplete because part of the collection was destroyed in a 1940 fire. FamilySearch combines surviving service records ("Burnt documents") and pension-related records ("Unburnt documents"), for about four million records in total. Appendix B describes parish assignment.

1900. For each factor, we compute a county-level accessibility measure based on the log of the population-weighted average distance from a county's parishes to the nearest factor. Appendix B.2 provides further details on sources and construction.

3.4. Sample selection and descriptive statistics

Our final parish-level dataset comprises 10,807 grouped parishes.[16] We retain patents geolocated at least to the city level and aggregate outcomes into 5-year periods from 1895–1899 to 1975–1979, yielding a balanced parish-by-period panel.

Our inventor-level panel spans the same period. We restrict the sample to 54,642 patentees (excluding firms and other non-natural entities) who registered at least one patent in the pre-war period (1895–1914) and track their patenting over time using PATSTAT inventor identifiers.[17] For each inventor, we construct time-invariant characteristics – residence, WWI mortality exposure, field, and patent characteristics – based on their first observed pre-war patent. The engineer indicator instead equals one if the inventor reports an engineering occupation on at least one pre-war patent.

Appendix Table A.1 presents summary statistics for key parish characteristics. In 1911, the average parish had a population of about 3,300 inhabitants and covered 14 square kilometres. On average, each parish contributed around 285 servicemen to the WWI mobilisation effort, implying a mobilisation rate of about 5.5 percent of the population.

Figure 2 provides a visual overview of the variation in WWI losses and patenting activity across parishes in our final dataset. Panel A shows the spatial distribution of soldier deaths per parish, revealing substantial geographic variation. Panel B presents a histogram of soldier deaths, highlighting that roughly one in nine parishes did not experience any soldier fatalities, and that the distribution is highly skewed. On average, parishes reported 54 soldier deaths, equivalent to about 1 percent of the 1911 population, as shown in Appendix Table A.1. Additionally, 6 percent of parishes had at least one soldier who attended an elite school and died in the war, while 11 percent experienced the death of a soldier with an engineering background.

Panel C of Figure 2 maps the average number of patents registered per 5-year period across parishes over the full sample. Patenting is highly spatially concentrated. Panel D shows that even after aggregation to 5-year periods, around three-quarters of parish-period observations have zero patents. Among parishes with patenting activity, the mode is a single patent, although the distribution features a long right tail. Due to this skewness, our main analysis focuses on a binary outcome variable indicating whether at least one patent was registered in a given 5-year period. Its sample mean is 0.24 (Appendix Table A.1).

Most of the variation in both mortality and patenting occurs within counties: 93 percent of the variance in log WWI deaths across parishes and 82 percent of the variance in average parish patenting lies within rather than between counties.[18] Patenting is also highly clus-

[16] For details on parish grouping, see Appendix B.

[17] Appendix B.5 provides details on the sample construction.

[18] These shares are one minus the R^2 from a regression of the parish-level variable on county fixed effects. For mortality, the sample includes only parishes with positive deaths, as in our baseline specification. For patenting,

tered within counties.[19] Using a county-level Herfindahl-Hirschman index, we estimate that innovation clusters effectively span only about 5 percent of a county's parishes, supporting the use of a geography finer than the county. Appendix Figure A.3 further shows that spatial concentration rises with patent importance: more consequential innovations are more geographically concentrated than less consequential ones.

4. Empirical strategy

4.1. Empirical specifications

Our baseline model is a continuous-treatment difference-in-differences (DiD) specification that compares changes in innovation across parishes exposed to different WWI mortality intensity:

$$y_{it} = \alpha + \beta log(d_i^{WWI}) \times Post\ WWI + \gamma' W_{it} + FE + \epsilon_{it}, \tag{1}$$

where y_{it} is an innovation outcome (e.g. an indicator for at least one patent) for parish i in 5-year period t, from 1895-1899 to 1975-1979. $log(d_i^{WWI})$ is the log number of servicemen from parish i killed in WWI, and $Post\ WWI$ indicates periods from 1915 onward.[20] We also include controls W_{it} and different sets of fixed effects that vary by specification.[21]

To study dynamic treatment effects and to evaluate pre-trends, we also estimate the following dynamic DiD model:

$$y_{it} = \alpha + \sum_{k=1895, k\neq 1910}^{1975} \beta_k log(d_i^{WWI}) \times D_k + \gamma' W_{it} + FE + \epsilon_{it}, \tag{2}$$

where D_k is an indicator for 5-year period k. As it is customary, we omit the pre-treatment period 1910-1914, which therefore becomes the baseline.

In addition to the parish-level model, we also use an inventor-level panel for 1895-1979. This allows us to study effects on pre-war inventors, heterogeneity by sector, inventor characteristics, patent complexity, and patent importance, and the composition of remaining patenting by field, location, and co-authorship. Using this panel, we estimate the following inventor-level continuous DiD model:

$$y_{ijt} = \alpha + \beta log(d_i^{WWI}) \times Post\ WWI + \gamma' W_{it} + FE + \epsilon_{ijt}, \tag{3}$$

---

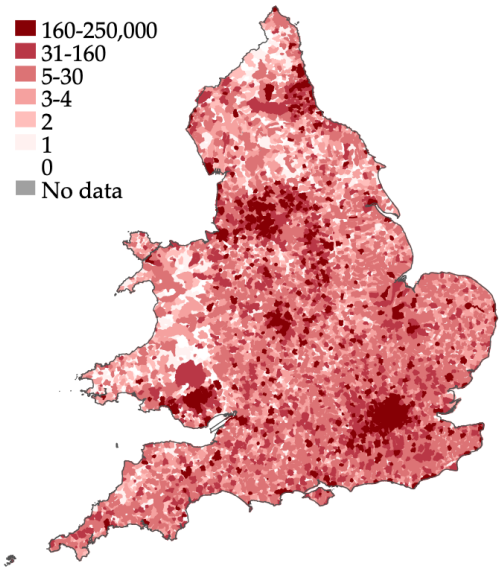
the variable is the parish mean of the "any patent" indicator over 1895–1979.

[19] Appendix Figure A.2 illustrates within-county variation for a random sample of ten counties.

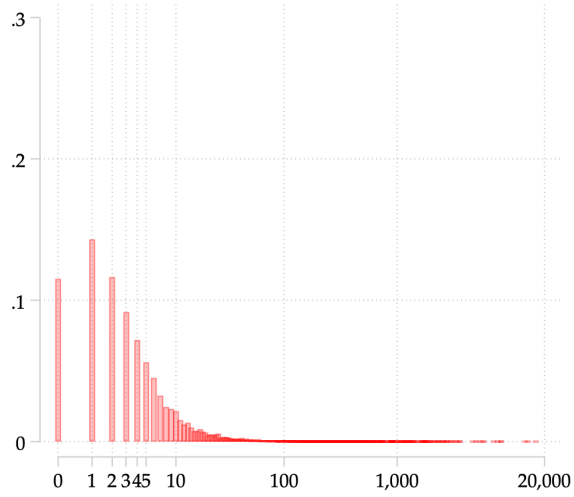
[20] Because we take logs, the baseline model excludes parishes with zero deaths. Baseline results are robust to using $log(1 + d_i^{WWI})$ (Appendix Table A.10), and to measuring mortality as the death rate (deaths per 1911 inhabitant) or as an indicator for above-median deaths (Appendix Table A.2).

[21] These controls are the log of WWI mobilisation at the parish level, as well as variables from the 1911 Census, always interacted with linear time trends. Specifically, we include total population, the share of households in the parish with no servants (as a proxy for income), the share of white-collar workers and a proxy for unemployment (to control for labour market conditions prevailing before the War). The set of controls is purposefully parsimonious because the difference-in-differences assumptions rely on parallel trends – for which there is evidence already in the model without controls –, and to avoid the inclusion of an excessive number of controls and fixed effects (that have to be interacted with linear trends or time effects).

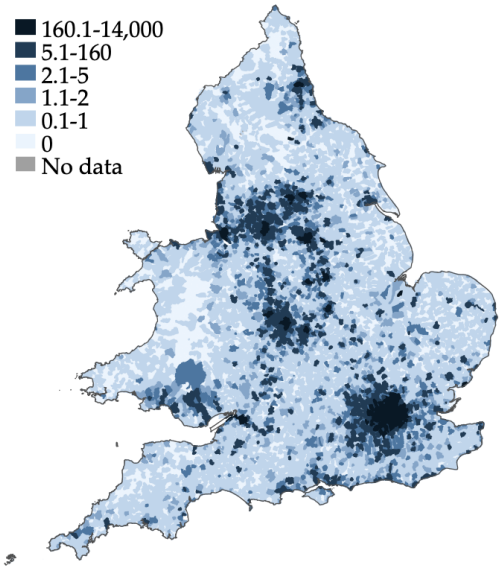
Figure 2
Patents registered and WWI deaths in England and Wales



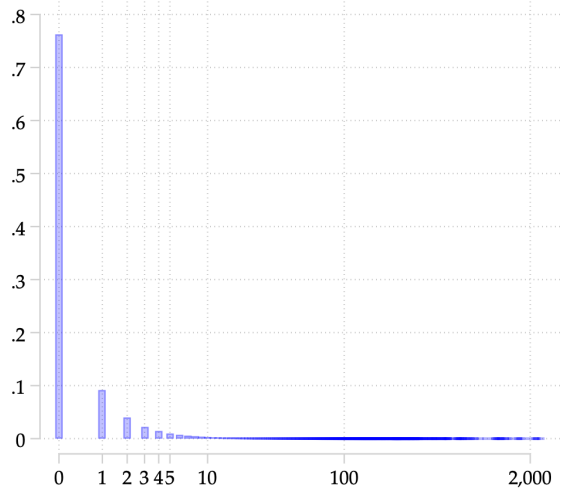
(a) Number of WWI soldier deaths



(b) Number of WWI soldier deaths – density



(c) Number of patents over 5-year period – average



(d) Number of patents over 5-year period – density

Notes: Historical (grouped) parishes in England and Wales. Panel A displays the number of soldiers killed in WWI originating in each parish. Panel B reports the density of the number of soldiers killed in WWI. Panel C shows the number of patents registered over a 5-year period, averaged over all periods in 1895-1979. Panel D shows the density of number of patents registered over a 5-year period (where each period counts as an observation).

where y_{ijt} is an indicator for inventor j from parish i having registered a patent in period t, and the rest is as above.[22]

4.2. Identification assumptions

Because (log) WWI mortality varies continuously across parishes, identification comes from differences in treatment intensity rather than from a treated-versus-untreated comparison. The key DiD assumption is that, conditional on controls and fixed effects, parishes with different mortality intensity would have followed parallel outcome trends absent differential mortality exposure. Under this assumption, the interaction between post-war indicators and $log(d_i^{WWI})$ identifies the marginal causal effect of WWI mortality on innovation outcomes (Callaway, Goodman-Bacon and Sant'Anna, 2024).

Since the log specification excludes zero-death parishes, our estimates do not identify the effect of moving from zero to positive exposure. Instead, they identify how outcomes vary with marginal changes in mortality intensity within the positive-mortality support.

Although the parallel trends assumption is not directly testable, we provide several validation checks in Section 8. First, we show no systematic pre-trend differences across deciles of $log(d_i^{WWI})$ and find similar conclusions when using a binary above-median treatment (Appendix Tables A.2–A.4, Panel C). Second, we show that results from a complementary design in which we instrument parish-level WWI deaths using a shift-share IV following Carozzi, Pinchbeck and Repetto (forthcoming) yield similar results. Specifically, we use as instrument for actual mortality predicted mortality that combines battalion-level mortality rates with each parish's distribution of servicemen across battalions. The exogeneity assumption requires that, conditional on controls, battalion-level mortality reflects differences in combat exposure rather than unobserved parish-level determinants of innovation. We show that the first stage is strong and that IV estimates closely track the baseline DiD in Section 8.

5. Main results

5.1. The local effect of human-capital shocks on innovation in the long run

Table 1 reports DiD estimates from equation 1 of the effect of parish-level WWI mortality on parish-level patenting.[23] Column 1 includes parish and time fixed effects only. The estimate implies that higher parish-level fatalities reduce the probability that residents produce any patent in subsequent decades. In column 2, we add (log) county-level mortality as a proxy for labour-market-wide mortality exposure, allowing us to compare highly localised war losses with broader regional shocks. The county-level measure has no statistically significant long-run effect on patenting in our setting.

[22] For the inventor-level analysis, the vector FE includes inventor fixed effects, time fixed effects, and year of first patent fixed effects interacted with time fixed effects.

[23] In this table and in those that follow, we restrict the sample to that used in column 4 of Table 2, the most demanding specification, which also controls for log parish population. Doing so keeps the number of observations constant across columns and avoids variation due to missing covariate values. The restriction is imposed purely for comparability; none of the results are sensitive to relaxing it.

In column 3, our preferred specification, we introduce county-time fixed effects, thus limiting comparisons to parishes within the same county and absorbing county-wide shocks. The coefficient on parish mortality becomes larger in absolute terms by around a quarter, suggesting that county-level factors correlated with parish-level mortality attenuate the estimated local effect. In this specification, a 10 percent increase in WWI deaths lowers the probability of patenting by about 0.12 percentage points, or roughly 0.5 percent relative to the baseline mean. Given the substantial variation in WWI mortality across parishes, these estimates imply economically meaningful differences across observed levels of exposure: moving from the 25th to the 75th percentile of the mortality distribution among affected parishes lowers the probability of patenting by about 2.2 percentage points, close to 9 percent of the baseline mean.

Columns 4 and 5 assess the robustness of this result to the inclusion of more flexible controls. Column 4 includes interactions of baseline controls (log 1911 population, share of households with no servants, share of white-collar workers, and share unemployed) with a linear time trend. Finally, column 5 controls more flexibly for initial parish size by including the three-way interaction between time, counties, and deciles of 1911 parish population within each county. The estimated effect remains negative across these specifications.

The negative effect of WWI deaths on patenting is not confined to the binary patenting outcome. Appendix Tables A.2, A.3, and A.4 report alternative specifications using patent counts, patents per capita, the parish WWI death rate, and an above-median mortality indicator. The estimates are consistently negative across outcomes and treatment definitions, and are similar when continuous variables are winsorized or trimmed at the 95th percentile, suggesting that the results are not driven by a small number of high-patenting or high-mortality parishes.

Figure 3 examines the dynamics of these effects by presenting estimates and 95 percent confidence intervals from the event study specification in equation 2 that includes parish and county-time fixed effects (i.e. the specification corresponding to column 3 of Table 1). Inspection of the figure reveals that all estimated pre-trend coefficients are small and not statistically different from zero. In the 5-year period that includes the Great War, patenting drops sharply in high-mortality parishes. The gap narrows in the early 1920s, when the coefficient is smallest and not statistically different from zero, although it does not close. The effect widens again from the mid-1920s and is again large and statistically significant by 1930–34, before the onset of WWII. Coefficients for 1935–39 and all subsequent periods are of the same order as, or larger than, the wartime drop, so that the decline persists to the end of the sample.[24]

For comparison, Appendix Figure A.4 plots an event-study specification that includes both parish- and county-level mortality (corresponding to column 2 of Table 1). The pattern for parish-level deaths is effectively identical to that in Figure 3. By contrast, county-level mortality has a different effect on parish innovation: larger county shocks are linked to no

[24] Pooling the four inter-war periods (1920–39) yields a coefficient of -0.008 ($p = 0.001$); pooling 1940–79 yields -0.016. In Section 8 we show that restricting the sample to pre-WWII periods leaves the estimates unchanged.

change, or even a modest increase, in patenting in the short to medium run, but to lower patenting in the long run. These results underscore that the spatial scale at which population shocks are measured is crucial for understanding their effects on places.

The results using our instrumental variables approach for WWI deaths closely mirror those from the difference-in-differences specification discussed here. Therefore, to maintain clarity and focus, we present the IV results separately in Section 8, and concentrate on the difference-in-differences specification throughout the main analysis.

TABLE 1
WWI DEATHS AND PATENTING – OLS ESTIMATES

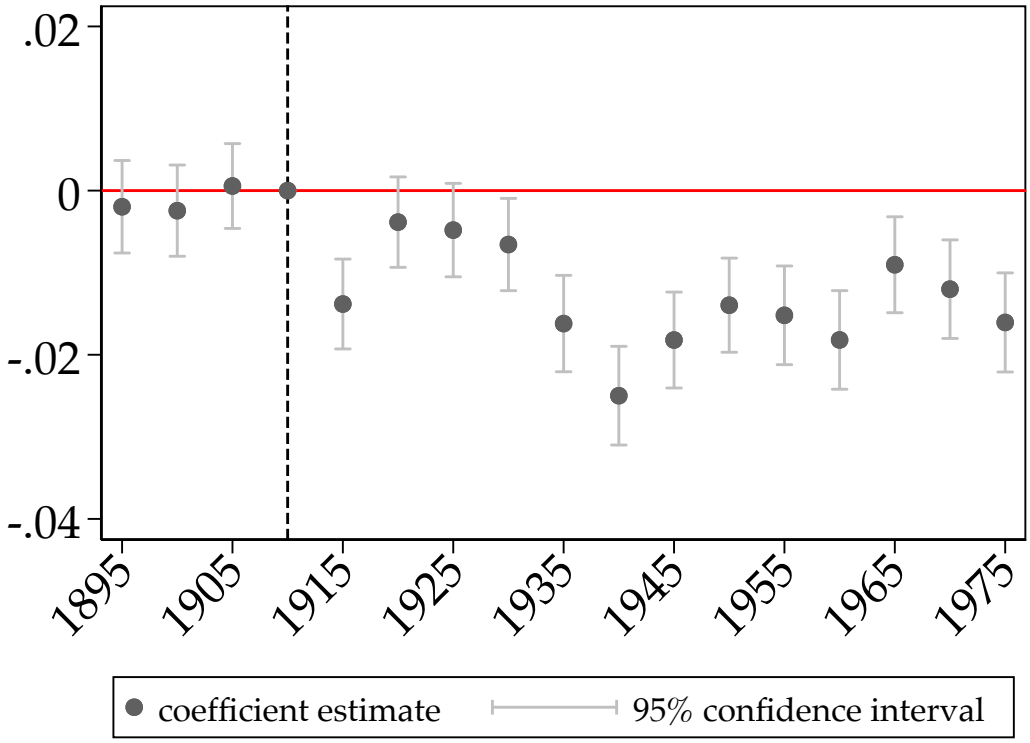
	(1) Any patent	(2) Any patent	(3) Any patent	(4) Any patent	(5) Any patent
$Log(d^{WWI})\times$ Post WWI	-0.009***	-0.009***	-0.012***	-0.008***	-0.007***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)
$Log(d^{County})\times$ Post WWI		0.001			
		(0.002)			
Mean dep.var.	0.26	0.26	0.26	0.26	0.26
R2	0.48	0.48	0.49	0.49	0.51
Observations	162021	162021	162021	162021	162021
Parish FE	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y
County $\times$ Time	N	N	Y	Y	Y
Controls $\times$ Linear trend	N	N	N	Y	N
County $\times$ Time $\times Q_n$ Pop	N	N	N	N	Y

Notes: OLS estimates of the effect of WWI deaths on the probability that at least one patent is registered by a resident of parish i during 5-year period t. Column 1 includes parish and time fixed effects. Column 2 adds log county-level mortality. Column 3 adds county-time fixed effects. Column 4 adds baseline controls interacted with a linear time trend. Column 5 controls for a three-way interaction between time, counties, and deciles of 1911 parish population within each county. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

5.2. Effects by patent quality

The results presented thus far demonstrate that WWI mortality led to a decline in the production of local innovation. However, this analysis centres on the *quantity* of innovation and does not consider its *quality* – that is, the technological significance and impact of each patent. It is plausible that WWI mortality altered the balance between the number and the importance of innovations, potentially reducing the overall volume of patenting without necessarily affecting the production of high-impact inventions. To investigate this question, we study whether WWI human-capital losses affected not just the output but also the quality of innovations. We measure patent quality using a text-based indicator of innovativeness constructed following Kelly et al. (2021) and briefly outlined earlier. This allows us to assess whether WWI mortality lowered the likelihood that communities generated more influential patents, in addition to reducing patenting activity overall.

FIGURE 3
WWI DEATHS AND PATENTING – EVENT-STUDY RESULTS

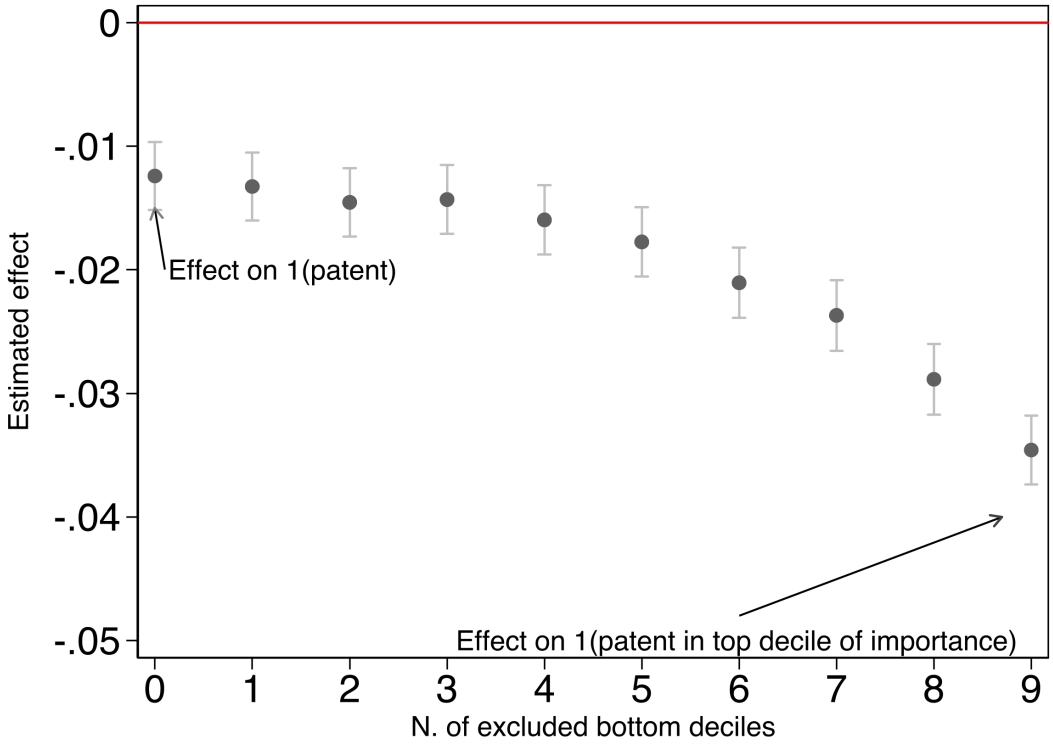


Notes: OLS estimates of the effect of WWI deaths on the probability that at least one patent is registered by a resident of parish i during 5-year period t. Estimates from equation 2, including time, parish, and county-time fixed effects. 95% confidence intervals constructed using standard errors clustered at the parish level.

Figure 4 presents a series of estimates from repeated applications of equation 1, where the dependent variable is progressively restricted to patents above increasing importance thresholds. For each 5-year window, we rank patents by their importance and construct a binary variable indicating whether parish i had at least one patent meeting or exceeding a given threshold during period t. The initial estimate includes all patents, treating any patent as "important". Subsequent iterations successively tighten the criterion: first to only patents in the top 90% of the importance distribution, then the top 80%, and so forth, culminating with an indicator that equals one only if a parish resident published at least one patent in the top 10% of importance within that window.

The results document that the negative impact of the WWI death shock on innovation potential grows with patent importance. In absolute terms, a higher exposure to the WWI mortality shock reduces the probability of publishing at least one patent in the top half of the importance distribution by approximately one-third more than the effect on the probability of publishing a patent of any importance. This relationship is even more pronounced in the right tail of the importance distribution: the effect of the WWI shock on the probability of producing a patent in the top 10 percent of importance is roughly three times as large as the effect on the probability of registering a patent of any importance, suggesting that WWI

FIGURE 4
WWI DEATHS AND PATENTING – IMPORTANT PATENTS



Notes: OLS estimates of the effect of WWI deaths on the probability that at least one patent whose innovativeness (measured as in Kelly et al. (2021)) is in the top deciles of the distribution of patents published at the same time is registered by a resident of parish i during 5-year period t. Each coefficient refers to a different regression, so that the left-most coefficient reflects the estimate reported in Table 1 and the right-most the estimate obtained when using an indicator equal to 1 if at least one patent registered belongs to the 10 percent of the most important patents published during the 5-year period t. Estimates include time, parish, and county-time fixed effects. 95% confidence intervals constructed using standard errors clustered at the parish level.

mortality affected both the quantity and the quality of the innovation produced by British communities.[25]

5.3. Effects on patents by established and new inventors

Highly localised war mortality could affect communities' innovative output in the long run in several ways. A natural starting point, grounded in recent theories of innovation, is to consider implications for the productivity of existing inventors and for the creation of new inventors (Bell et al., 2019*a*; Akcigit et al., 2018). In both cases, mechanical channels are possible. War deaths may directly reduce innovation if inventors are killed. Even when they are not, a large mortality shock could persistently reduce local population, for example if

[25]By construction, the mean of the dependent variable – critical for assessing the relative magnitude of the estimated coefficients – declines with each iteration, as the threshold for a patent being classified as important increases. The fact that we estimate a gradient between effect size and patent importance even without explicitly accounting for this by rescaling the coefficients and the confidence intervals by the mean of the dependent variable in the corresponding estimation sample indicates that, in relative terms, the effect is substantially concentrated among breakthrough patents.

villages are abandoned. To the extent that new inventors scale with population, that would mechanically reduce future innovation.

We explore these channels in Table 2. We define a "new" inventor as a patentee who did not appear in the sample earlier, and a pre-existing inventor as one who registered at least one patent in the pre-WWI period.[26] Column 1 estimates the effect of WWI deaths on patenting by pre-WWI inventors; the effect is large and negative, suggesting that war mortality negatively affected previously active inventors. The magnitude is especially large for pre-existing inventors: relative to a baseline probability of 15 percent, the coefficient implies that doubling the number of WWI deaths reduces patenting by previously active inventors by about 3.7 percentage points, roughly one quarter of the mean. In column 2 we construct our outcome restricting to patents registered by inventors who we can confirm were not killed in WWI,[27] and find that results are broadly unchanged. Columns 3 and 4 show that WWI mortality also reduced patenting by new inventors, and that the results survive controlling for time-varying population. While population may be an outcome and therefore a "bad control", its inclusion does not change our main estimate.

Taken together, the results suggest that the war shock reduced both the productivity of existing inventors and the rate at which parishes produce successful new inventors, for reasons beyond mechanical death or population effects. We explore the mechanisms that underpin these results further in the next sections.

6. Human-capital shocks and the productivity of established inventors

The parish-level results indicate that WWI mortality affected not only the entry of new inventors but also the innovative activity of inventors who were already active before the war. We therefore turn to an inventor-level analysis using our individual panel of inventors already active before WWI and tracking them over the period 1895–1979. The detailed information available in this dataset allows us to examine heterogeneity in the effects of WWI mortality along several dimensions. In particular, we study how these effects vary with inventors' backgrounds and fields of specialisation, and we describe the composition of remaining patenting by field, location, and co-authorship.

6.1. Baseline inventor-level results

We start by estimating the baseline DiD model in equation 3 by OLS. Column 1 of Table 3 reports the results and shows that inventors initially located in higher-mortality parishes are less likely to patent after WWI, in line with our parish-level estimates.[28] In this specification,

[26] In our data we observe the details of patents but not the decisions of individuals to become inventors. This means our variable capturing a patent by a new patentee may incorrectly include patents by inventors who last patented before our sample period starts.

[27] We match inventors to the CWGC WWI fatalities list using the full first name and surname. This approach is conservative, in that we may misclassify some surviving inventors as killed when they share a name with a deceased soldier.

[28] Two related sample-construction concerns arise because we select inventors who patented in 1895–1914 and then follow them in a balanced panel. First, it is possible (although unlikely) that inventors in high-mortality parishes enter the sample earlier or later than those in low-mortality parishes; this would mechanically affect

TABLE 2
WWI DEATHS AND PATENTING – OLS ESTIMATES – NEW AND EXISTING INVENTORS

	(1) Any patent existing inv.	(2) Any patent existing inv.	(3) Any patent new inv.	(4) Any patent new inv.
$Log(d^{WWI})\times$ Post WWI	-0.053***	-0.041***	-0.018***	-0.018***
	(0.002)	(0.002)	(0.001)	(0.001)
Log Population				0.054***
				(0.003)
Sample	Full	Not killed in WWI	Full	Full
Mean dep. var.	0.15	0.13	0.19	0.19
R2	0.46	0.46	0.45	0.45
Observations	162021	162021	162021	162021

Notes: OLS estimates of the effect of WWI deaths on parish-level patenting. Column 1: outcome is an indicator equal to one if at least one patent is registered by a resident of parish i during 5-year period t who was already patenting before 1915. Column 2: same outcome, excluding inventors matched to CWGC fatalities (to assess mechanical death effects). Column 3: outcome is an indicator equal to one if at least one patent is registered by a resident of parish i during 5-year period t who is a "new" inventor (i.e., one who did not appear in the sample earlier). Column 4: same as column 3, adding a control for log parish population. Parish, time, and county-time fixed effects are included in all specifications. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

a 10 percent increase in WWI deaths lowers an inventor's probability of patenting by about 0.012 percentage points, or roughly 0.1 percent relative to the baseline mean. Moving from the 25th to the 75th percentile of the mortality distribution among affected parishes lowers this probability by about 0.22 percentage points, close to 2 percent of the baseline mean.

Importantly, this decline is not driven by the mechanical disappearance of inventors who were themselves killed in the war. In column 2, we restrict the sample to patentees who were not killed in WWI.[29] The estimated effect remains negative and is larger in magnitude: the same 25th-to-75th comparison lowers the probability of patenting by about 0.33 percentage points, about 2.5 percent of the mean. Thus, WWI mortality reduced the subsequent innovative activity of surviving pre-war inventors, indicating that the consequences of the local human-capital shock extended beyond the direct loss of inventors themselves. This finding is consistent with recent evidence that worker deaths can have important effects on surviving coworkers, reflecting limited substitutability between workers and substantial replacement frictions (Jäger, Heining and Lazarus 2026).[30]

pre-war patenting trends. Reassuringly, this is not the case in our data: the distribution of first-patent years is very similar across high- and low-mortality areas (Appendix Figure A.5), and the year of first patent and the probability of entering only in 1910–1914 are essentially unrelated to parish WWI mortality. Second, inventors who first appear late in the pre-war window contribute with mechanical zeros in earlier periods even before they are observed as inventors. Results are robust to restricting each inventor to periods at or after their first patent (Appendix Table A.7, column 2, and Appendix Figure A.6).

[29]To identify these, we match patentees to the CWGC WWI fatalities list using the full first name and surname. This approach is conservative, in that it may incorrectly classify some surviving inventors as having been killed when they share a name with a deceased soldier. Alternative matching methods (e.g., using only the initial or the first few letters of the first name) yield very similar results; see Appendix Table A.6.

[30]A further concern is that pre-trend tests in the full inventor sample are mostly informative for inventors who entered early enough to have a meaningful pre-war patenting record after their first patent. To address this, we

TABLE 3
WWI DEATHS AND PATENTING – INVENTOR-LEVEL ESTIMATES

	(1) Any patent	(2) Any patent	(3) Any breakthr. patent	(4) Any highly cited patent
$Log(d^{WWI})\times$ Post WWI	-0.001***	-0.002***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Sample	Full	Not killed in WWI	Full	Full
Mean dep. var.	0.12	0.13	0.03	0.02
R2	0.60	0.58	0.19	0.16
Observations	924460	489532	924460	924460

Notes: OLS estimates of the effect of WWI deaths on an indicator equal to one if inventor j from parish i registers a patent during 5-year period t. Column 1 presents results from equation 3 for the full sample, whereas in column 2 the sample is restricted to individuals who do not appear in the CWGC dataset of WWI fatalities, matched using the full first name and surname. In columns 3 and 4, the outcomes are, respectively, an indicator for the inventor registering a "breakthrough" patent or a highly-cited patent (see text for details). Individual, time, and year of first patent by time fixed effects are included in all specifications. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

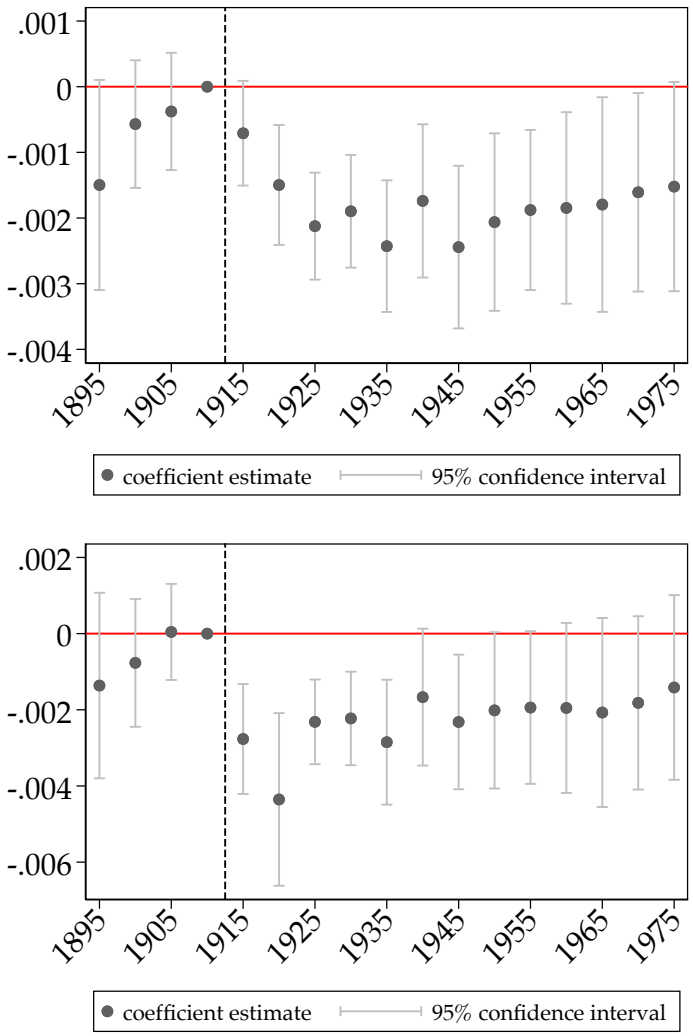
Figure 5 shows the corresponding event-study estimates for our main outcome, patenting of any type, for both the full sample of pre-war inventors and the sample excluding inventors who were killed in WWI. Following WWI, the estimates are negative and remain below zero throughout the post-war period. The pattern is very similar when inventors who were killed in WWI are excluded, providing further evidence that the decline in innovative activity is not mechanically driven by inventor deaths. The inventor-level dynamics also differ in an informative way from the parish-level event study in Figure 3. In particular, the inventor-level estimates do not exhibit the renewed decline around WWII that is visible in the parish-level results. This suggests that the persistent reduction in patenting among pre-war inventors is not driven by places with high WWI mortality also being systematically more exposed to wartime disruptions during WWII.

We next examine whether the decline in patenting also extends to particularly important innovations. In column 3 of Table 3, the outcome is an indicator for registering a "breakthrough" patent, defined as one whose importance (calculated following Kelly et al. 2021) exceeds the 80th percentile of the distribution of all patents registered in the past 5 years. In column 4, we use a citation-based measure of importance, defining a highly cited patent as one that receives more citations than 80 percent of all patents registered in the past 5 years.[31] Under both measures, we find that inventors in higher-mortality areas are less

re-estimate the baseline specification on inventors whose first patent is in 1909 or earlier, or in 1904 or earlier (Appendix Table A.7, columns 3 and 4). In both restricted samples, the estimated effect remains negative, statistically significant, and comparable in magnitude to the baseline. The corresponding event studies are similar and are omitted for brevity.

[31] This measure is imperfect because citation counts in PATSTAT have been systematically recorded only since the 1980s, so highly influential patents that lost relevance before then may go undetected.

Figure 5
Inventor-level event study



Notes: Inventor-level OLS estimates of the effect of WWI deaths on the probability that at least one patent is registered during 5-year period t. Estimates from equation 3. Individual, time, and year of first patent by time fixed effects are included in all specifications. 95% confidence intervals constructed using standard errors clustered at the parish level. Upper panel: all inventors. Lower panel: inventors who were not killed in WWI (defined as those we could not match to the CWGC dataset of WWI fatalities using the full first name and surname).

likely to produce highly innovative patents.[32] Compared with the sample mean of each outcome, the point estimates in columns 3 and 4 imply that the decline in high-quality patenting is proportionally larger than the decline in patenting of any type (column 1). A doubling of the number of parish deaths causes a reduction of about 2 percent relative to the mean for both outcomes, compared with a relative effect of less than 1 percent that we obtain when focusing on patenting of any type.

[32]Using one- or ten-year windows instead of 5 years yields similar results, which we omit for brevity.

6.2. Heterogeneous effects across inventors and fields

By concentrating on inventors who were already active before WWI, we can identify which fields of knowledge experienced a greater impact from the mortality shock. To do this, we estimate an extended version of equation (3) in which the $log(d_i^{WWI}) \times Post$ term is interacted with a series of indicators, one for each inventor's pre-war field of specialisation, measured by the one-digit CPC code of the first patent they produced in the pre-war period. This approach enables us to break down the average effect of WWI mortality by field, revealing how the shock differentially affected areas that require different skills and expertise to generate innovation.

The results are presented in Figure 6, which reports the estimated coefficients for each interaction term. Results show substantial heterogeneity across fields, with the largest effects in electricity, mechanics, transportation, physics and chemistry. These science-intensive sectors, at the forefront of innovation at the time, required highly trained scientists and the execution of increasingly complex tasks. The need to commercialise scientific breakthroughs made laboratory research, specialised skills, and coordinated teams and infrastructure essential (Edgerton and Horrocks, 1994).[33]

Figure 7 examines heterogeneity in the effects of mortality on patenting by different measures of inventor characteristics, patent complexity, and patent importance. The figure illustrates that the negative effect of WWI deaths on patenting is not uniform: it is significantly more pronounced among inventors whose early work marked them as being specialised in producing more complex innovation.

We start by contrasting inventors who reported engineering as their occupation in at least one patent registered before WWI with those who did not. Results indicate that engineers experienced a notably larger negative impact from mortality, underscoring that the loss of human capital due to the War was particularly damaging to highly skilled inventors.

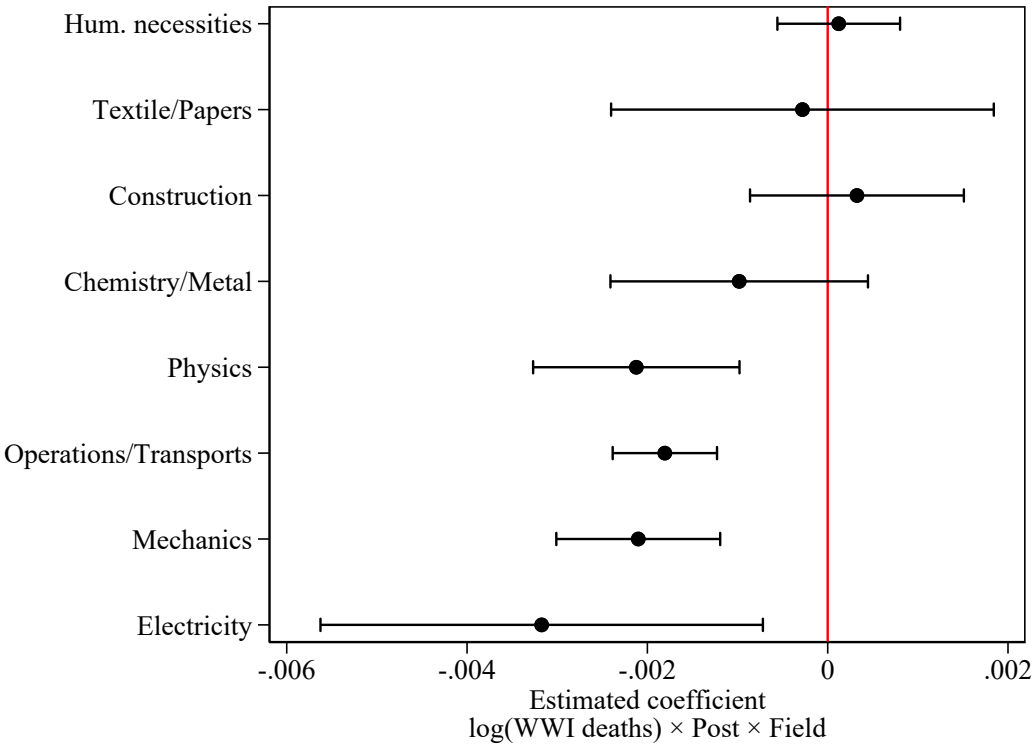
In addition, we examine heterogeneity by dividing inventors according to different characteristics of their first pre-war patent. Specifically, inventors whose initial patent falls above the median in patent breadth, abstract length, or number of sentences experience greater declines in subsequent patenting activity. For comparison, we also report heterogeneity by patent importance, measured using the metric from Kelly et al. 2021, which captures a distinct dimension of patent quality.[34] The final coefficients reported in the figure show the results obtained by dividing the sample based on the first principal component of the engineer indicator, patent breadth, abstract length, and number of sentences, providing a summary index of these dimensions. These estimates reiterate that the negative impact of the mortality shock is most pronounced among inventors whose early work was more technologically

[33]These sectors were also those in which engineers were most active, as shown in Appendix Figure A.8. Appendix Figure A.9 reports scatter plots of the share of engineers among inventors against a proxy for technological complexity, the length of the patent abstract (panel A), and patent importance, computed following Kelly et al. (2021) (panel B). In both cases, we observe a strong positive relationship: patents that are more complex and more influential were disproportionately produced by engineers.

[34]Throughout our study period, the length of patent abstracts was not regulated, resulting in considerable variation.

FIGURE 6

HETEROGENEOUS EFFECTS BY FIELD OF SPECIALISATION



Notes: Inventor-level OLS estimates of the effect of WWI deaths on an indicator equal to one if at least one patent is registered during 5-year period t, interacted with the field of specialisation of each inventor, as proxied by the field of specialisation of the first patent produced by the inventor in the pre-war period. Estimates from equation 3 where the term $log(d_i^{WWI}) \times Post$ has been interacted with each field, as reported on the vertical axis. Individual, time, and year of first patent by time fixed effects are included. 95% confidence intervals constructed using standard errors clustered at the parish level. Fields of specialisation are assigned according to the NLP technique explained in Appendix D.

complex.

Taken together, these results indicate that the decline in innovation after WWI was disproportionately concentrated among inventors working in science- and engineering-intensive fields and among those whose pre-war inventions were more innovative or complex. This pattern is consistent with human-capital losses being especially consequential for forms of innovation that depend more heavily on specialised skills and complementary inputs.

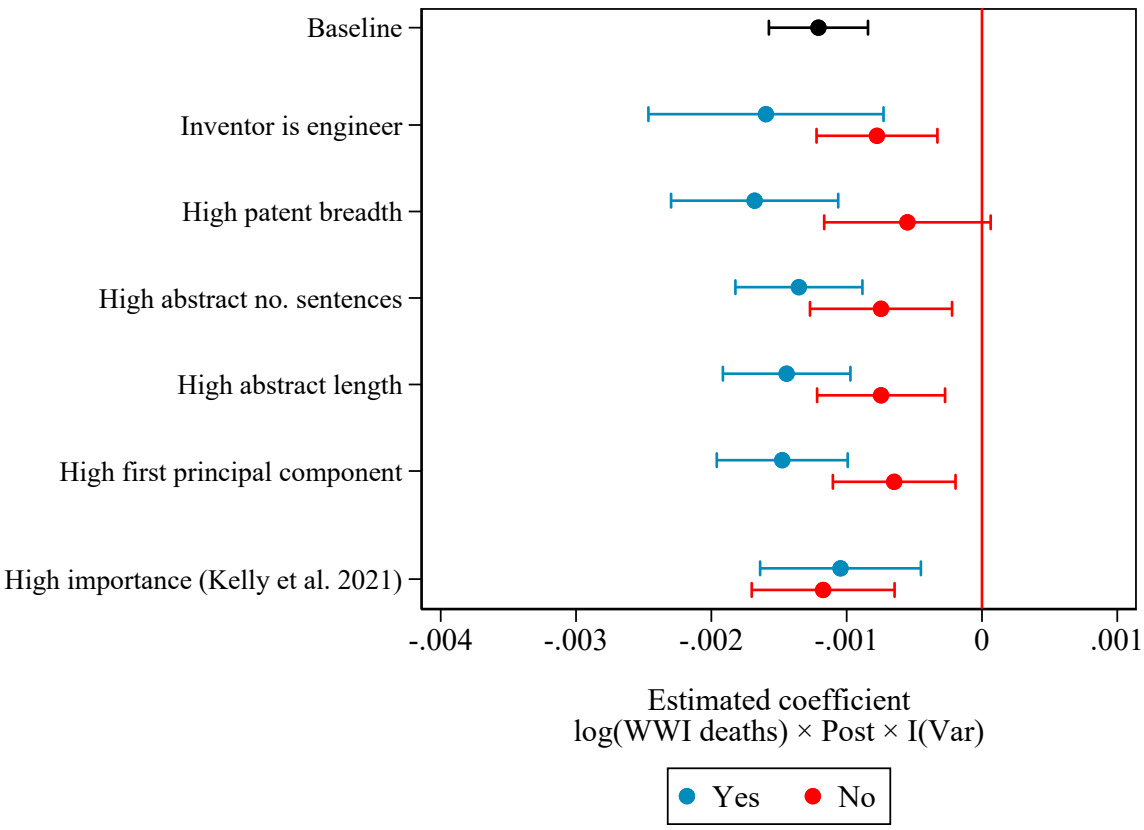
6.3. The effect of mortality on the composition of patenting activity

The persistent decline in patenting among surviving pre-war inventors suggests that WWI mortality affected more than the direct survival of inventors themselves. Local human-capital losses may also have reduced the productivity of established inventors by disrupting access to complementary workers, collaborators, and other inputs into the innovation process. We therefore examine whether mortality changed the type of patenting activity of established inventors in terms of field, location, or co-authorship.

Figure 8 presents the results. In each column, we redefine the dependent variable to

Figure 7

Heterogeneous effects by patent complexity



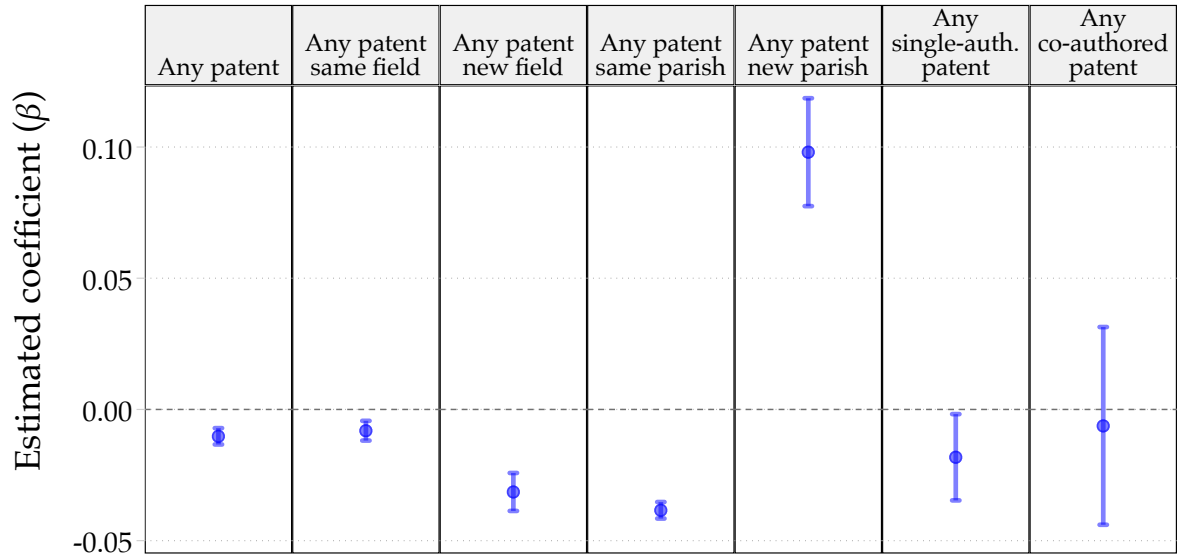
Notes: This figure presents inventor-level OLS estimates of the impact of WWI deaths on patenting outcomes, examining how effects vary according to characteristics of the inventor and of the inventor's first pre-war patent (see main text for details). The first row displays the overall (uninteracted) effect of WWI mortality. The following rows show estimated effects separately for inventors who reported being engineers in at least one patent registered before WWI versus non-engineers. Additional rows present effects for inventors whose first patent was of high or low complexity, defined as above or below the median based on measures such as patent breadth (defined in Appendix E), abstract length or sentence count. The figure further distinguishes between high and low patent importance, and includes a summary using the first principal component of the engineer indicator, patent breadth, abstract length, and sentence count. The outcomes are binary indicators for having patented in a given period, and coefficients are reported for each interaction term. Individual, time, and year of first patent by time fixed effects are included in all specifications. 95% confidence intervals constructed using standard errors clustered at the parish level.

capture different types of patenting activity.[35] Column 1 reports our baseline estimate for comparison. In column 2, we redefine the patenting indicator to take value 1 if the inventor registers at least one patent within the inventor's original field, as defined by the field of the inventor's first pre-WWI patent, while in column 3 the dependent variable is equal to 1 if the inventor registers at least one patent in a different field. Both estimates are negative and similar to the baseline effect, suggesting that remaining patenting does not shift into new fields.

[35]To facilitate interpretation across columns, all coefficients are rescaled by the corresponding sample mean of the outcome. See Appendix Table A.8 for the full, unrescaled estimates.

Figure 8

Inventor-level estimates – field, location, and co-authorship



Notes: $N = 924,460$. Inventor-level OLS estimates of the effect of WWI deaths on patenting from eq. 3. All coefficients are rescaled by the corresponding sample mean of the outcome to facilitate comparability across specifications. See Appendix Table A.8 for unrescaled estimates. Column 1 reports the baseline estimate (analogous to column 1 of Table 3) for comparison. In column 2 (resp., 3), the dependent variable is an indicator taking the value 1 if inventor j registers at least one patent during 5-year period t within the same sector (resp., within a different sector) as the one the inventor first patented in. In column 4 (resp., 5), the dependent variable is an indicator taking the value 1 if inventor j registers at least one patent during 5-year period t in the same parish (resp., in a different parish) as the one in which they registered their first patent. Finally, in column 6 (resp., 7), the dependent variable is an indicator taking the value 1 if inventor j registers at least one single-authored (resp., co-authored) patent during 5-year period t. Notice that inventor j in the same 5-year period t can produce patents within different fields, in multiple locations, and with different co-authorship teams. Individual fixed effects, time fixed effects, and year of first patent by time fixed effects are included in all specifications. 95% confidence intervals constructed using standard errors clustered at the parish level.

In columns 4 and 5, we perform a similar decomposition based on location. In column 4, the dependent variable takes value 1 if the inventor registers at least one patent while residing in the same parish as when registering their first patent. In column 5, instead, the dependent variable is equal to 1 if the inventor registers at least one patent upon moving to a different parish. Here we observe a clear pattern: higher WWI mortality reduces the probability of patenting in the original parish but raises the probability of patenting in a different parish.

Importantly, relocation cannot mechanically account for the negative baseline inventor-level estimate. Inventors' exposure to WWI mortality is fixed at their pre-war location, and subsequent patents enter the baseline outcome irrespective of where they are registered. An inventor who leaves a high-mortality parish therefore remains assigned to that parish's mortality, and any patents produced after the move still count as patenting.

Finally, in columns 6 and 7, the dependent variables take value 1 if the inventor registers at least one single-authored patent and at least one co-authored patent, respectively.

The negative mortality effect on single-authored patenting closely follows the baseline estimate, whereas the effect on co-authored patenting is statistically indistinguishable from zero. Complementary workers and team-specific human capital are central to inventive activity, and their loss can reduce the productivity of remaining inventors (Jaravel, Petkova and Bell 2018; Jäger, Heining and Lazarus 2026). The pattern in columns 6–7 is consistent with collaboration providing some insulation from local human-capital losses. However, it is also consistent with a shift in the composition of remaining patents – for example if those who stop patenting were predominantly working alone, or if some patents that would have been filed alone are now filed with a co-author. Our estimates cannot distinguish between those interpretations.

Taken together, WWI mortality changes the location and the form of remaining patents. Inventors from higher-mortality parishes become less likely to patent in their original location and more likely to patent from a different parish, and the decline is concentrated in single-authored rather than co-authored patents. Switching fields does not show an analogous effect.

7. Human-capital shocks and the supply of new inventors

In this section, we return to the parish-level analysis to explore how human-capital shocks affect the local supply of new inventors. As shown earlier in Table 2, patenting by new inventors declined at the parish level in response to exposure to WWI mortality and this effect is not driven by changes in the scale of local population. The evolution of this outcome measure may capture either entry into and exit from innovative activity or changes in the productivity of researchers who have yet to file a patent. As such, we can use it to shed light on mechanisms that affect at least one of these phenomena, although typically we will be unable to distinguish between them. We start by examining how the wartime loss of social elites and skilled engineers affects the cultivation of successful new inventors. We then explore the extent to which pre-existing institutions and innovation ecosystems can mitigate the effects of mortality on the same outcome.

7.1. The importance of social elites and engineers

As previously discussed, a very high proportion of inventors in Britain in the first half of the twentieth century classified themselves as engineers. Although we do not have any data to confirm it, it also seems possible that many of those who would go on to become inventors had attended university. We therefore expect that losing social elites and men with engineering skills could reduce the supply of future inventors. Besides this, we may also expect that engineers could serve as role models for potential future innovators (Bell et al., 2019*b*). To test the extent to which the loss of talented and skilled individuals affects the production of new inventors, we therefore extend our baseline model of eq. 1 (as estimated in column 3 of Table 1) as follows:

$$y_{it}^{New} = \alpha + \beta_1 \mathbb{1}(d_i^{WWI} > med) \times Post + \beta_2 \mathbb{1}(HC_i) \times Post + FE + \epsilon_{it}, \tag{4}$$

where $\mathbb{1}(HC_i)$ is an indicator variable for human capital that takes the value one when the parish experienced the death of at least one individual with skills and characteristics associated with the generation of new ideas. We will define this indicator in different ways using the information contained in the mortality records. To match the nature of these variables, we include total WWI deaths in this regression using an indicator for above-median value. The coefficient of interest is β_2, which measures the effect on innovation of the loss of these skills, conditional on the parish being a high or low death location.

Results are reported in Table 4. In column 1, we define $\mathbb{1}(HC_i)$ as an indicator for the parish having lost at least one individual belonging to the social elite, defined as those who studied at Oxford or Cambridge, or an elite school. The coefficient is negative and significant and indicates that the loss of a member of the social elite reduces the likelihood of subsequent patenting by more than 3 percentage points. In columns 2 and 3, we examine the effect of deaths of officers. In column 2, we replace the social elite indicator with an indicator for the death of at least one officer, finding again a negative effect. In column 3, we divide officers into two groups, engineers and other officers. The results indicate that the loss of an officer with an engineering background has a much larger negative effect on subsequent patenting than the loss of other types of officers, for which the effect is small and insignificant.

In column 4, we add back our indicator for losing a social elite and show that both social-elite and engineer deaths matter. The loss of these groups thus has effects on subsequent patenting over and above the loss of other men. The estimate on engineers is in line with recent evidence from the UK and the Americas showing that engineers are key drivers of innovation and growth and produce more patents than other types of innovators (Hanlon, 2025; Maloney and Valencia Caicedo, 2022). In column 5, we address the concern that the results might merely reflect pre-war differences in the local supply of engineers and white-collar workers by adding interactions between the $Post$ dummy and (i) the share of adult men in engineering occupations in 1911 and (ii) the share of adult men in white-collar occupations in 1911. The estimates are unchanged, and the coefficient on overall WWI mortality remains precisely estimated. This suggests that our mortality measure captures both the loss of skilled men that we observe in the data and unobserved variation in the loss of talented men.

7.2. Heterogeneity and the mitigating role of institutions

Previous work demonstrates that the presence of other inventors, transport links, communications technologies, and institutions can drive the generation of new ideas (see e.g., Akcigit, Grigsby and Nicholas 2017*b*; Bell et al. 2019*b*; Andrews 2023). Local access to these resources in the period prior to WWI could potentially mitigate the effects of war deaths on innovation. Specifically, if these ecosystem factors can substitute for human-capital inputs in the production of new ideas, we may expect to find that the effect of war deaths is smaller in places where they are more accessible. We conduct heterogeneity analysis to test this hypothesis using four county-level accessibility measures and our parish-level data. In all cases, we separate out counties with above-median access to an ecosystem factor, and test for differential effects relative to the overall average effect by adding a triple interaction term to

TABLE 4
WWI DEATHS AND PATENTING – OLS ESTIMATES – IMPORTANCE OF ELITES

	Any patent by new inventor				
	(1)	(2)	(3)	(4)	(5)
I(Social elite) x Post	-0.033***			-0.017*	-0.017*
	(0.009)			(0.010)	(0.010)
I(Officer) x Post		-0.010**			
		(0.004)			
I(Engineer off.) x Post			-0.051***	-0.047***	-0.045***
			(0.008)	(0.008)	(0.008)
I(Non eng off) x Post			-0.006		
			(0.004)		
$I(d^{WWI} > med)\times$ Post	-0.032***	-0.030***	-0.023***	-0.024***	-0.024***
	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)
Mean dep. var.	0.19	0.19	0.19	0.19	0.19
R2	0.45	0.45	0.45	0.45	0.45
Observations	162021	162021	162021	162021	160950
1911 Occupation x Post controls	N	N	N	N	Y

Notes: OLS estimates of the effect of WWI deaths on the probability that at least one patent is registered by "new" inventors (patentees who did not appear in the sample earlier) resident of parish i during 5-year period t. Social elite is an indicator for losing one or more individuals who studied at Oxford or Cambridge, or at an elite school. Similarly, officer is an indicator for losing one or more officers. Engineer is an indicator for losing one or more officers with engineering background. Parish, time, and county-time fixed effects are included in all specifications. Column 5 additionally controls for $Post$ interacted with i) share of adult men in engineering occupations in 1911 and ii) share of adult men in white-collar occupations in 1911. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

our baseline specification. Note that the county-time effects in this specification absorb the two-way interaction between above median accessibility and post WWI so we are effectively conducting comparisons strictly within places with comparable baseline accessibility to each ecosystem factor. In addition, to absorb the effects of density in general we also control for post times above-median accessibility measure constructed using distance to churches.

Results are reported in Table 5. In column 1, we find that the effect of war deaths is significantly smaller – approximately half as large in absolute terms – in counties with above-median access to engineering role models (proxied by members of engineering institutes). In column 2, we examine proximity to universities and find a similar pattern. However, in columns 3 and 4 we find there is little or no effect of proximity to rail stations and post offices. Notably, across all columns and as expected, we find that the mean of the dependent variable is considerably larger in places with better access to ecosystem factors, consistent with the idea that these factors are positively associated with patenting. Overall, these results indicate that the effects of war deaths are in fact considerably smaller in some places, and in particular suggest that physical proximity to universities and networks of experts can mitigate the effects of shocks, at least to some degree.

To conclude, this section shows that the impact of war deaths on innovation is amplified by the loss of talented and skilled individuals, and mitigated by the presence of institutions

TABLE 5
WWI DEATHS AND PATENTING – OLS ESTIMATES – HETEROGENEITY

	Any patent by new inventor			
	(1)	(2)	(3)	(4)
$Log(d^{WWI})\times$ Post WWI	-0.025***	-0.028***	-0.024***	-0.023***
	(0.002)	(0.002)	(0.002)	(0.002)
x I(Access $>$ med)	0.005**	0.008***	0.002	0.000
	(0.003)	(0.002)	(0.002)	(0.002)
Access factor	Engineers	Univer.	Stations	Post Off.
Mean dep.var.(Access $<$ med)	0.12	0.13	0.12	0.16
Mean dep.var.(Access $>$ med)	0.25	0.24	0.26	0.22
R2	0.45	0.45	0.45	0.45
Observations	162021	162021	162021	162021

Notes: OLS estimates of the effect of WWI deaths on the probability that at least one patent is registered by "new" inventors (patentees who did not appear in the sample earlier) resident of parish i during 5-year period t, by pre-WWI characteristics defined at the county level. Parish, time, and county-time fixed effects are included in all specifications. In all cases we additionally control for the interaction between post and above median access to churches. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

and ecosystem factors that can substitute for or sustain human capital in the production of new ideas. These results suggest that the loss of talented and technically skilled servicemen had persistent local effects: communities lost not only potential future inventors and role models, but also part of the human-capital base that supports the production of new ideas. While this evidence is indirect, it is consistent with the idea that the loss of talented men in the war affected the production of talented individuals in the next generation, by reducing the local supply of innovators and rendering these places less attractive for new inventors.

8. Robustness checks

8.1. Strong parallel trends

Difference-in-differences designs that use a continuous treatment variable require a stronger version of the parallel-trends assumption (Callaway, Goodman-Bacon and Sant'Anna, 2024), in which the average evolution of patenting if all parishes had experienced a given mortality intensity equals the path actually followed by parishes that experienced that intensity. Although we cannot test this assumption directly, we can assess its plausibility by examining outcome trends across the distribution of mortality, before and after treatment. To do so, we first divide the sample into deciles of the distribution of $\log(d_i^{WWI})$ and then estimate a separate event-study specification for each pair of deciles (i, z) with $i < z \leq 10$, yielding 45 pairwise comparisons. In each specification, the binary treatment indicator equals one for observations in the higher decile of the pair (z). For each regression, we formally test for the statistical significance of the pre-war interaction coefficients and save the resulting p-value. Appendix Figure A.7 presents the distribution of these p-values. Only one of the 45 regressions shows a statistically significant imbalance in pre-treatment coefficients ($p = 0.048$).

These tests provide no evidence of differential pre-treatment trends across mortality deciles, thus reassuring us that the strong parallel trends assumption is likely to hold.

8.2. IV results

In this subsection, we discuss an alternative instrumental variable strategy that does not rely on any of the parallel trends assumptions needed for the difference-in-differences approach. Specifically, we instrument the log of WWI deaths with the log of predicted deaths based on battalion-level mortality rates. The instrument is constructed by aggregating battalion-level mortality to the parish level, using the share of a parish's servicemen assigned to each battalion as weights. The intuition behind this approach is that during WWI, servicemen were assigned to battalions – fighting units roughly 1,000 strong – in a sequential manner based on enlistment order (Bet-El, 2009; Carozzi, Pinchbeck and Repetto, forthcoming). This sequential assignment is plausibly exogenous, as it is driven by the order of arrival rather than by underlying parish characteristics. Consequently, serving in one battalion rather than in the next is unlikely to be correlated with pre-existing determinants of innovation at the parish level.[36]

Specifically, indexing battalions with b and parishes with i, we instrument $\log(d_i^{WWI})$ with

$$z_i = \log\left(m_i \sum_{b=1}^{B} \alpha_{ib}\tilde{\delta}_{ib}\right),$$

where m_i is mobilisation from parish i; $\alpha_{ib} = m_{ib}/m_i$ is the share of parish i's mobilised who served in battalion b (with m_{ib} mobilised from i in b); d_b and m_b are deaths and mobilisation in battalion b; and $\tilde{\delta}_{ib} = \frac{d_b - d_{ib}}{m_b - m_{ib}}$ is battalion b's death rate excluding parish i (leave-out mean), with d_{ib} denoting deaths from parish i in battalion b.

This instrument has a shift-share structure, with shares α_{ib} and shocks δ_b, estimated by the leave-out means $\tilde{\delta}_{ib}$. Because, in this setting, variation in mortality across battalions is likely to be driven mostly by where battalions were deployed and by the fortunes of war, we follow the approach that assumes shocks are exogenous (Borusyak, Hull and Jaravel, 2022). Formally, our identification assumption is that our measures of battalion-level mortalities are conditionally uncorrelated with other parish-level determinants of patenting activity.[37]

In Appendix Figure A.10, we provide evidence in favour of the instrument orthogonality assumption. In panel A, we regress the instrument on several parish-level pre-WWI characteristics.[38] Most coefficient estimates are close to zero and statistically insignificant at conventional levels, indicating that the instrument is not correlated with observable characteristics that could affect patenting activity. In panel B of Appendix Figure A.10, we conduct

[36] Construction of the instrument is possible only for battalions and parishes that appear in both the CWGC-FWR data and the FamilySearch data. As a consequence, the sample size is reduced by about one third in the IV results. For additional technical details on the instrument's construction, we refer the reader to Carozzi, Pinchbeck and Repetto (forthcoming).

[37] This assumption is weaker than requiring random assignment of servicemen to battalions. In fact, the necessary and sufficient condition for instrument exogeneity is that the death rates δ_{ib} not be systematically related to any other pre-WWI parish-level characteristic that determines present or future innovation.

[38] All specifications include the logarithms of 1911 population and WWI mobilisation, and historic county dummies.

a similar exercise after aggregating the data at the battalion level, using the shares of servicemen from each parish serving in a battalion as weights (Borusyak, Hull and Jaravel, 2022).[39] This amounts to regressing battalion-level mortality shocks (δ_{ib}) on the economic, demographic or geographical characteristics of the community of origin of soldiers that make up these battalions. Once again, we find evidence in favour of the orthogonality assumption. Battalion-level death rates are uncorrelated with pre-determined parish level characteristics.

We confirm the relevance of the instrument for our parish-level analysis in Panel A of Table 6, which shows large first-stage coefficients across all four specifications that mirror those presented in Table 1. F-statistics are well above 1000 in all specifications, suggesting the instrument is strong throughout. [40]

Panel B of Table 6 presents the main parish-level IV estimates. Across the four specifications, the estimated effects of WWI deaths are very similar to the corresponding OLS estimates reported in Table 1, but slightly larger in magnitude. The corresponding IV event-study estimates are reported in Appendix Figure A.11. They are qualitatively similar to their OLS counterpart in Figure 3, with coefficients close to zero before WWI and negative estimates throughout most of the post-war period. Taken together, the IV results provide reassurance that endogeneity of WWI deaths is not a primary concern for our main findings. The instrumental-variable estimates for the inventor-level analysis also confirm the OLS difference-in-differences results and are reported in Appendix Table A.9.

8.3. Additional robustness checks

We next summarise additional robustness checks that address several potential concerns. First, the results in Table 1 and Figure 3 focus on a binary measure of innovation, because most of the variation in patenting is between parishes that patent and those that do not. However, our parish-level findings are not confined to the extensive margin. In Appendix Figure A.12 we estimate a series of regressions where the dependent variable is an indicator for whether the parish patent rate (number of patents per capita, using the 1911 Census population) exceeds various percentiles of the overall distribution. Across almost all percentiles, higher WWI mortality reduces the probability that a parish reaches higher patenting rates.[41] Notably, the negative effects on being above the 80th or 90th percentile are larger than the effect at the extensive margin, suggesting that WWI mortality had a sizeable impact not just on whether parishes patent at all, but also on the intensity of innovation (the intensive margin).

Second, Appendix Tables A.2–A.4 show that our baseline results are robust to alternative

[39] This aggregation is performed using the *ssaggregate* command in Stata 19.

[40] The identification result in Borusyak, Hull and Jaravel (2022) also requires that the number of shocks be sufficiently large and that they be sufficiently dispersed in terms of their average exposure. We follow their recommendation and report the inverse of the Herfindahl Index of shock-level average exposure as a measure of the effective sample size. In our dataset this equals 141, indicating that shocks are well dispersed and that the asymptotic results in that paper apply.

[41] For percentiles below the 77th, these coefficients are equivalent to those in column 1 of Table 1, because the probability of exceeding such percentiles coincides with the probability of registering at least one patent.

TABLE 6

WWI DEATHS AND PATENTING – PARISH-LEVEL IV ESTIMATES

	(1) $Log(d^{WWI})$ × Post	(2) $Log(d^{WWI})$ × Post	(3) $Log(d^{WWI})$ × Post	(4) $Log(d^{WWI})$ × Post	(5) $Log(d^{WWI})$ × Post
A. First-stage					
$z\times$ Post	0.926***	0.924***	0.911***	0.799***	0.538***
	(0.007)	(0.007)	(0.007)	(0.009)	(0.013)
F-stat	19906	18936	16330	8208	1675
Observations	111239	111239	111239	111239	110598
	Any patent	Any patent	Any patent	Any patent	Any patent
B. IV (second stage)					
$Log(d^{WWI})\times$ Post	-0.014***	-0.014***	-0.015***	-0.013***	-0.014**
	(0.002)	(0.002)	(0.002)	(0.003)	(0.006)
$Log(d^{County})\times$ Post		0.000			
		(0.003)			
Mean dep.var.	0.31	0.31	0.31	0.31	0.31
Observations	111239	111239	111239	111239	111239
Parish FE	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y
County × Time	N	N	Y	Y	Y
Controls × Linear trend	N	N	N	Y	N
County × Time $\times Q_n$ Pop	N	N	N	N	Y

Notes: Panel A: IV first-stage estimates of the log number of WWI deaths on the instrument, both interacted with an indicator for time periods after 1915-19. Panel B: IV estimates of the effect of WWI deaths on the probability that at least one patent is registered by a resident of parish i during 5-year period t. Specifications follow those in Table 1. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

outcome measures, including the total number of patents and patents per 1,000 inhabitants (based on the 1911 Census population). Panel B replaces d^{WWI} with the parish-level death rate, defined as the ratio of d^{WWI} to the 1911 population, and shows that results in rates are similar but slightly larger in magnitude for the probability of any patenting and for patent counts. A one-percentage-point increase in the parish death rate is associated with an approximately 0.9–1.3 percentage point reduction in the probability of producing any patent. Panel C instead defines treatment as an indicator for mortality in the parish being above the median, $I(d^{WWI} > med)$, and estimates a binary difference-in-differences specification. These tables also report Poisson pseudo maximum likelihood estimates, following Silva and Tenreyro (2006), which are well suited to count outcomes with a large share of zeros (Chen and Roth, 2023). Across the raw, winsorized, and trimmed specifications, the estimates remain consistent with our baseline results using the binary outcome.

We also address the possibility that our parish-of-origin measure combines residence- and birthplace-based assignment in a way that introduces measurement error. In Appendix

Table A.5, we re-estimate the baseline specification separately using deaths assigned by parish of residence and deaths assigned by parish of birth. The estimates are similar across the two measures, suggesting that our results are not driven by the fallback from residence to birthplace in the construction of parish-level mortality.

Third, our baseline treatment uses the log of WWI deaths, which excludes parishes with zero fatalities. Appendix Table A.10 and Table A.11 re-define treatment as $\log(1 + \text{deaths})$, thereby retaining zero-death parishes (and inventors active there). The results are unchanged.

Fourth, London accounts for a disproportionate share of both WWI deaths and patents. Because London registers patents in every period, it does not contribute to the identification of the parish fixed-effects specification with a binary outcome, but it could affect the estimates through the county-time effects and the controls. Appendix Table A.12 shows that the estimates are unchanged when we exclude London (Panel A) or all 35 parishes with a 1911 population above 100,000 inhabitants (Panel B).

Fifth, the main analysis aggregates outcomes to 5-year periods. Appendix Figures A.13 (parish-level) and A.14 (inventor-level) replicate the event-study specifications using two-year intervals, and the qualitative pattern mirrors the baseline.

Lastly, in Appendix Table A.13 and Table A.14, we restrict the sample to periods before World War II (WWII) to rule out the possibility that our results are driven by the effect of WWI mortality on WWII mortality (Carozzi, Pinchbeck and Repetto, forthcoming). Consistent with our dynamic specifications, which show that the effect of WWI mortality on innovation materialises shortly after the war, the results are unchanged under this restriction.

9. Conclusion

More than 750,000 British servicemen lost their lives in World War I. The losses varied substantially across space and fell disproportionately on young men, including substantial numbers of highly educated and technically trained individuals, giving rise to the notion of a British "lost generation." We study the long-run consequences of this human-capital loss for innovation. Linking parish-level military deaths to patent data, we find that communities with higher mortality experienced persistent declines in inventive activity. These declines were larger for breakthrough and technologically complex innovation, indicating that the shock affected not only the quantity but also the characteristics of local inventions. Importantly, the effect does not merely reflect the loss of existing inventors during the war. Exposure to higher mortality also reduces both the subsequent patenting of surviving pre-war inventors and the emergence of new inventors.

Our results are consistent with a disruption to the local supply of ideas. Conditional on overall mortality, patenting by new inventors fell more in parishes that lost engineers or alumni of elite educational institutions, providing evidence in favour of the "lost generation" narrative. Conversely, the decline was smaller in places with stronger pre-existing innovation ecosystems, which prior work has shown to matter for the production of inventors. Our inventor-level results also point to several margins of adjustment: patenting falls in in-

ventors' original parishes while rising elsewhere, and the decline is concentrated in single-authored rather than co-authored patents. By contrast, patenting falls both within and outside inventors' original technological fields. Together, these patterns highlight the role of local complementarities and innovation networks in shaping the consequences of human-capital losses.

Our findings speak to and complement recent studies showing that shocks that create labour scarcity at broader regional scales can stimulate labour-saving innovation. We highlight a distinct margin operating at much finer spatial scales: local mortality shocks that disproportionately deplete innovation-relevant human capital can depress innovative capacity for decades.

References

Abramitzky, Ran, Adeline Delavande, and Luis Vasconcelos. 2011. "Marrying up: the role of sex ratio in assortative matching." *American Economic Journal: Applied Economics*, 3(3): 124–57.

Acemoglu, Daron. 2002. "Directed Technical Change." *Review of Economic Studies*, 69(4): 781–809.

Acemoglu, Daron. 2007. "Equilibrium bias of technology." *Econometrica*, 75(5): 1371–1409.

Aghion, Philippe, and Peter Howitt. 1992. "A Model of Growth Through Creative Destruction." *Econometrica*, 60(2): 323–351.

Akcigit, Ufuk, John Grigsby, and Tom Nicholas. 2017*a*. "Immigration and the rise of American ingenuity." *American Economic Review*, 107(5): 327–331.

Akcigit, Ufuk, John Grigsby, and Tom Nicholas. 2017*b*. "The rise of american ingenuity: Innovation and inventors of the golden age." National Bureau of Economic Research.

Akcigit, Ufuk, Santiago Caicedo, Ernest Miguelez, Stefanie Stantcheva, and Valerio Sterzi. 2018. "Dancing with the stars: Innovation through interactions." National Bureau of Economic Research.

Andersson, David, Mounir Karadja, and Erik Prawitz. 2022. "Mass migration and technological change." *Journal of the European Economic Association*, 20(5): 1859–1896.

Andrews, Michael J. 2023. "How do institutions of higher education affect local invention? Evidence from the establishment of US colleges." *American Economic Journal: Economic Policy*, 15(2): 1–41.

Andrews, Michael J, and Alexander Whalley. 2022. "150 years of the geography of innovation." *Regional Science and Urban Economics*, 94: 103627.

Audretsch, David B, and Maryann P Feldman. 1996. "R and D spillovers and the geography of innovation and production." *The American Economic Review*, 86(3): 630–640.

Bailey, Roy E, Timothy J Hatton, and Kris Inwood. 2023. "Surviving the Deluge: British servicemen in World War I." *Economics & Human Biology*, 49: 101216.

Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen. 2019*a*. "Joseph Schumpeter Lecture, EEA Annual Congress 2017: Do Tax Cuts Produce more Einsteins? The Impacts of Financial Incentives VerSus Exposure to Innovation on the Supply of Inventors." *Journal of the European Economic Association*, 17(3): 651–677.

Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen. 2019*b*. "Who becomes an inventor in America? The importance of exposure to innovation." *The Quarterly Journal of Economics*, 134(2): 647–713.

Bergeaud, Antonin, and Cyril Verluise. 2024. "A new dataset to study a century of innovation in Europe and in the US." *Research Policy*, 53(1): 104903.

Bergeaud, Antonin, Jean-Baptiste Chaniot, and Clément Malgouyres. 2025. "Escaping Labor Scarcity: Innovation and Human Capital after WW1 in France." Working Paper.

Bet-El, Ilana R. 2009. *Conscripts: forgotten men of the Great War.* The History Press.

Boehnke, Jörn, and Victor Gay. 2020. "The Missing Men World War I and Female Labor Force Participation." *Journal of Human Resources*, 0419–10151R1.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2022. "Quasi-experimental shift-share research designs." *The Review of Economic Studies*, 89(1): 181–213.

Brainerd, Elizabeth. 2017. "The lasting effect of sex ratio imbalance on marriage and family: Evidence from World War II in Russia." *Review of Economics and Statistics*, 99(2): 229–242.

Brakman, Steven, Harry Garretsen, and Marc Schramm. 2004. "The strategic bombing of German cities during World War II and its impact on city growth." *Journal of Economic Geography*, 4(2): 201–218.

BWO. 1922. "Statistics of the Military Effort of the British Empire During the Great War, 1914-1920."

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro HC Sant'Anna. 2024. "Difference-in-differences with a continuous treatment." National Bureau of Economic Research.

Carey, Gordon Vero. 1921. *The war list of the University of Cambridge: 1914-1918.* CUP Archive.

Carozzi, Felipe, Edward Pinchbeck, and Luca Repetto. forthcoming. "Scars of War: The Legacy of WWI Deaths on Civic Capital and Combat Motivation." *American Economic Journal: Applied Economics*.

Chen, Jiafeng, and Jonathan Roth. 2023. "Logs with Zeros? Some Problems and Solutions*." *The Quarterly Journal of Economics*, qjad054.

Chupilkin, Maxim, and Zsoka Koczan. 2022. "The economic consequences of war: Estimates using synthetic controls." European Bank for Reconstruction and Development Working Paper 271.

Commonwealth War Graves Commission. 2023. "Online Casualty Database." [Online; accessed 1-Jan-2021].

Correia, Sergio. 2015. "Singletons, cluster-robust standard errors and fixed effects: A bad mix." *Technical Note, Duke University*, 7.

Correia, Sergio, Paulo Guimarães, and Tom Zylkin. 2020. "Fast Poisson estimation with high-dimensional fixed effects." *The Stata Journal*, 20(1): 95–115.

Craig, Edwin Stewart. 1920. *Oxford University roll of service.* Clarendon Press.

Davis, D. R., and D. E. Weinstein. 2002. "Bones, bombs, and break points: The geography of economic activity." *American Economic Review*, 92(5): 1269–1289.

Edgerton, David EH, and Sally M Horrocks. 1994. "British industrial research and development before 1945 1." *The Economic History Review*, 47(2): 213–238.

European Patent Office. 2023. "PATSTAT Global – Autumn 2023." European Patent Office, Vienna, Austria.

FamilySearch. 2023. "United Kingdom World War I Service Records." [Online; accessed 1-Jan-2021].

Forces War Records. 2023. "Online Database." [Online; accessed 1-Jan-2021].

Franck, Raphaël. 2024. "Labor scarcity, technology adoption and innovation: evidence from the cholera pandemics in 19th century France." *Journal of Economic Growth*, 1–41.

Glaeser, Edward L, Hedi D Kallal, Jose A Scheinkman, and Andrei Shleifer. 1992. "Growth in cities." *Journal of political economy*, 100(6): 1126–1152.

Griliches, Zvi. 1990. "Patent Statistics as Economic Indicators: A Survey." *Journal of Economic Literature*, 28(4): 1661–1707.

Habakkuk, H. J. 1962. *American and British Technology in the Nineteenth Century: The Search for Labour-Saving Inventions.* Cambridge:Cambridge University Press.

Hanlon, W. Walker. 2015. "Necessity Is the Mother of Invention: Input Supplies and Directed Technical Change." *Econometrica*, 83(1): 67–100.

Hanlon, W Walker. 2025. "The Rise of the Engineer: Inventing the Professional Inventor During the Industrial Revolution." *The Economic Journal*, 135(670): 1749–1781.

Hicks, John R. 1932. *The Theory of Wages.* London:Macmillan.

Jäger, Simon, Jörg Heining, and Nathan Lazarus. 2026. "How Substitutable Are Workers? Evidence from Worker Deaths." *American Economic Review*. Conditionally accepted.

James, Brigadier EA. 2012. *British Regiments 1914-1918.* Andrews UK Limited.

Jaravel, Xavier, Neviana Petkova, and Alex Bell. 2018. "Team-specific capital and innovation." *American Economic Review*, 108(4-5): 1034–1073.

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy. 2021. "Measuring technological innovation over the long run." *American Economic Review: Insights*, 3(3): 303–320.

Khan, B. Zorina. 2013. "Selling Ideas: An International Perspective on Patenting and Markets for Technological Innovations, 1790–1930." *Business History Review*, 87(1): 39–68.

Lin, Erin. 2022. "How war changes land: Soil fertility, unexploded bombs, and the underdevelopment of Cambodia." *American Journal of Political Science*, 66(1): 222–237.

Maloney, William F, and Felipe Valencia Caicedo. 2022. "Engineering growth." *Journal of the European Economic Association*, 20(4): 1554–1594.

Minnesota Population Center. 2019. "Integrated Public Use Microdata Series, International: Version 7.2 [dataset]."

Moretti, Enrico. 2021. "The effect of high-tech clusters on the productivity of top inventors." *American Economic Review*, 111(10): 3328–3375.

Nagaoka, Sadao, Kazuyuki Motohashi, and Akira Goto. 2010. "Patent statistics as an innovation indicator." In *Handbook of the Economics of Innovation*. Vol. 2, 1083–1127. Elsevier.

Neffke, Frank M. H. 2019. "The value of complementary co-workers." *Science Advances*, 5(12): eaax3370.

Nicholas, Tom. 2011. "Independent invention during the rise of the corporate economy in Britain and Japan 1." *The Economic History Review*, 64(3): 995–1023.

Pretel, David. 2018. "The Global Rise of Patent Expertise During the Late Nineteenth Century." *Technology and Globalisation: Networks of Experts in World History*, , ed. David Pretel and Lino Camprubí, 129–157. Cham:Springer International Publishing.

Riaño, Juan Felipe, and Felipe Valencia Caicedo. 2024. "Collateral damage: The legacy of the secret war in Laos." *The Economic Journal*, ueae004.

Romer, Paul M. 1990. "Endogenous technological change." *Journal of political Economy*, 98(5, Part 2): S71–S102.

Schürer, Kevin, and Edward Higgs. 2014. "Integrated Census Microdata (I-CeM), 1851-1911 [data collection]." *UK Data Service. SN*, 7481.

Shrosbree, Colin. 1988. *Public Schools and Private Education: The Clarendon Commission, 1861–64, and the Public Schools Acts.* Manchester:Manchester University Press.

Silva, JMC Santos, and Silvana Tenreyro. 2006. "The log of gravity." *The Review of Economics and statistics*, 88(4): 641–658.

Smil, Vaclav. 2005. *Creating the twentieth century: Technical innovations of 1867-1914 and their lasting impact.* Oxford University Press.

University of Portsmouth. 2011. "Great Britain Historical GIS Project." [Online; accessed 1-Feb-2022].

Voth, Hans-Joachim. 2021. "Persistence–myth and mystery." In *The handbook of historical economics*. 243–267. Elsevier.

Voth, Hans-Joachim, Bruno Caprettini, and Alex Trew. forthcoming. "Fighting for Growth: Labor Scarcity and Technological Progress During the British Industrial Revolution." *American Economic Review*.

Waldinger, Fabian. 2010. "Quality matters: The expulsion of professors and the consequences for PhD student outcomes in Nazi Germany." *Journal of Political Economy*, 118(4): 787–831.

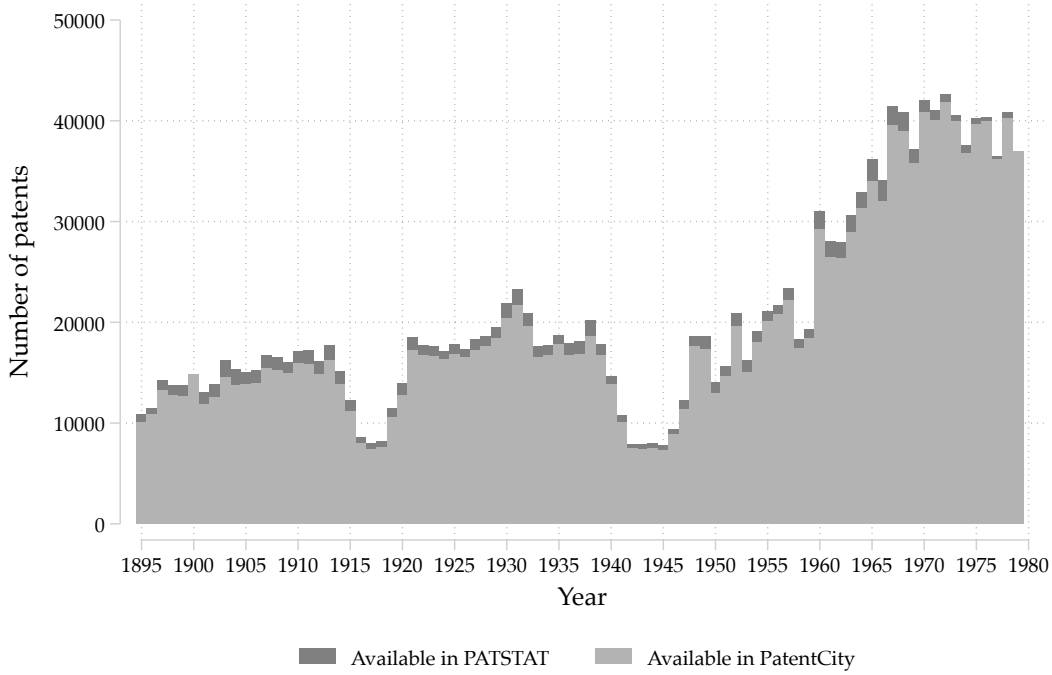
Waldinger, Fabian. 2016. "Bombs, brains, and science: The role of human and physical capital for the creation of scientific knowledge." *Review of Economics and Statistics*, 98(5): 811–831.

Winter, Jay M. 1977. "Britain's 'Lost Generation'of the First World War." *Population Studies*, 31(3): 449–466.

Online Appendix

A. Additional Figures and Tables

FIGURE A.1
PATENTCITY AND PATSTAT DATA



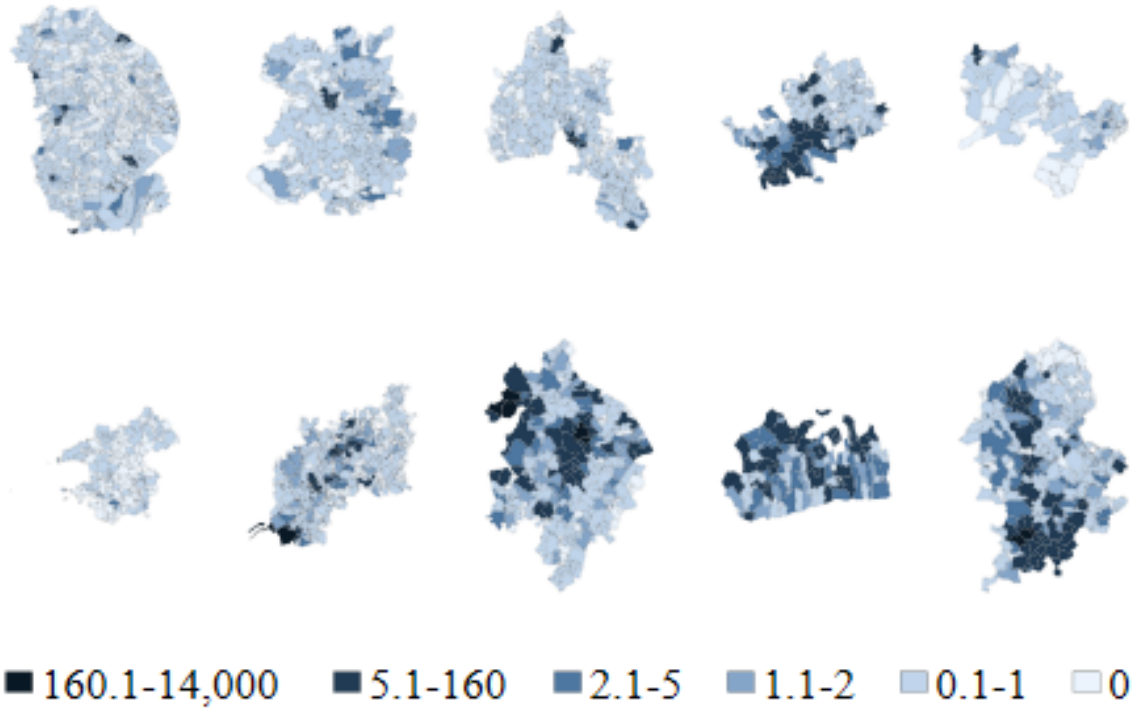
Notes: This figure reports, for each year in our sample, the number of patents registered at the Great Britain patent office and reported in PatentCity (light gray) and in PATSTAT (dark gray), respectively.

TABLE A.1
DESCRIPTIVE STATISTICS

	Mean	Std. dev.	Min	Max
A. Parish-level data				
Population 1911	3,332.30	69218.13	3	7004730
Area (sq. km)	14.25	19.65	0	1314
Population density 1911	252.87	839.97	1	23507
Share of households with servants 1911	0.14	0.09	0	1
Male ratio 1911	0.49	0.07	0	1
N. Mobilised in WWI	277.02	6497.92	0	637773
N. Mobilised/Population in WWI (%)	5.29	4.00	0	32
N. Dead in WWI	53.27	1144.72	0	113285
N. Dead/Population WWI (%)	0.97	0.96	0	10
At least one WWI dead (dummy)	0.88	0.32	0	1
At least one WWI dead had elite educ. (dummy)	0.07	0.25	0	1
At least one engineering-related WWI death (dummy)	0.11	0.31	0	1
Ever registered a patent (1895-1979)	0.66	0.47	0	1
Observations	10807			
B. Parish-level panel data (1895-1979)				
Any patent registered (over 5-year period)	0.24	0.43	0	1
N. patents registered (over 5-year period)	4.13	206.46	0	29501
N. patents reg. by existing inv. (over 5-year period)	1.86	97.39	0	17121
N. patents reg. by first-time inv. (over 5-year period)	1.12	45.61	0	7861
Observations	183719			
C. Inventor-level data				
Co-authored first patent (dummy)	0.33	0.47	0	1
Highly cited first patent (dummy)	0.12	0.33	0	1
Breakthrough first patent (dummy)	0.18	0.38	0	1
Engineer before WWI (dummy)	0.43	0.50	0	1
Observations	54642			

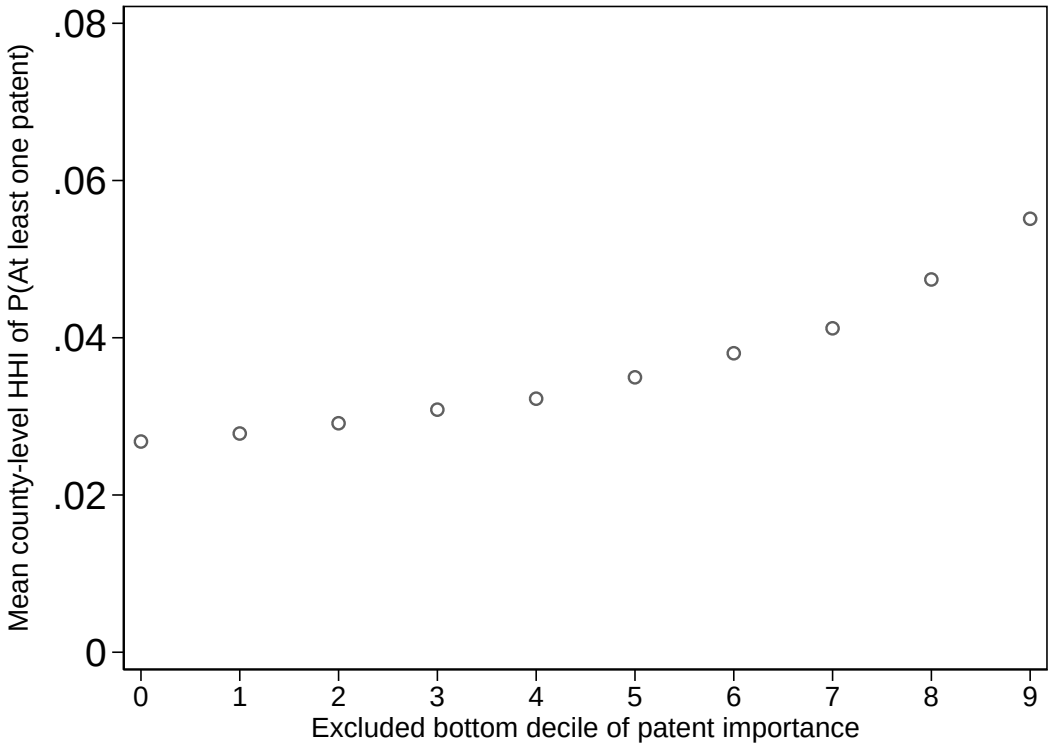
Notes: Panel A: Descriptive statistics at the (grouped) parish level. Data are either from the 1911 Census or from the WWI war records, as indicated. Panel B: Descriptive statistics from the parish-level panel, where each observation is a parish observed over a 5-year period. The panel covers all 5-year periods in 1895-1979. Any patent is an indicator equal to one if at least one patent is registered by a patentee residing in the parish during a given 5-year period. Similarly, we define the number of registered patents as the average number of patents registered in the parish over a given 5-year period, n. patents by existing inventors (i.e., those who registered at least one patent in the pre-war period), n. patents by first-time inventors (those who appear in the data for the first time). Panel C: Descriptive statistics from the inventor-level dataset. Co-authored, highly cited, and breakthrough are indicators for the first patent being co-authored, highly cited, or breakthrough. Engineer before WWI is an indicator equal to one if the patentee reports engineering as their occupation in at least one patent registered before WWI.

FIGURE A.2
PATENTS REGISTERED ACROSS TEN RANDOMLY DRAWN COUNTIES IN ENGLAND AND WALES



Notes: Historical (grouped) parishes for ten randomly selected counties in England and Wales. The figure shows the number of patents registered over a 5-year period, averaged over all periods in 1895-1979.

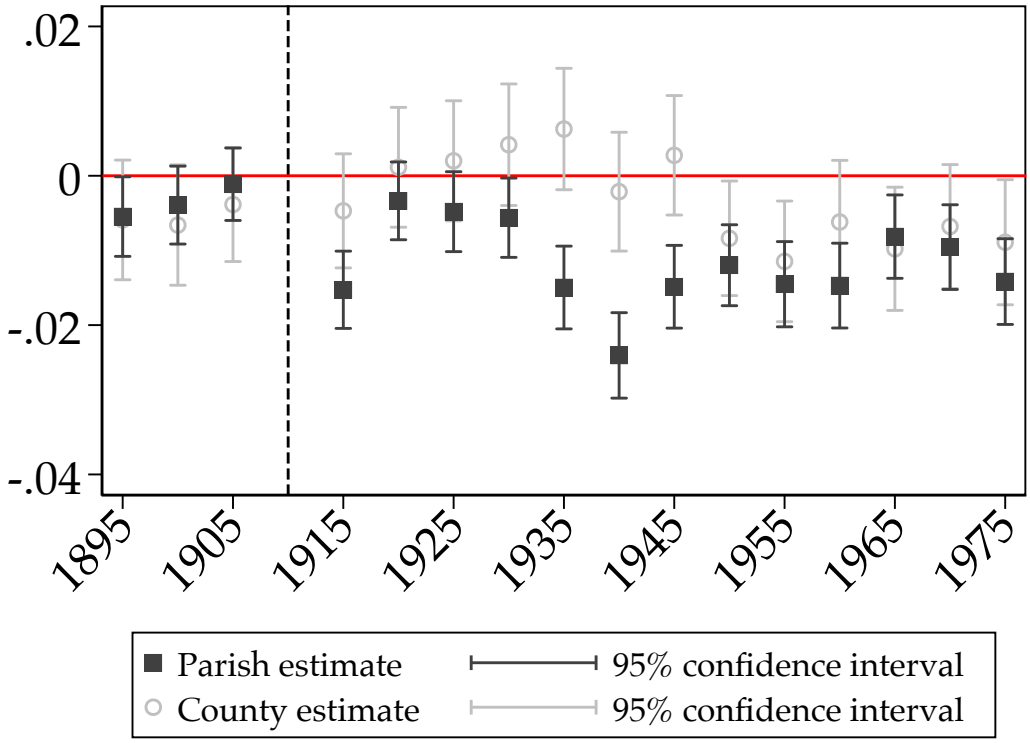
FIGURE A.3
SPATIAL CONCENTRATION OF PATENTING ACTIVITY ACROSS PARISHES OF THE SAME COUNTY



Notes: This figure reports the average Herfindahl–Hirschman Index (HHI) of the spatial concentration of patenting activity across parishes within counties, averaged over all counties in the sample. We first compute, for each parish, the historical average of our main dependent variable – an indicator equal to one if at least one patent is registered by an inventor residing in parish i during 5-year period t – averaged across all periods in the sample. The HHI is then calculated using this historical average distribution of patenting activity across parishes within each county. The analysis focuses on relatively more important patents, as measured by the patent importance index proposed by Kelly et al. (2021). Specifically, patents are ranked within publication cohorts, and we progressively restrict the sample to patents belonging to higher deciles of the importance distribution. Each coefficient in the figure corresponds to a different HHI calculation based on these restrictions: the left-most coefficient uses all patents, while the right-most coefficient measures the spatial concentration of the top 10 percent most innovative patents published during 5-year period t.

Figure A.4

County and Parish level deaths – Event-study results



Notes: OLS estimates of the effect of WWI deaths at parish and county scales on the probability that at least one patent is registered by a resident of parish i during 5-year period t. Estimates from equation 2, including time and parish fixed effects. 95% confidence intervals constructed using standard errors clustered at the parish level.

Table A.2

WWI deaths and patenting – other specifications, raw variables

	(1) Any patent	(2) N. patents	(3) Patents p.c.
A. Log deaths			
$log(d^{WWI})\times$ Post WWI	-0.01***	-0.10***	-0.00
	(0.001)	(0.010)	(0.012)
Observations	162021	121414	162021
B. Death rate p.c.			
$d^{WW1}p.c.\times$ Post WWI	-0.88***	-13.57***	-3.87
	(0.247)	(2.684)	(2.420)
Observations	162021	121414	162021
C. Deaths above median			
$I(d^{WW1} > med)\times$ Post WWI	-0.02***	-0.37***	-0.03
	(0.005)	(0.082)	(0.045)
Observations	162021	121414	162021

Notes: Each cell in this table represents a separate regression. Estimates in columns 1 and 3 are obtained by OLS. Estimates in column 2 are obtained by Poisson pseudo maximum likelihood estimation using the command *ppmlhdfe* in Stata (Correia, Guimarães and Zylkin, 2020). All regressions control for parish fixed effects and county-time fixed effects. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively. The number of observations is lower throughout column 2 because *ppmlhdfe* drops singletons and observations separated by fixed effects (Correia, 2015). Patents p.c. are patents per capita multiplied by 1,000. All non-binary variables are used without winsorizing or trimming.

TABLE A.3

WWI DEATHS AND PATENTING – OTHER SPECIFICATIONS, 95-PERCENT WINSORIZED VARIABLES

	(1) Any patent	(2) N. patents	(3) Patents p.c.
A. Log deaths			
$log(d^{WWI})\times$ Post WWI	-0.01***	-0.10***	-0.02***
	(0.002)	(0.008)	(0.003)
Observations	162021	121414	162021
B. Death rate p.c.			
$d^{WW1}p.c.\times$ Post WWI	-1.28***	-8.12***	-0.45
	(0.317)	(1.274)	(0.594)
Observations	162021	121414	162021
C. Deaths above median			
$I(d^{WW1} > med)\times$ Post WWI	-0.02***	-0.23***	-0.04***
	(0.005)	(0.035)	(0.008)
Observations	162021	121414	162021

Notes: Each cell in this table represents a separate regression. Estimates in columns 1 and 3 are obtained by OLS. Estimates in column 2 are obtained by Poisson pseudo maximum likelihood estimation using the command *ppmlhdfe* in Stata (Correia, Guimarães and Zylkin, 2020). All regressions control for parish fixed effects and county-time fixed effects. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively. The number of observations is lower throughout column 2 because *ppmlhdfe* drops singletons and observations separated by fixed effects (Correia, 2015). Patents p.c. are patents per capita multiplied by 1,000. Continuous outcomes and treatments are winsorized at the 95th percentile.

Table A.4

WWI deaths and patenting – other specifications, 95-percent trimmed variables

	(1) Any patent	(2) N. patents	(3) Patents p.c.
A. Log deaths			
$log(d^{WWI})\times$ Post WWI	-0.01***	-0.09***	-0.03***
	(0.002)	(0.011)	(0.003)
Observations	153960	109062	146173
B. Death rate p.c.			
$d^{WW1}p.c.\times$ Post WWI	-1.16***	-9.34***	-1.06**
	(0.394)	(2.351)	(0.506)
Observations	152949	106431	145153
C. Deaths above median			
$I(d^{WW1} > med)\times$ Post WWI	-0.02***	-0.22***	-0.05***
	(0.005)	(0.035)	(0.006)
Observations	162021	112881	153743

Notes: Each cell in this table represents a separate regression. Estimates in columns 1 and 3 are obtained by OLS. Estimates in column 2 are obtained by Poisson pseudo maximum likelihood estimation using the command *ppmlhdfe* in Stata (Correia, Guimarães and Zylkin, 2020). All regressions control for parish fixed effects and county-time fixed effects. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively. The number of observations is lower throughout column 2 because *ppmlhdfe* drops singletons and observations separated by fixed effects (Correia, 2015). Patents p.c. are patents per capita multiplied by 1,000. Observations above the 95th percentile of the relevant continuous outcome or treatment are excluded.

Table A.5

WWI deaths and patenting – residence and birthplace measures

	(1)	(2)	(3)	(4)	(5)
	Any patent	Any patent	Any patent	Any patent	Any patent
A. Residence					
$Log(d^{Residence})\times$ Post WWI	-0.010***	-0.010***	-0.013***	-0.009***	-0.007***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)
$Log(d^{County})\times$ Post WWI		0.001			
		(0.002)			
Mean dep.var.	0.27	0.27	0.27	0.27	0.27
R2	0.48	0.48	0.49	0.49	0.51
Observations	155062	155062	155062	155062	155062
	(1)	(2)	(3)	(4)	(5)
	Any patent	Any patent	Any patent	Any patent	Any patent
B. Birthplace					
$Log(d^{Birthplace})\times$ Post WWI	-0.010***	-0.010***	-0.012***	-0.008***	-0.007***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)
$Log(d^{County})\times$ Post WWI		0.001			
		(0.002)			
Mean dep.var.	0.26	0.26	0.26	0.26	0.26
R2	0.48	0.48	0.49	0.49	0.51
Observations	158904	158904	158904	158904	158904
Parish FE	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y
County $\times$ Time	N	N	Y	Y	Y
Controls $\times$ Linear trend	N	N	N	Y	N
County $\times$ Time $\times Q_n$ Pop	N	N	N	N	Y

Notes: OLS estimates of the effect of WWI deaths on the probability that at least one patent is registered by a resident of parish i during 5-year period t. Panel A uses deaths assigned by parish of residence, while Panel B uses deaths assigned by parish of birth. Column 2 includes county-level mortality. Columns 3–5 include increasingly demanding fixed effects and controls as in Table 1. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

Table A.6

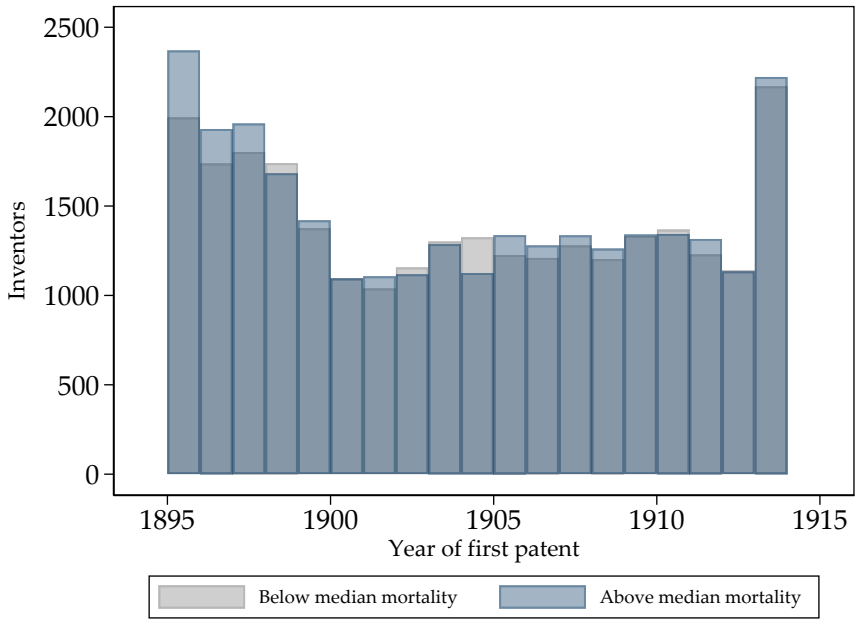
WWI deaths and patenting – Inventor-level estimates – excluding WWI deaths

	(1) Not killed in WWI	(2) Not killed in WWI	(3) Not killed in WWI	(4) Active after WWI
$Log(d^{WWI})\times$ Post WWI	-0.002***	-0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Match on	Surn.+3 dgts.	Surn.+2 dgts	Surn.+1 dgt	–
Mean dep.var.	0.13	0.14	0.15	0.23
R2	0.58	0.57	0.55	0.43
Observations	466429	423317	351883	244647

Notes: OLS estimates of the effect of WWI deaths on an indicator equal to one if inventor j from parish i registers a patent during 5-year period t. The sample is restricted to patentees who registered at least one patent before WWI. Columns 1–3 present results from equation 3, restricting the sample to patentees who do not appear in the CWGC dataset of WWI fatalities under different matching criteria: column 1 requires a match on surname and the first 3 letters of the forename; column 2, surname and the first 2 forename letters; column 3, surname and the forename initial to match exactly. Column 4 restricts the sample to active patentees (those who registered at least one patent in the post-WWI period). Inventor, time, and year of first patent by time fixed effects are included in all specifications. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

Figure A.5

Year of first patent by WWI mortality



Notes: Histogram of the year of first patent among individual inventors in the inventor-level sample (those who patented at least once in 1895–1914). High- and low-mortality groups are defined by whether the inventor's first-patent parish has above- or below-median $\log(d^{WWI})$. The two distributions largely overlap, indicating that inventors from high- and low-mortality parishes do not systematically enter the sample at different times.

Table A.7

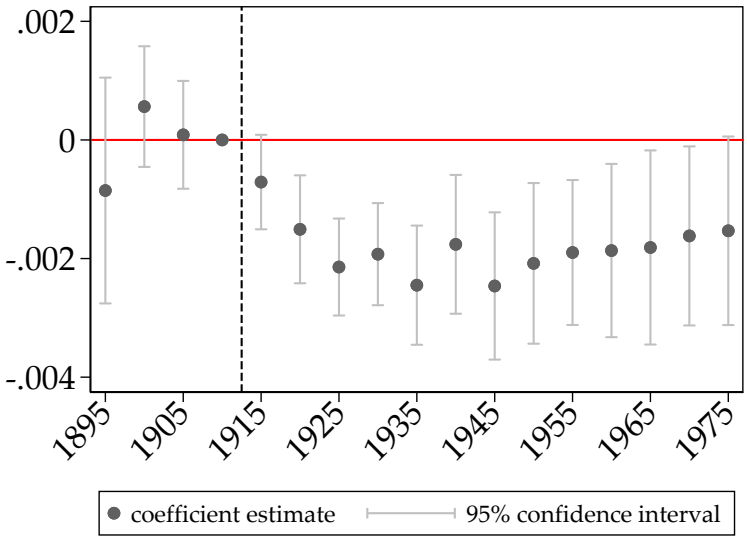
WWI deaths and patenting – Inventor-level estimates – sample restrictions

	(1) Baseline	(2) Drop pre-entry	(3) First by 1909	(4) First by 1904
$Log(d^{WWI})\times$ Post WWI	-0.001***	-0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Sample	All periods	At/after entry	First patent by 1909	First patent by 1904
Mean dep. var.	0.117	0.127	0.122	0.129
Inventors	54381	54381	42461	29659
R2	0.60	0.61	0.59	0.58
Observations	924477	851469	721837	504203

Notes: OLS estimates of the effect of WWI deaths on an indicator equal to one if inventor j from parish i registers a patent during 5-year period t. Column 1 replicates the baseline specification in column 1 of Table 3, keeping all 5-year periods from 1895 onwards. Column 2 drops periods before each inventor's first patent, so that a zero corresponds to an inventor who has already entered the sample and does not patent in that period. Columns 3 and 4 restrict the sample to inventors whose first patent is in 1909 or earlier, or in 1904 or earlier, respectively. Individual, time, and year of first patent by time fixed effects are included in all specifications. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

Figure A.6

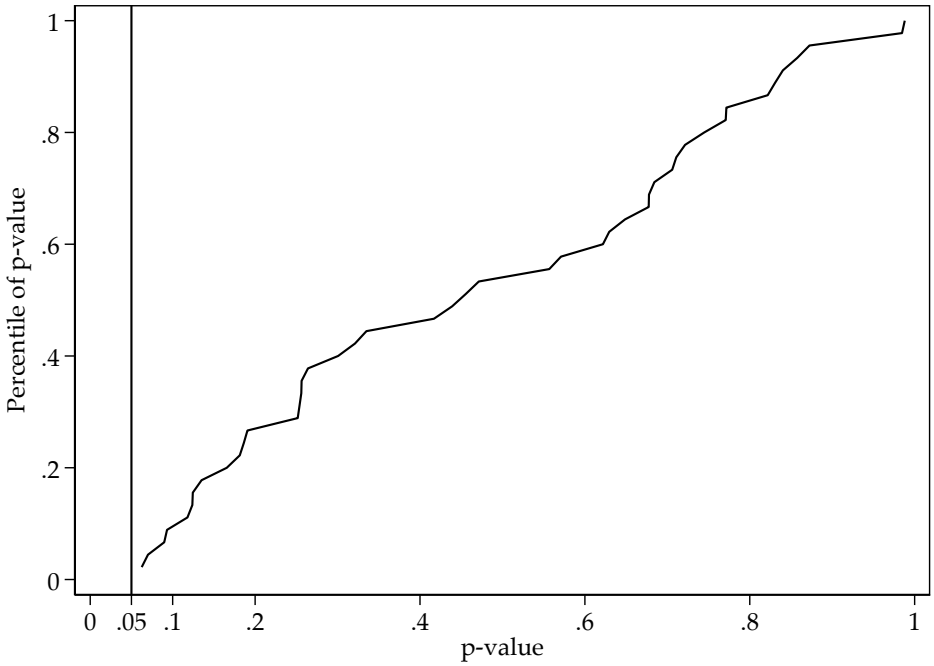
Inventor-level event study – dropping pre-entry periods



Notes: Inventor-level OLS estimates of the effect of WWI deaths on the probability that at least one patent is registered during 5-year period t, restricting each inventor to periods at or after the 5-year bin of their first patent. Estimates from equation 3. Individual, time, and year of first patent by time fixed effects are included. 95% confidence intervals constructed using standard errors clustered at the parish of first patent. The omitted period is 1910–1914.

Figure A.7

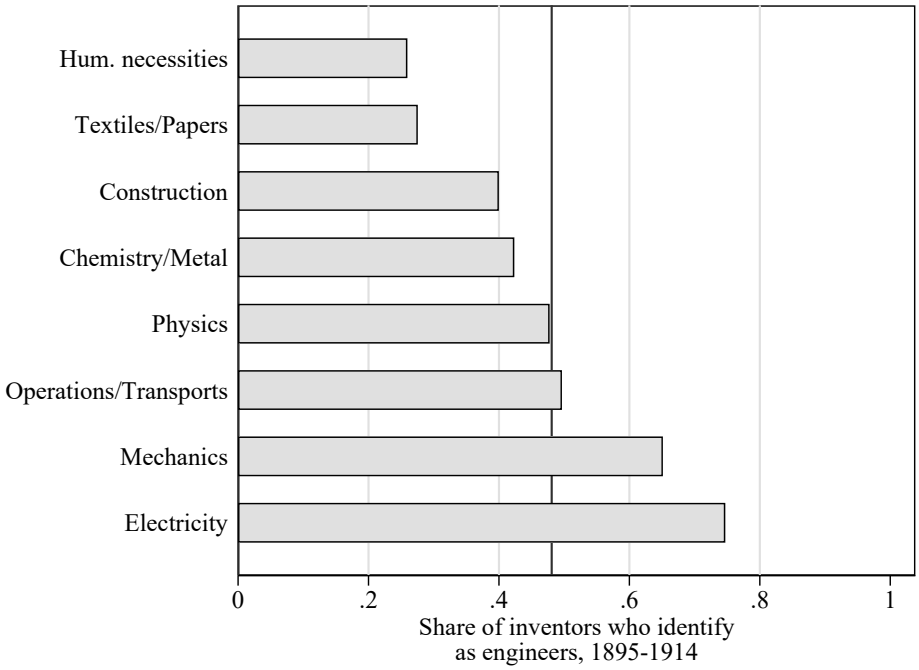
Parallel trends - additional evidence



Notes: The figure plots the empirical CDF of p-values from pairwise "strong" parallel-trends tests across deciles of $\log$(number dead). For each admissible pair of deciles (i, z) (with $i < z \leq 10$), we estimate an event-study regression of the outcome on $\mathbf{1}\{\text{decile} = z\} \times \mathbf{1}\{t\}$ indicators (omitted year 1910), with parish and county-time fixed effects and standard errors robust to clustering at the parish level. We test $H_0 : \beta_{1895} + \beta_{1900} + \beta_{1905} = 0$ and plot the distribution of the resulting p-values; the vertical line indicates $p = 0.05$.

Figure A.8

Prevalence of engineers among inventors by field

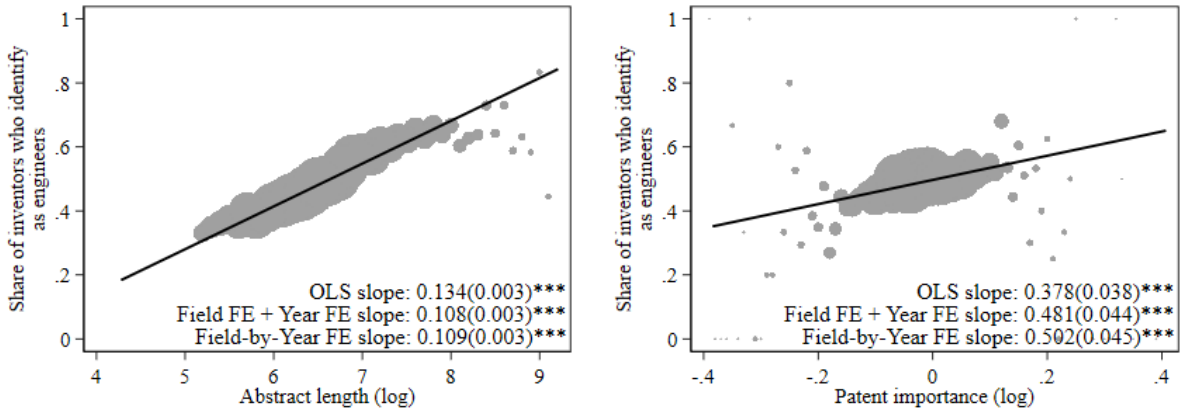


Notes: This figure reports the incidence of engineers among inventors across different fields of specialisation in the pre-WWI period. For each major technological sector, we compute the share of inventors who report their occupation as being "engineer". The figure illustrates the extent to which engineering expertise was predominant among British inventors of the period across various technological domains.

Figure A.9

Engineers, patent complexity and importance

(a) Abstract length (b) Patent importance



Notes: Scatter plots of the abstract length (panel A) and patent importance (panel B – measured following Kelly et al. 2021) against the share of authors who were engineers. Estimated OLS slopes (with or without year and field × year fixed effects) are also reported, with robust standard errors in parentheses.

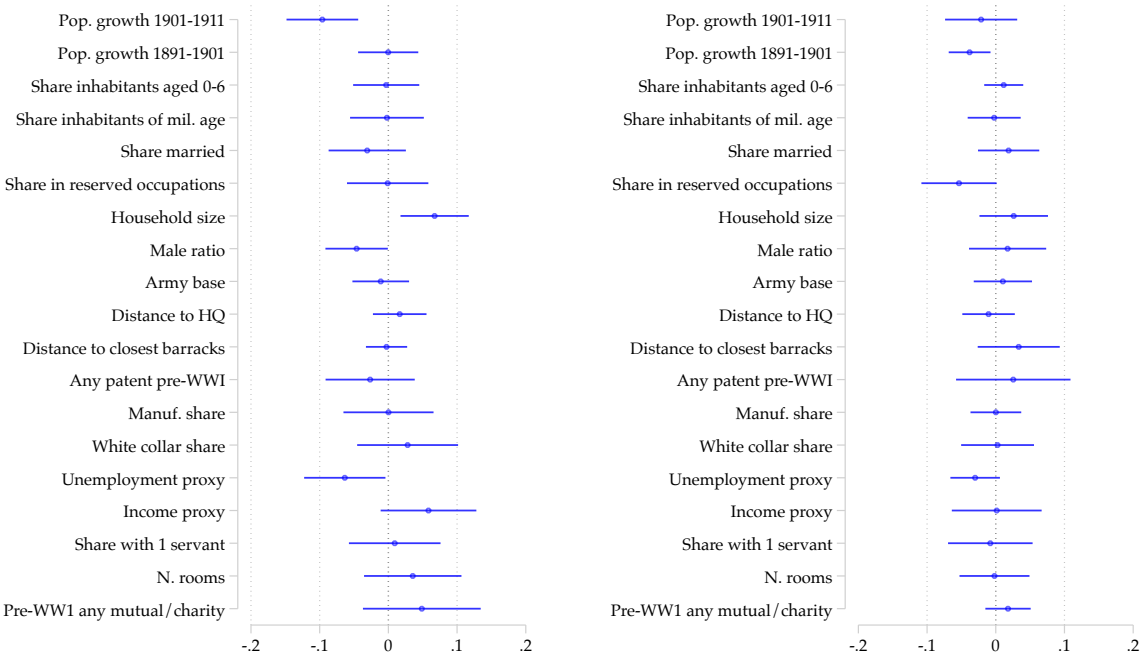
Table A.8

Inventor-level estimates – field, location, and co-authorship

		Same vs. new field		Same vs. new parish		Authorship	
	Any patent	Any patent same field	Any patent new field	Any patent same parish	Any patent new parish	Any single-auth. patent	Any co-authored patent
$Log(d^{WWI})\times$ Post	-0.001***	-0.001***	-0.002***	-0.003***	0.004***	-0.002**	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Mean dep. var.	0.117	0.084	0.049	0.083	0.043	0.083	0.048
R2	0.60	0.74	0.23	0.75	0.23	0.45	0.27
Observations	924460	924460	924460	924460	924460	924460	924460

Notes: Inventor-level OLS estimates of the effect of WWI deaths on patenting from eq. 3. Column 1 reports the baseline estimate (analogous to column 1 of Table 3) for comparison. In column 2 (resp., 3), the dependent variable is an indicator taking the value 1 if inventor j registers at least one patent during 5-year period t within the same sector (resp., within a different sector) as the one the inventor first patented in. In column 4 (resp., 5), the dependent variable is an indicator taking the value 1 if inventor j registers at least one patent during 5-year period t in the same parish (resp., in a different parish) as the one in which they registered their first patent. Finally, in column 6 (resp., 7), the dependent variable is an indicator taking the value 1 if inventor j registers at least one single-authored (resp., co-authored) patent during 5-year period t. Notice that inventor j in the same 5-year period t can produce patents within different fields, in multiple locations, and with different co-authorship teams. Individual, time, and year of first patent by time fixed effects are included in all specifications. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

Figure A.10
Instrumental Variable Balancing Checks

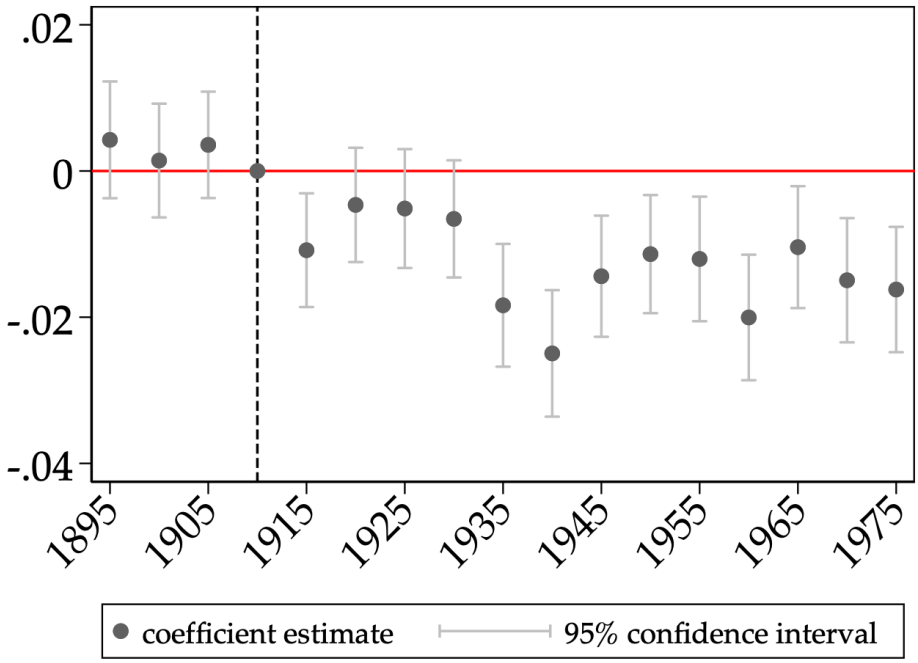


(A) Parish-level (B) Battalion-level

Notes: Panel A: OLS estimates of individual regressions of different variables on the instrument z_i, with 95% confidence intervals. All outcomes are standardised to have mean zero and unit standard deviation. All specifications control for the log of 1911 population, the log of WWI mobilisation, and historic county fixed effects. Panel B: OLS estimates of individual regressions of the battalion-level death rate δ_b on different variables. All variables have been aggregated at the battalion level using the *ssaggregate* command in Stata 19 (residualised using log of population and log of mobilisation) following Borusyak, Hull and Jaravel (2022). Population growth rates are trimmed at the 1st and 99th percentile, while household size, the number of rooms per person, and area are trimmed at the 99th percentile. Share of inhabitants of military and share of married men are calculated restricting to individuals aged 14-35 at the time of the 1911 Census. Income proxy is the first principal component of white collar share, unemployment, share with 1 servant, and n. rooms per person. Standard errors are clustered at the historic county level (panel A) or regiment level (panel B).

Figure A.11

WWI deaths and patenting – Event-study – IV



Notes: IV estimates of the effect of WWI deaths on the probability that at least one patent is registered by a resident of parish i during 5-year period t. Estimates include time and parish fixed effects and county-time fixed effects. 95% confidence intervals constructed using standard errors clustered at the parish level.

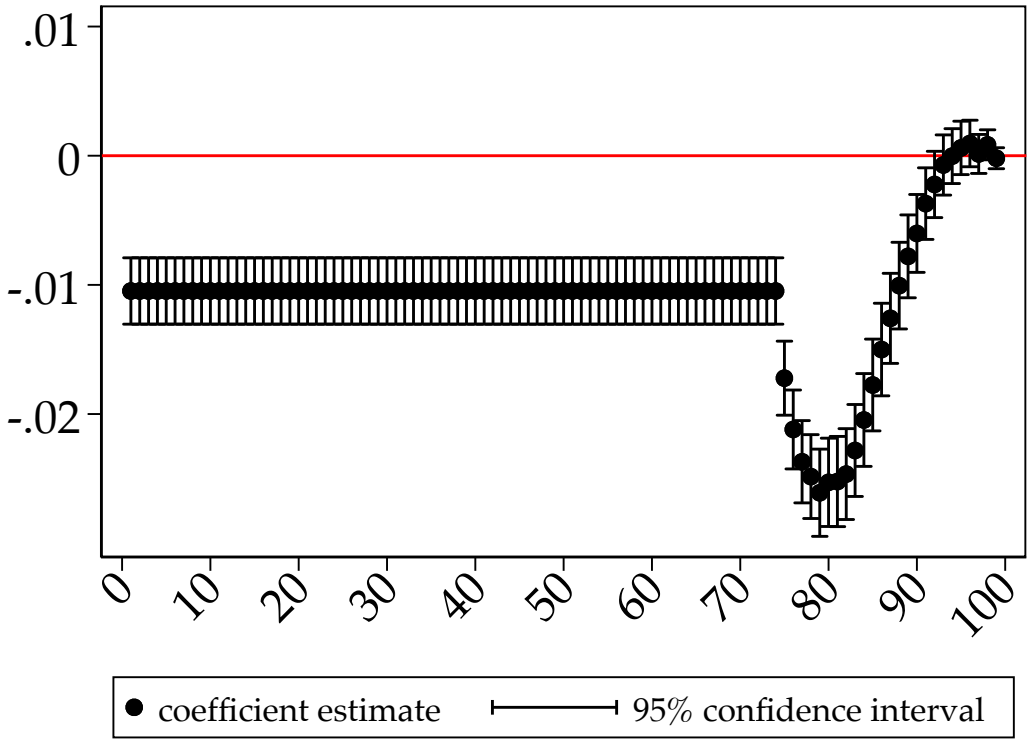
Table A.9

WWI deaths and patenting – inventor-level IV estimates

	(1) Any patent	(2) Any patent	(3) Any breakthr. patent	(4) Any highly cited patent
$Log(d^{WWI})\times$ Post WWI	-0.001***	-0.002***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Sample	Full	Not killed in WWI	Full	Full
Mean dep.var.	0.12	0.13	0.03	0.02
R2	0.00	0.00	0.00	0.00
Observations	896546	475456	896546	896546

Notes: IV estimates of the effect of WWI deaths (instrumented using predicted deaths as described in Section 8.2) on an indicator equal to one if inventor j from parish i registers a patent during 5-year period t. The sample is restricted to patentees who registered at least one patent in the pre-war period. Column 1 presents results from equation 3 for the full sample, whereas in column 2 the sample is restricted to individuals who do not appear in the CWGC dataset of WWI fatalities, matched using the full first name and surname. In columns 3 and 4, the outcomes are, respectively, an indicator for the inventor registering a "breakthrough" patent or a highly-cited patent (see text for details). Individual, time, and year of first patent by time fixed effects are included in all specifications. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

FIGURE A.12
WWI DEATHS AND PATENTING – OLS ESTIMATES – PATENT RATE ABOVE EACH PERCENTILE OF DISTRIBUTION OF PATENT RATES



Notes: OLS estimates of the effect of WWI deaths on the probability that the number of patents registered by residents of parish i during 5-year period t, divided by the population of parish i at the 1911 Census, is above each percentile of the distribution of patent rates in our sample. Each coefficient is a different regression, where the dependent variable refers to the percentile specified on the horizontal axis. Estimates are based on equation 1, including parish and time fixed effects and a control for log population. 95% confidence intervals constructed using standard errors clustered at the parish level.

Table A.10

WWI deaths and patenting – Parish-level estimates including parishes without any WWI deaths

	(1) Any patent	(2) Any patent	(3) Any patent	(4) Any patent
$Log(1+d^{WWI})\times$ Post WWI	-0.008***	-0.013***	-0.008***	-0.009***
	(0.001)	(0.001)	(0.002)	(0.002)
Mean dep.var.	0.24	0.24	0.24	0.24
R2	0.48	0.49	0.49	0.49
Observations	182350	182350	182350	182350
Parish FE	Y	Y	Y	Y
Time FE	Y	Y	Y	Y
County $\times$ Time	N	Y	Y	Y
Controls $\times$ Linear trend	N	N	Y	Y
Population (log)	N	N	N	Y

Notes: OLS estimates of the effect of WWI deaths on the probability that at least one patent is registered by a resident of parish i during 5-year period t. The treatment variable in this regression is log(1+number of deaths). Column 2 adds county-time fixed effects. Column 3 adds controls interacted with a linear trend. Column 4 also controls for the log of population (interpolated). Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

Table A.11

WWI deaths and patenting – Inventor-level estimates including parishes without any WWI deaths

	(1) Any patent	(2) Any patent	(3) Any breakthr. patent	(4) Any highly cited patent
$Log(1+d^{WWI})\times$ Post WWI	-0.001***	-0.002***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Sample	Full	Not killed in WWI	Full	Full
Mean dep.var.	0.12	0.13	0.03	0.02
R2	0.00	0.00	0.00	0.00
Observations	928897	492133	928897	928897

Notes: OLS estimates of the effect of WWI deaths on an indicator equal to one if inventor j from parish i registers a patent during 5-year period t. The sample is restricted to patentees who registered at least one patent in the pre-war period. The treatment variable in this regression is log(1+number of deaths), which implies that inventors active in parishes without any WWI deaths are not dropped from the sample. Inventor and time fixed effects are included in all specifications. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

Table A.12

WWI deaths and patenting – Parish-level estimates excluding London and large cities

	(1) Any patent	(2) Any patent	(3) Any patent	(4) Any patent
A. Excluding London				
$Log(d^{WWI})\times$ Post WWI	-0.009***	-0.012***	-0.008***	-0.009***
	(0.001)	(0.001)	(0.002)	(0.002)
Mean dep.var.	0.26	0.26	0.26	0.26
Observations	162004	162004	162004	162004
B. Excluding parishes with 1911 population above 100,000				
$Log(d^{WWI})\times$ Post WWI	-0.010***	-0.013***	-0.009***	-0.010***
	(0.001)	(0.001)	(0.002)	(0.002)
Mean dep.var.	0.26	0.26	0.26	0.26
Observations	161429	161429	161429	161429
Parish FE	Y	Y	Y	Y
Time FE	Y	Y	Y	Y
County $\times$ Time	N	Y	Y	Y
Controls $\times$ Linear trend	N	N	Y	Y
Population (log)	N	N	N	Y

Notes: OLS estimates of the effect of WWI deaths on the probability that at least one patent is registered by a resident of parish i during 5-year period t. Panel A excludes London (the grouped conurbation, with a 1911 population of 7.0 million and 113,285 WWI deaths). Panel B excludes the 35 parishes with a 1911 population above 100,000 (London, Liverpool, Manchester, Birmingham, Bristol, and other large cities). Column 2 adds county-time fixed effects. Column 3 adds controls interacted with a linear trend. Column 4 also controls for the log of population (interpolated). Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

Table A.13

WWI deaths and patenting – Parish-level estimates removing post-WWII period

	(1) Any patent	(2) Any patent	(3) Any patent	(4) Any patent
$Log(d^{WWI})\times$ Post WWI	-0.006***	-0.008***	-0.006**	-0.006**
	(0.001)	(0.001)	(0.002)	(0.002)
Mean dep.var.	0.26	0.26	0.26	0.26
R2	0.52	0.52	0.53	0.53
Observations	85967	85967	85967	85967
Parish FE	Y	Y	Y	Y
Time FE	Y	Y	Y	Y
County $\times$ Time	N	Y	Y	Y
Controls $\times$ Linear trend	N	N	Y	Y
Population (log)	N	N	N	Y

Notes: OLS estimates of the effect of WWI deaths on the probability that at least one patent is registered by a resident of parish i during 5-year period t. Sample period is restricted to 1895–1939. Column 2 adds county-time fixed effects. Column 3 adds controls interacted with a linear trend. Column 4 also controls for the log of population (interpolated). Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

Table A.14

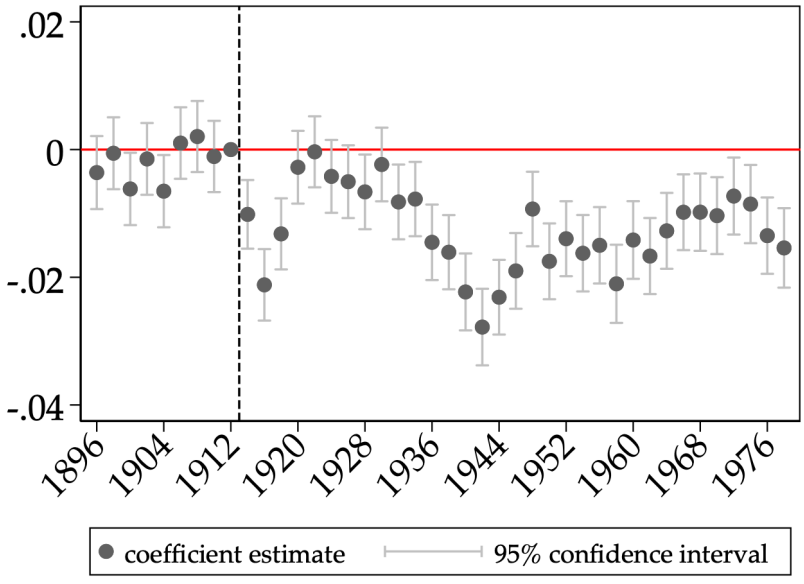
WWI deaths and patenting – Inventor-level estimates removing post-WWII period

	(1) Any patent	(2) Any patent	(3) Any breakthr. patent	(4) Any highly cited patent
$Log(d^{WWI})\times$ Post WWI	-0.001***	-0.002***	-0.000**	-0.001**
	(0.000)	(0.001)	(0.000)	(0.000)
Sample	Full	Not killed in WWI	Full	Full
Mean dep.var.	0.21	0.23	0.04	0.04
R2	0.65	0.63	0.25	0.22
Observations	489420	259164	489420	489420

Notes: OLS estimates of the effect of WWI deaths on an indicator equal to one if inventor j from parish i registers a patent during 5-year period t. The sample is restricted to patentees who registered at least one patent in the pre-war period. Sample period is then restricted to 1895-1939. Individual, time, and year of first patent by time fixed effects are included in all specifications. Standard errors are clustered at the parish level. *, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

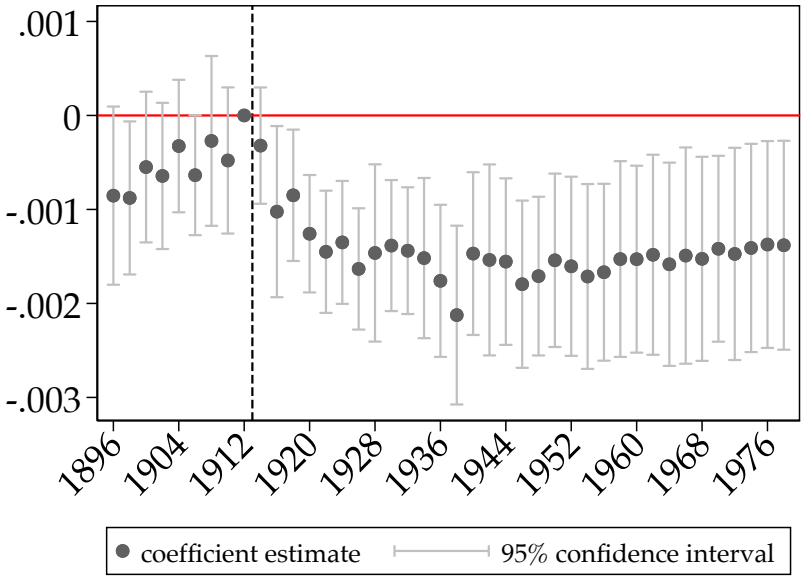
Figure A.13

WWI deaths and patenting – Two-year intervals



Notes: OLS estimates of the effect of WWI deaths on the probability that at least one patent is registered by a parish resident during a 2-year period. Estimates are based on equation 2 and include time, parish, and county-time fixed effects. 95% confidence intervals constructed using standard errors clustered at the parish level.

FIGURE A.14
WWI DEATHS AND PATENTING – TWO-YEAR INTERVALS – INVENTOR-LEVEL



Notes: Inventor-level OLS estimates of the effect of WWI deaths on the probability that at least one patent is registered by an inventor during a 2-year period. Estimates are based on equation 3. Individual, time, and year of first patent by time fixed effects are included. 95% confidence intervals constructed using standard errors clustered at the parish of first patent level.

B. Data

B.1. Military records

Data on British service personnel killed in WWI were obtained from the Commonwealth War Graves Commission (CWGC) (Commonwealth War Graves Commission, 2023), an intergovernmental organisation dedicated to marking, recording and maintaining the graves, memorials and memories of the men and women of the Commonwealth forces who died in both World Wars. Open data from this organisation contain information on names, time of death, rank, fighting unit, age and a string from which we can extract the location of origin of dead soldiers. Data on locations is augmented using information from Forces War Records (FWR), a military genealogy specialist website (Forces War Records, 2023). To implement this merge, we searched FWR for all soldiers in CWGC, and recorded place of birth and residence whenever available. After merging the datasets we perform a number of cleaning operations and restrict attention to soldiers from England and Wales who were killed between 1 August 1914 and 31 December 1918. We then geolocate soldiers to 1911 parish of origin by a combination of matching location strings to parish names and batch geolocation. When assigning soldiers to parishes, we prioritise the FWR data and residential location such that we only use birthplace when residence is missing or cannot be matched to a parish. The geolocation procedure assigns parishes to 73% (585,371) of the soldiers killed.

We next use ancillary data to identify members of the social elite in our CWGC-FWR data. Our primary method to do so relies on the Oxford University Roll of Service (Craig, 1920), and The War List of the University of Cambridge 1914-1918 (Carey, 1921). These documents name all faculty, enrolled students, and alumni of these pre-eminent universities that served during the war. We digitise these documents and match individuals listed to soldiers recorded in the processed CWGC-FWR data using surnames, initials, and rank. We successfully match 3305 individuals (23%) from the Oxford and 3260 (23%) from the Cambridge lists to our soldier-level fatalities dataset in this way.[42] As this definition of elites is highly selective, we then specify an alternative measure of social elites by using the information in the rank field of the CWGC-FWR data to construct a dummy that proxies for a killed soldier being an officer (either commissioned or non-commissioned).[43]

Data on 4,135,026 war records of soldiers mobilised during WWI are obtained from FamilySearch, a not-for-profit organisation which offers on-line access to large genealogical datasets (FamilySearch, 2023). FamilySearch draws its information from the British Army Service Records for 1914 to 1920. These records contain information on enrolled soldiers including names, place of residence, birthplace, age at the time of enlistment, year and unit in which the soldier was enlisted.[44] When cleaning and processing this information, we use as

[42]The match rate reflects the high mortality in these groups, well above the mortality in the general population of servicemen.

[43]The ranks recorded in this field are highly heterogeneous for officers but not for privates, so we define officers as those whose rank does not coincide with a set of predetermined strings (e.g. “Private”, “Rifleman").

[44]Digitised versions of these records can be consulted at www.ancestry.co.uk. The FamilySearch collection, which includes the extracted data, is called “United Kingdom, World War I Service Records, 1914-1920”. The original sources of this information are the “Burnt documents” (record code WO 363) and the “Unburnt collection”

reference the Table of Organisation of each regiment as detailed in James (2012).

B.2. Further data sources

Individual-level information on the English and Welsh population before the Great War is obtained from the 1911 Census of population. The data we use originate from Schürer and Higgs (2014) and is distributed by IPUMS (Minnesota Population Center, 2019). We use this data both at the individual level and to construct aggregates at the parish level. From this source, we obtain information on several income proxies including the number of servants and the number of rooms per household. We obtain aggregate area-level information for Census years 1901-1931, as well as digital maps for parishes, districts and constituencies from "A Vision of Britain through Time" (VoB), an online library of spatial data created by the Geography Department at the University of Portsmouth (University of Portsmouth, 2011). The parish-level population counts in the VoB data come from the *Census Reports* that were published following each Census. There are known to be discrepancies between the population counts in this source and the more recently published micro-data, for example because not all records have survived or there is ambiguity in the true parish in the individual level records. Consequently, in general we use the counts from the *Census Reports* where available. Further, to minimise discrepancies we also implement the corrections to assigned parishes in the 1911 micro-data using the look-up tables published on the I-CeM website.[45]

We obtain membership lists of four engineering trade bodies in pre WWI years – the Institution for Mechanical Engineers, 1896; Iron and Steel Institute, 1888; Institution of Automobile Engineers, 1910; Institution of Electrical Engineers, 1887 – from the website Grace's Guide, a list of pre-WWI universities from Wikipedia, a list of rail stations in 1910 from a GIS first constructed by Jordi-Marti Henneberg, and a georeferenced list of all post offices in 1900 from the GB1900 Gazetteer. We geocode the first two sources using batch geolocation of address information. We then use these sources to construct a series of separate county-level ecosystem accessibility measures. Counties are large – there are 52 in England and Wales – and nest parishes. To construct each measure, we first compute the linear distance between each parish and the nearest ecosystem factor (e.g., a university), then obtain the average distance to that factor in each county by taking the weighted average of the parish level distances. Our weights for this purpose are 1911 parish populations. We then take the natural log of this weighted average distance.

B.3. Spatial Units of Analysis and Reconciliation

Our main analysis is based on a 1911 parish-level dataset covering England and Wales. We take 1911 as our reference year because it was the last Census conducted before the onset of the Great War in 1914. The civil parishes we use in our analysis are administrative units

(record code WO 364), which are kept in the National Archives at Kew in London. The Burnt Documents are roughly 2.5 million records on WWI soldiers which survived the fire resulting from an incendiary bomb hitting the War Office Record Store in 1940. The Unburnt Collection is made of soldier information obtained from pension claims. This collection was stored separately in 1942 and, therefore, did not suffer the fate of many of the Burnt Documents.

[45] Available at https://www.essex.ac.uk/-/media/newparids11.txt?la=en, accessed on May 5, 2023.

corresponding to the lowest level of local government in the United Kingdom. Civil parishes evolved from ecclesiastical parishes during the 19^{th} century, but by 1880 had no religious or ecclesiastical duties. In 1911, the territories of England and Wales were divided into 14,664 parishes, of which 13,404 were in England and 1,260 were in Wales. We drop all parishes that had zero population in 1911 – usually parcels of empty land in remote rural areas – and 10 additional parishes that have repeated names within the same county.

Parishes are nested within local government districts, of which there were 1,861 in 1911, and in turn within 52 counties. Parish boundaries change over time and in some cases variables are only available at other (higher) levels of aggregation. In order to aggregate or re-weight information to common boundaries we use a spatial matching procedure based on the assumption of uniform population distribution within parishes. Because our main spatial units (parishes) are relatively small (10 sq. km on average) and parish boundaries are often quite stable in the 30-year period we study, we expect the measurement error induced by making this assumption to be limited. We selectively group parishes using a semi-automated approach that groups those with similar names and that lie in close proximity. For example, we group High Abbotside with Low Abbotside and we also group Ledbury Urban with Ledbury Rural. In addition to this, we group a small number of suburban parishes with the corresponding city parish. For example, we group the suburb of Sculcoates with Hull. Reflecting that soldiers from the London area often report their location as being London, we also group together London parishes based on the historical conurbation definition available from Vision of Britain. Finally, we exclude 26 parishes, of which 25 are parcels of empty land and the other is unnamed. After restrictions and grouping, our final parish set encompasses 10,807 parishes.

Our data on 1911 parishes come from two different sources: the 1911 Census micro-data from I-CeM and the *Census Reports* from VoB. These sources use different parish codes and contain a slightly different set of parishes, so we create a mapping file and reconcile the data before conducting analysis.

B.4. Geolocation Procedure, Measurement Error, and Validation

Our empirical analysis requires geolocating soldiers based on information on their place of birth and residence from the sources described above. Here we provide details of this procedure.

The CWGC dataset on soldiers killed during WWI includes 796,601 records, from which, because our analysis focuses on England and Wales, we remove servicemen born in Scotland, Ireland, and abroad. We then extract information for residence or birthplace (or both) from either the birthplace and residence fields in FWR or the "additional information" string included in the CWGC source.

Geolocation of WWI fatalities proceeds by combining a) direct string matches with parish names based on data from FWR on historic county and location of birthplace/residence, b) direct string matching as above but based on the CWGC additional information field, and c) latitudes and longitudes obtained from a batch geolocating service to which we input the

FWR locations. For the batch geolocation process, we use a service provided by the company OpenCageGeo, which is based on OpenStreetMap and is available across platforms.

The data on parish of origin (birthplace or residence) of mobilised men in WWI – obtained from FamilySearch – have a slightly different structure and, therefore, we use a different procedure from the one used for CWGC/FWR data.[46] To match the FamilySearch records to an individual parish we combine: a) a direct string match with parish names for records that have both an historic county and a location, b) direct string matching with parish names for records that include no county information (only match to parishes with unique names), c) hand matching of a fraction of remaining records carried out by identifying locations via GoogleMaps. We are able to geolocate just over 2.6 million of these records.

When using this data together with the CWGC information on deaths to construct our instrument, we further exclude 2.05 million records for which the battalion is missing. Finally, we drop 57,795 entries that are duplicates in terms of all variables, and 72,131 records dated before 1905 or after 1920. Of the remaining sample, we then drop 735,768 individuals that could not be geolocated, as well as 28,578 from regiments with zero or negligible mortality, such as the Hussars. Finally, to ensure we have enough observations to construct the shares serving in each battalion, we drop 38,281 soldiers from battalions with less than 100 servicemen in the data.

Owing to measurement error in the geolocation process and the incompleteness of the FamilySearch records, some parishes exhibit mobilisation or WWI death counts that are unusually large relative to their population. To ensure that these potential outliers do not drive our results, we identify all parishes in which the per capita number of mobilised or of deaths lies above the 99^{th} percentile of the respective distribution. We then replace these figures with imputed values obtained by multiplying the 1911 parish population by the corresponding district-level death or mobilisation rate.[47] Results are robust to not applying this correction.

Various validation checks that support our geolocation approaches are provided in Carozzi, Pinchbeck and Repetto (forthcoming).

B.5. Patents data

We draw on two principal sources of patent micro-data: PATSTAT and PatentCity.

PATSTAT is the European Patent Office's global patent database, providing standardised bibliographic data on over 100 million patents collected from more than 100 authorities worldwide. For our purposes, we focus on the subset of patents filed at the UK Intellectual Property Office between 1852 and 1979, encompassing information on inventors, applicants, technical classifications, titles and – where available – textual abstracts (European Patent Office, 2023).

PatentCity is a historical database of British patents covering the years 1895-1979 and

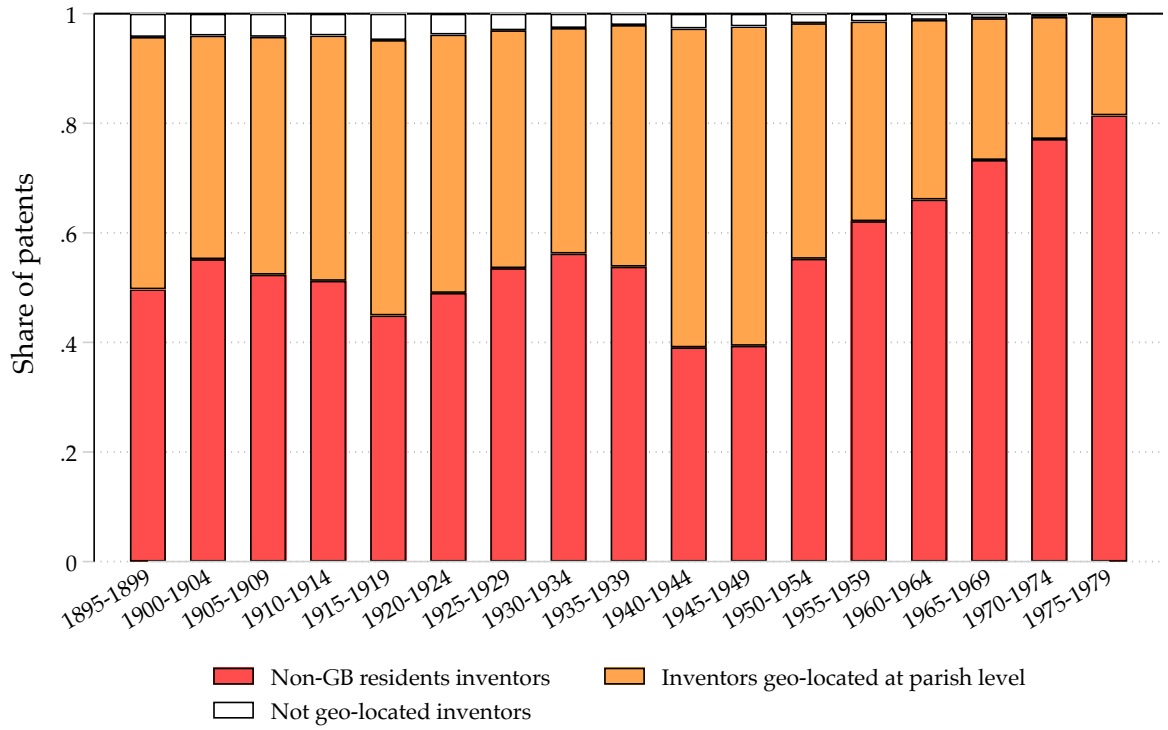
[46] For example, the batch geocoding procedure that we used and validated when using FWR data on locations for killed soldiers yields very poor results when used with the FamilySearch strings.

[47] We use the historical county when district and grouped parish boundary coincide.

specifically designed for spatial analysis of inventors' locations (Bergeaud and Verluise, 2024).[48] For our purposes, we focus on the 3.8 million records registered at the UK Intellectual Property Office. For about 1.2 million of these, at least one inventor can be matched to a location within the United Kingdom at the parish, district, or city level. Of these, we are able to assign a parish in England or Wales to 1.13 million of them. The combination of PATSTAT and PatentCity provides a comprehensive panel of British patents, allowing us to observe inventor identities and their location, as well as information on patent title, abstract, and sector.

B.5.1. Assignment of inventors to parishes

FIGURE B.1
INVENTORS GEO-LOCATED AT PARISH LEVEL



Notes: This figure reports, for each 5-year period in our sample, the share of patents available in PatentCity that are granted to inventors not residing in Great Britain (in red), the share of patents produced by inventors residing in Great Britain that we are able to geo-locate at the parish level (in orange), and the share of inventors residing in Great Britain for which the precision of the information available in PatentCity does not allow a geo-location to England and Wales parishes. Our estimation samples are based on inventor-patent pairs that fall in the orange category.

We assign each inventor–patent tuple to a parish using latitude and longitude coordinates provided by PatentCity, which reflect PatentCity's internal geocoding of inventor addresses.

[48] In the original dataset there are patents dating as early as 1784 and as late as 2021. However, coverage before 1895 is incomplete, and the source used after 1979 is different. For these reasons, we restrict to the period 1895-1979.

We retain observations with sufficiently precise geolocation quality – those geolocated to a city, district, street, house-number, or postal-code – and restrict to patentees located in Great Britain. These coordinates are then spatially matched to historical 1911 parish boundaries via polygon intersection using QGIS, discarding non-matching points.

More specifically, we discard 56 percent of patents because the patentee is located outside the United Kingdom and 2 percent because the georeferenced information provided in PatentCity is not sufficiently detailed or the coordinates thereby reported do not match with the map of 1911 parishes of England and Wales. Figure B.1 reports the share of patents registered by foreign inventors, the share of patents for which we are able to assign a parish to at least one of the inventors, and the share of patents that we are not able to geolocate, for each 5-year window in our sample.

B.5.2. Inventor-level link between PatentCity and PATSTAT

PATSTAT and PatentCity have a common patent identifier, so we can merge the two sources at the patent level. However, they do not share a common inventor identifier, hence a match between the two sources at the inventor level is not immediate. This match is essential for constructing the individual-level panel of inventors because the PATSTAT name and identifier allow us to trace inventors over time. When the patent is single-authored, we merge inventor-level records by the exact patent identifier across the two sources. When the patent is multi-authored, we link the two sources at the inventor level using a fuzzy matching procedure.

To this end, we follow a three-step approach. First, within each patent we generate the full cross-product of PatentCity and PATSTAT inventor entries, producing a long-format list of admissible inventor–inventor candidate pairs (the two sources share a unique patent identifier, so patents can be merged exactly). Second, we clean name strings and then use Stata's *matchit* command to compute fuzzy similarity scores between each PatentCity name and each candidate PATSTAT name within the same patent. For each PatentCity inventor, we select the PATSTAT inventor with the highest similarity score, requiring a minimum score of 0.25; lower-scoring pairs are left unmatched. Third, for matched records, we import the cleaned PATSTAT inventor name and assign it a unique inventor ID based on PATSTAT's numeric inventor identifier. When the same PATSTAT inventor is matched multiple times within a patent, we keep only the highest-scoring match to enforce a one-to-one mapping within each patent. The resulting dataset provides a harmonised inventor-level linkage that combines PatentCity's geolocation and contextual information with PATSTAT's standardised names and numeric inventor identifiers.

Based on these criteria, we match PatentCity and PATSTAT inventors for approximately 750,000 inventor–patent tuples in our main PatentCity data (more than 92 percent of all inventor–patent tuples). We find all PatentCity inventors in PATSTAT for 551,063 patents; for the remaining 49,153 patents, at least one inventor is unmatched.

To construct the inventor-level panel, we exclude patents identified by PATSTAT as registered by companies (non-individual inventors) and we keep only inventors that registered at least one patent in the pre-war period (1895-1914). These restrictions leave us with 54,642

unique individuals, which comprise our final sample.

C. Construction of the patent importance measure

This measure follows Kelly et al. (2021).

C.1. Construction of the Textual Representation

The construction of the patent-importance measure starts from the universe of patents in PATSTAT registered at the Great Britain patent office (with no further restrictions to maximise predictive power) and keeps only records with an abstract. This yields about 1.6 million patents. We transform each abstract into a TF–IDF vector using `scikit-learn`.[49] Each vector is then normalised to unit L^2 norm so that cosine similarity between any two patents equals the dot product of their normalised TF–IDF vectors.

C.2. Similarity Windows and Mean Cosine Similarity

For each patent k filed in year t, we compute its similarity to two sets of comparator patents. The backward set consists of all patents filed during the 5 years preceding t, that is $(t-5,\ldots,t-1)$, while the forward set consists of all patents filed during the ten years after t, that is $(t+1,\ldots,t+10)$. For each of these windows, we compute the mean cosine similarity between patent k and the patents in the corresponding set. Because the TF–IDF vectors are normalised, this step reduces to computing dot products between the vector for patent k and the vectors of all patents in the relevant window.

To make this computation feasible at scale, we follow the centroid-based approach introduced by Kelly et al. (2021). For each year, we construct the sum of the normalised vectors of all patents falling in the backward window and, separately, the sum of the normalised vectors of all patents falling in the forward window. The mean similarity of patent k to a window is then obtained by taking the dot product of the vector for k with the corresponding sum vector and dividing by the number of patents in the window.

C.3. Importance Measure

Following the definition in Kelly et al. (2021), we define the technological importance of patent k as the ratio between its similarity to future patents and its similarity to prior patents,

$$\text{Importance}_k = \frac{\text{ForwardSimilarity}_k}{\text{BackwardSimilarity}_k}$$

This measure is high when the textual content of patent k is more similar to language that appears in subsequent innovations than to the terminology used in earlier ones, indicating that the ideas embodied in k anticipate future developments in the technological frontier.

[49] The vocabulary is restricted to the 1,000 most frequent terms in the corpus and English stopwords are removed.

D. CPC prediction algorithms

This section provides details on the Natural Language Processing (NLP) models used to predict the one-digit Cooperative Patent Classification (CPC) code from patent titles.

We use the patent's title as the input for our NLP algorithms. For each exercise, we restrict attention to patents for which a unique one-digit CPC code is observed in PATSTAT. We divide these patents randomly into a training sample (80%) and a held-out testing sample (20%), using a fixed random seed to guarantee replicability. The training sample is used to estimate the relationship between textual features extracted from the patent title and the CPC field, while the testing sample provides an unbiased evaluation of out-of-sample predictive accuracy.[50] As the name suggests, the training sample is used to "teach" the algorithm about the relationship between the text of each patent's title and the one-digit CPC code.

We estimate four supervised classifiers commonly used in text classification: *Naive Bayes*, *Decision Tree*, *Random Forest*, and *Support Vector Machine* (SVM). All models are trained on the same TF–IDF representation of the cleaned patent titles, where text is lowercased, punctuation is removed, English stopwords are dropped, and terms are weighted using term-frequency inverse-document-frequency with a maximum vocabulary size of 5,000 features.[51] Each model is then used to predict the CPC field for the patents in the testing sample, and its performance is evaluated by comparing predicted and observed CPC codes.

Lastly, we utilise the same algorithm to predict the one-digit CPC code also for patents published before 1910 (for which a CPC code is not available in our sources) and for patents reporting multiple CPC codes (summing up to ≈ 1.3 million patents). Among the algorithms, the SVM achieves the highest predictive accuracy ($\approx 85\%$), and therefore we use the SVM predictions as our preferred measure of the most likely field.

For one-digit CPC codes, we present and discuss three sets of diagnostics based on the testing sample for each of the four classifiers. First, in Figure D.1, we compare the marginal distribution of predicted CPC fields with the true empirical distribution of fields in the test sample. For each classifier, Figure D.1 plots overlaid histograms showing that the predicted and observed distributions are broadly similar, although the degree of alignment varies across algorithms. In all panels, red bars reflect the distribution of one-digit CPC codes as they are observed in the test sample while blue bars report the distribution of predicted one-digit CPC codes in the test sample. All classifiers perform well according to this metric, perhaps with the exception of Naive Bayes, which tends to overestimate the share of patents in sector B (Performing operations and transports) and to underestimate most other fields.

Second, we examine the full pattern of misclassifications by computing, for each observed field, the share of patents that are predicted into each possible field. For every algorithm,

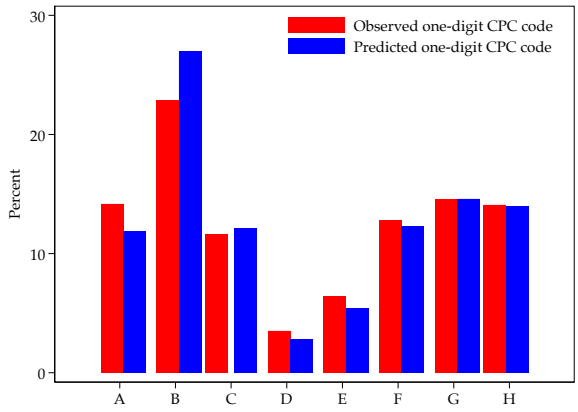
[50] Notice that this exercise is performed on all patents registered at the Great Britain patent office and available on PATSTAT, with no restriction about publication year and/or geo-reference to parishes in PatentCity. We follow this route in order to maximise the accuracy of the prediction algorithm. The total number of patents used is 3,420,207.

[51] These transformations are implemented in Python using the `nltk` and `scikit-learn` libraries.

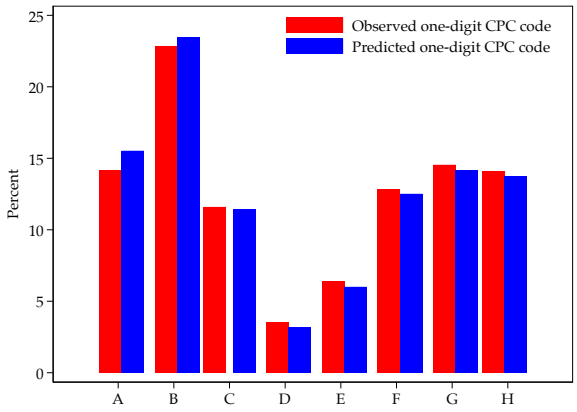
Figure D.1

Comparison of observed and predicted CPC distributions across classifiers

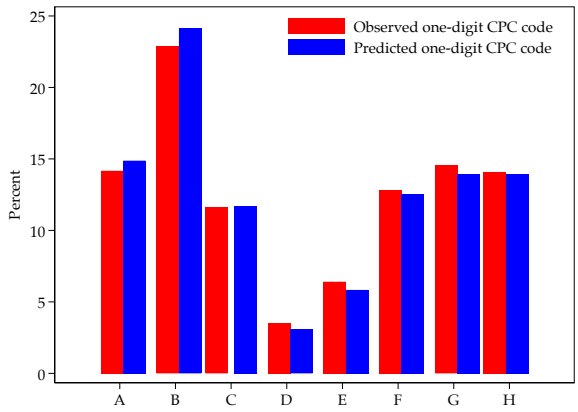
(a) Naive Bayes



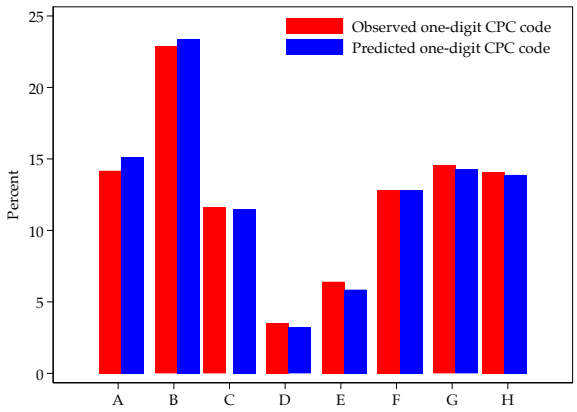
(b) Decision Tree



(c) Random Forest



(d) Support Vector Machine



Notes: Each panel compares the empirical distribution of observed one-digit CPC codes (red) with the distribution predicted by the corresponding machine-learning classifier (blue), computed on the held-out test sample. A closer alignment indicates better predictive performance.

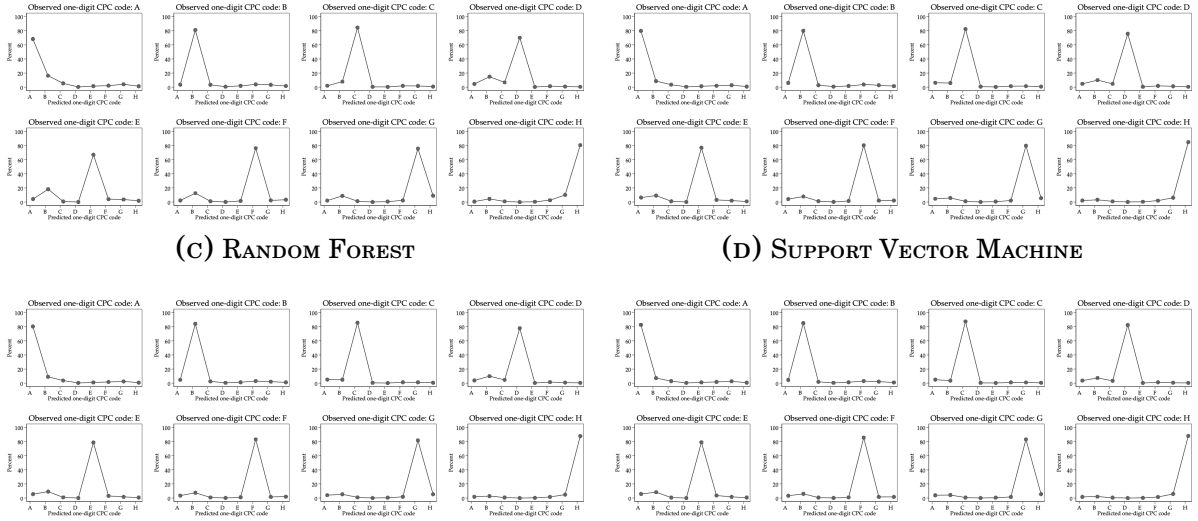
Figure D.2 reports a panel of eight plots (one per CPC field), which can be interpreted as a visual confusion matrix. In a well-performing classifier, each plot displays a dominant peak at the correct CPC category. The evidence presented in Figure D.2 is reassuring about the validity of the prediction exercise, as for all CPC codes and classifiers we consistently see a spike aligned with the observed field. Similarly to the results presented in Figure D.1, the Naive Bayes classifier visibly underperforms compared to the other alternatives.

Third, Tables D.1–D.4 report the detailed classification performance for each of the four supervised models considered in the validation exercise: Naive Bayes, Decision Tree, Random Forest, and Support Vector Machine (SVM). For every one-digit CPC field (A–H), we compute two standard metrics used in multi-class text classification: *recall* and *precision*. Recall is defined as $\Pr(k \in \hat{F} \mid k \in F)$ and measures the share of patents k truly belonging to field F that the algorithm correctly assigns to that field. Precision is defined as $\Pr(k \in F \mid k \in \hat{F})$ and measures the share of patents predicted to be in field F that are in fact observed in that field. These metrics summarise, respectively, how often each model

Figure D.2

Conditional distribution of predicted CPC codes by classifier

(a) Naive Bayes (b) Decision Tree



(c) Random Forest (d) Support Vector Machine

Notes: Each panel summarises, for a given classifier, the conditional distribution of predicted one-digit CPC codes given the observed CPC field. Within each subfigure, the eight small plots correspond to patents whose true CPC is $A, \ldots, H$; on the horizontal axis we report the predicted CPC code, and on the vertical axis the share (in percent) of patents in that true field that are assigned to each predicted field. Thus, for a patent with true field F, the plotted values trace out the distribution of predicted fields $\hat{F}$. For each classifier, these shares are computed as the number of patents with true field F and predicted field $\hat{F}$, divided by the total number of patents with true field F, multiplied by 100. A well-performing classifier displays a dominant peak and relatively low mass on other entries.

misses patents belonging to a field (low recall) and how often it over-assigns patents to a field (low precision). Each table also reports an "Average" row showing the simple average of recall and precision across fields, which describes the model's balanced performance across the CPC spectrum. Lastly, the final row reports the overall *multi-class accuracy*, computed as the share of patents in the held-out test sample for which the predicted one-digit CPC code exactly matches the observed CPC code. This metric serves as our primary criterion for algorithm selection. The Support Vector Machine achieves the highest multi-class accuracy among all classifiers.

Table D.1

Field-level recall and precision: Naive Bayes

CPC field	Recall	Precision
A	0.681	0.812
B	0.809	0.686
C	0.842	0.805
D	0.700	0.872
E	0.668	0.791
F	0.762	0.797
G	0.756	0.753
H	0.805	0.813
Average	0.753	0.791
Accuracy (multi-class)	0.767	

Notes: Rows A–H report recall and precision for each one-digit CPC field. Recall is defined as $\Pr(k \in \hat{F} \mid k \in F)$ and precision as $\Pr(k \in F \mid k \in \hat{F})$. The "Average" row reports averages of recall and precision across fields. The final row reports the overall *multi-class accuracy*, defined as the share of patents in the test sample for which the predicted CPC code equals the observed one.

Table D.2

Field-level recall and precision: Decision Tree

CPC field	Recall	Precision
A	0.796	0.727
B	0.799	0.779
C	0.822	0.834
D	0.756	0.839
E	0.768	0.819
F	0.803	0.822
G	0.796	0.818
H	0.847	0.869
Average	0.798	0.813
Accuracy (multi-class)	0.805	

Notes: Rows A–H report recall and precision for each one-digit CPC field. Recall is defined as $\Pr(k \in \hat{F} \mid k \in F)$ and precision as $\Pr(k \in F \mid k \in \hat{F})$. The "Average" row reports averages of recall and precision across fields. The final row reports the overall *multi-class accuracy*, defined as the share of patents in the test sample for which the predicted CPC code equals the observed one.

Table D.3
Field-level recall and precision: Random Forest

CPC field	Recall	Precision
A	0.802	0.764
B	0.840	0.796
C	0.855	0.849
D	0.778	0.892
E	0.785	0.862
F	0.831	0.850
G	0.815	0.851
H	0.876	0.885
Average	0.823	0.844
Accuracy (multi-class)	0.831	

Notes: Rows A–H report recall and precision for each one-digit CPC field. Recall is defined as $\Pr(k \in \hat{F} \mid k \in F)$ and precision as $\Pr(k \in F \mid k \in \hat{F})$. The "Average" row reports averages of recall and precision across fields. The final row reports the overall *multi-class accuracy*, defined as the share of patents in the test sample for which the predicted CPC code equals the observed one.

Table D.4
Field-level recall and precision: Support Vector Machine

CPC field	Recall	Precision
A	0.826	0.774
B	0.850	0.831
C	0.873	0.882
D	0.823	0.902
E	0.789	0.863
F	0.854	0.854
G	0.830	0.846
H	0.878	0.891
Average	0.840	0.856
Accuracy (multi-class)	0.846	

Notes: Rows A–H report recall and precision for each one-digit CPC field. Recall is defined as $\Pr(k \in \hat{F} \mid k \in F)$ and precision as $\Pr(k \in F \mid k \in \hat{F})$. The "Average" row reports averages of recall and precision across fields. The final row reports the overall *multi-class accuracy*, defined as the share of patents in the test sample for which the predicted CPC code equals the observed one.

E. Construction of the patent breadth measure

We construct a text-based measure of technological breadth using patent abstracts and the one-digit Cooperative Patent Classification (CPC) fields. The intuition is that a technologically narrow patent should use language primarily associated with a single technological field, whereas a broader patent should draw more evenly on language associated with several fields.

We start from the universe of patents registered at the Great Britain patent office for which an abstract is available. Before constructing the textual representation, we remove explicit references to patent-classification headings from the abstracts, so that the measure is not mechanically driven by strings revealing a patent's technological classification. We transform each abstract into a TF–IDF vector using `scikit-learn`. The vocabulary includes unigrams and bigrams and is restricted to at most 10,000 features. English stopwords are removed, terms appearing in fewer than 20 patents or in more than 98 percent of patents are excluded, and each TF–IDF vector is normalised to unit L^2 norm, so that cosine similarity between vectors equals their dot product.

For each patent k published in year t, we measure the similarity of its abstract to the language associated with each of the eight one-digit CPC fields, $f \in \{A, \ldots, H\}$. For each field, the reference set consists of patents assigned by the procedure described in Appendix D to that field and published between $t-10$ and t. We construct the mean normalised TF–IDF vector of the patents in each field and compute the cosine similarity between patent k and each of these eight field-specific reference vectors.

When patent k itself belongs to field f, we remove it from the corresponding field reference before computing the similarity, so that the measure is not mechanically affected by the patent's inclusion in its own comparison set.

Let s_{kf} denote the resulting mean cosine similarity between patent k and the language associated with field f. We transform the eight similarities into shares,

$$p_{kf} = \frac{s_{kf}}{\sum_{g=A}^{H} s_{kg}},$$

so that $\sum_f p_{kf} = 1$. These shares capture how the linguistic content of patent k is distributed across technological fields. We then summarise the concentration of these similarity shares using a Herfindahl–Hirschman Index,

$$HHI_k = \sum_{f=A}^{H} p_{kf}^2.$$

The index is high when the language of patent k is concentrated in a small number of technological fields and low when its language is distributed more evenly across fields. We therefore interpret lower values of HHI_k as indicating greater technological breadth. In the empirical analysis, we use this measure only to distinguish relatively broad from relatively narrow patents. We define *High patent breadth* as an indicator equal to one when HHI_k is below the median of its distribution, and *Low patent breadth* otherwise.