Science Education and Upper-Tail Human Capital: Evidence from the Dutch HBS

Bas Machielsen*

July 6, 2026

Abstract:

The *Hogere Burgerschool* (HBS) was a new science-oriented secondary school aimed explicitly at the Dutch middle class. This paper exploits its staggered rollout to estimate the causal effect of access to scientific secondary education on the formation of upper-tail human capital. I link archival HBS establishment records to the student registers of all five Dutch universities, a national biographical dictionary, and father–son occupational pairs, and find that opening an HBS raised the number of a municipality's children who went on to university by roughly sixteen per treated municipality per five-year cohort bin, and raised the population-scaled enrolment rate by about 3.7 per 10,000 inhabitants, more than doubling the baseline rate among never-treated municipalities, with the response concentrated in the science and medical faculties that the HBS's beta-heavy curriculum targeted. The production of scientists and biographically notable persons also rose, while broad-population occupational-attainment effects are absent, consistent with an effect concentrated in the upper tail of the human-capital distribution rather than a general rise in average mobility. The results identify modern, science-oriented secondary schooling as a channel through which upper-tail human capital was produced.

JEL Classification: I24, J62, N33, N34
Keywords: Hogere Burgerschool, upper-tail human capital, university enrolment, difference-in-differences, educational expansion, Netherlands, nineteenth century

*Utrecht University School of Economics, Kriekenpitplein 21–22, 3584 EC Utrecht, the Netherlands; e-mail: a.h.machielsen@uu.nl.

1 Introduction

In the opening years of the Nobel Prize, the Netherlands, a country of barely five million people, laid claim to four of the first five science Nobels awarded to its citizens in physics and chemistry: Jacobus Henricus van 't Hoff (Chemistry, 1901), Hendrik Lorentz and Pieter Zeeman (Physics, 1902), and Heike Kamerlingh Onnes (Physics, 1913). Willem Einthoven followed in 1924, and Jan Tinbergen and Paul Crutzen added two more over the decades that followed. Surprising is not only that so much talent was concentrated in a small country, so much that contemporaries spoke of a "second Golden Age" (Willink, 1998), but also that these laureates shared a common education. Nearly all of these men were educated not at the classical gymnasium that had trained the Dutch elite for centuries, but at the *Hogere Burgerschool* (HBS), a new type of secondary school created by Thorbecke's Middle Education Act of 1863 and aimed squarely at the commercial and industrial middle class.[1]

This paper takes up the question whether this Dutch cluster of scientific achievement reflects a broader effect of the HBS on the formation of upper-tail human capital rather than a handful of fortunate biographies. Whereas existing work has gauged the upper tail almost exclusively through its most extreme cases, the encyclopaedia subscribers, patentees, and biographical eminences of the pre-industrial era, I reconstruct it across the entire population of municipalities a modern schooling reform touched, linking each place to the university students, biographically notable persons, recognised scientists born there, and the general population. Identification comes from the fact that the HBS network was assembled gradually, town by town, over nearly four decades rather than imposed everywhere at once, so that municipalities that received a school earlier, later, or never provide the counterfactual against which the effect of local access can be read. This design lets me ask not merely whether the HBS produced a few more Lorentzes, but whether it thickened the upper tail of human capital across a whole society.[2]

A growing literature in economic history distinguishes between *average* human capital, the literacy and basic numeracy captured by enrolment rates and signatures that has been the object of most historical human-capital measurement (Cappelli et al., 2023), and *upper-tail* human capital: the knowledge elites, skilled professionals, and notable innovators who disproportionately drive technological and institutional change (Mokyr, 2009, 2016). The two need not move together, and a recurring finding of this work is that it is

[1]Section 2 documents this cluster of achievement, and Dutch scientific performance more broadly, in detail.

[2]The HBS occupied a deliberately new position in the Dutch educational landscape. Thorbecke's Middle Education Act of 1863 created a five-year secondary school with an explicit technical and scientific orientation, aimed at the commercial and industrial middle class rather than the classical learned professions the existing Latin schools and gymnasia served. Funded partly by the state and partly by municipalities that established their own schools with state subsidy, the network grew from a small founding cohort into a nationwide system over nearly four decades, through the staggered, town-by-town rollout this paper exploits. Section 2 details the Act, the rollout's institutional and curricular history, and the political economy of its municipality-by-municipality timing.

the upper tail, rather than mean schooling, that best predicts long-run growth. Far less is understood, however, about what *causes* upper-tail human capital to form in the first place. The evidence that exists points to early-life exposure and local institutions as decisive (Bell et al., 2019), but it is drawn overwhelmingly from the pre-industrial or early-industrial period, or rests on measures of the extreme right tail alone. Whether a modern-style expansion of secondary schooling can raise upper-tail human capital remains an open question. Existing evidence on schooling reforms speaks mainly to the average, the margin on which a student population as a whole moves (Oreopoulos et al., 2006); it says little about whether such a reform can thicken the upper tail across the full population it serves, not merely among its most exceptional graduates.

I exploit the staggered municipal rollout of the HBS between 1864 and 1900 as quasi-random variation in access to scientific secondary education, comparing outcomes across municipalities that received an HBS earlier, later, or not at all. The obvious objection is that municipal (*gemeentelijke*) HBS placement was a local choice that could have been timed to local economic booms, and that treated municipalities are disproportionately large towns. Three features of the design speak to this, developed fully in Section 3.4: the fifteen state-run *rijks-HBS* were sited through national parliamentary negotiation rather than local lobbying, offering a placement margin plausibly free of local timing; the event-study path shows broadly flat pre-treatment coefficients, and formal sensitivity analysis bounds how much of the effect a remaining differential trend could explain; and the composition of the comparison group is confronted directly: the not-yet-treated control group used for the main estimates pools eventual adopters that have not yet received their school with the never-treated municipalities, while the doubly robust estimator's propensity-score weighting bears down on comparison units far from the treated towns' covariate profile, and the headline outcomes are re-estimated against control reservoirs restricted to untreated towns of comparable pre-treatment population.

Opening an HBS raised the number of a municipality's children who went on to university by roughly sixteen per treated municipality per five-year cohort bin, and raised the population-scaled enrolment rate by about 3.7 per 10,000 inhabitants, more than doubling the roughly 2.3-per-10,000 baseline rate among never-treated municipalities. It also raised the production of biographically notable persons and recognised scientists, but left broad-population occupational attainment largely unchanged. This is consistent with an effect concentrated in the upper tail of the human-capital distribution rather than a general rise in average mobility: the HBS examined only 7,100 candidates nationally over three decades (Section 2.1), a narrow slice of any cohort, so even a large per-graduate effect would be invisible in population-wide mobility statistics. The null result on broad occupational attainment is therefore consistent with the scale and aim of the institution.

This paper's contributions are threefold. First, it provides the first causal estimates of a modern, *science*-oriented secondary-schooling expansion on upper-tail human capital, extending a literature that has so far studied either general schooling expansions at the average margin or upper-tail formation in pre-industrial settings with no comparable school-content variation. Second, it makes a measurement contribution: three linked layers (university registers, a national biographical dictionary, and academy membership) that cover the *full* municipality universe a schooling reform touched, rather than only the extreme cases (encyclopaedia subscribers, patentees, biographical eminences) this literature has typically relied on. Third, it shows the effect is upper-tail-specific: university enrolment and notable-person production rise while broad-population occupational attainment does not move, evidence that speaks directly to the claim, central to Mokyr (2009) and Squicciarini and Voigtländer (2015), that the upper tail rather than the mean is the margin on which growth turns.

The results speak to a literature that has so far measured upper-tail human capital almost exclusively before the age of mass secondary schooling, using biographical and prosopographic databases as a spatial proxy for the knowledge elite. Squicciarini and Voigtländer (2015) use *Encyclopédie* subscriptions as a spatial measure of knowledge elites in eighteenth-century France, showing that upper-tail knowledge, but not average literacy, predicts industrial growth. Dittmar and Meisenzahl (2020) use entries in the *Deutsche Biographie* to measure upper-tail human capital across German cities, finding that public-goods institutions adopted during the Protestant Reformation raised the production of notable individuals. David de la Croix's research program has done the most to extend this measurement approach across a longer historical arc: de la Croix et al. (2018) model apprenticeship institutions as a channel of preindustrial human-capital transmission, de la Croix et al. (2024) build a prosopographic database of some 79,000 European scholars active between 1000 and 1800 to map the academic market that produced them, and de la Croix and Licandro (2015) track the lifespans of roughly 300,000 historically notable individuals to document the deep roots of rising longevity among elites. This paper's own outcomes, university enrolment, entries in a national biographical dictionary, and membership of the Royal Academy of Sciences, apply the same biographical-database logic to a nineteenth-century secondary-schooling reform and, unlike most of this literature, measure upper-tail human capital across the full population of municipalities the reform touched, not only its most exceptional individuals.

A second strand asks what *causes* upper-tail human capital to form, and it is here that the present paper speaks most directly. Closest in spirit, Squicciarini (2020) isolates the *content* of schooling as the cause of depressed subsequent industrial development: French districts slow to replace religious instruction with a modern technical curriculum saw weaker industrial development a decade later. Whereas she studies a society held back from a science-oriented curriculum, and measures the payoff on the broad margin of industrial development, I study a reform built expressly to deliver a science-oriented curriculum and measure its effect on the upper tail of the human-capital distribution. Meisenzahl and Mokyr (2016) document that the most productive inventors of the British Industrial Revolution were disproportionately educated and apprenticed in elite settings, and Cantoni and Yuchtman (2014) show that medieval universities shaped the geography of commercial and legal expertise centuries before industrialisation; both point to formal and informal educational institutions, rather than talent alone, as upper-tail-forming. Bell et al. (2019) similarly find early-life exposure and local institutions decisive for the formation of inventors, though within a single country and mostly in the twentieth century. This causal literature has so far found its evidence in early-industrial or narrower settings, while the modern-schooling-reform literature speaks almost entirely to the average margin (Oreopoulos et al., 2006). This paper's contribution is to bring a modern, population-wide secondary-schooling reform into the causal literature on upper-tail formation, and to show that its effect is concentrated exactly where that literature would predict: in the upper tail, not the mean.

This paper is structured as follows. Section 2 details the historical setting. Section 3 describes the data and the empirical strategy. Section 4 presents results across various layers of upper-tail human capital: university enrolment, notable-person production, and broad-population occupational attainment. Section 5 concludes.

2 Historical Setting

2.1 Origin and Growth of HBS Schools

Thorbecke's Middle Education Act of 1863 carved out a new tier of Dutch education between the elementary school and the university. Where the existing Latin schools and gymnasia trained the "learned estate" for university study, the new system of secondary education was meant to serve the commercial and industrial middle class.[3] The flagship institution, the five-year *Hogere Burgerschool* (HBS), combined a broad academic curriculum with an explicit technical and scientific orientation. The law mandated fifteen state-run *rijks-HBS* schools, distributed across the provinces; municipalities could additionally establish their own municipal HBS (*gemeentelijke HBS*) with central-government subsidy. This legal distinction mapped onto a sharp difference in placement politics from the outset: the rijks-HBS network was negotiated nationally in parliament, while municipal openings depended on local initiative, fiscal capacity, and the cooperation of municipal councils. A complementary three-year "citizens' school" (*burgerschool*) for less advanced pupils, and the new Polytechnic School in Delft (*Polytechnische School*, the future Technical University), completed the institutional architecture (Willink, 1998).

The HBS rollout proceeded steadily through the second half of the nineteenth century. The first cohort sat the final examination in 1866, just 38 candidates, but by 1867 there were roughly twenty schools, staffed by 88 teachers holding a doctorate (Willink, 1998). Between 1864 and 1895, 7,100 candidates sat the HBS examination and 5,900 (roughly 83 percent) passed; of those, approximately 2,600 went on to the Polytechnic School in

[3] Willink (1998) reproduces the parliamentary record in which Thorbecke explicitly defends the separation of the gymnasium and the HBS along social rather than age lines: the gymnasium prepares "the future learners," the HBS the "future active society."

Delft (Willink, 1998). The number of secondary-school pupils per thousand children aged 11–14 rose from 8.3 in 1870 to 10.8 in 1880 and 11.6 in 1890, narrowing the gap with Germany (12.6 in 1890) almost entirely (Mandemakers, 1996). By 1900 there were 64 HBS establishments scattered across the country, roughly three times the number a generation earlier.[4]

Three forces governed where and when an HBS opened. First, a loose population threshold of about ten thousand inhabitants served as a rule of thumb for whether a municipality could plausibly support a five-year secondary school; the threshold was not legally binding, but it shaped expectations and central-government subsidies. Second, the fifteen rijks-HBS locations specified in the 1863 law were the product of national parliamentary negotiation rather than local lobbying: Thorbecke's allies in The Hague placed schools to balance provincial coverage, including provinces in which local demand was weak. Third, religious and political opposition mattered, particularly in the south: Willink notes that in Noord-Brabant there was initially little appetite for the HBS (Willink, 1998, 28), and the Catholic clergy in Limburg actively obstructed HBS openings until well into the 1870s. Municipal (*gemeentelijke*) adoption, in contrast, often turned on local circumstance rather than either of the first two forces: in Tilburg, the HBS was installed in King Willem II's former palace after the king's estate transferred the building to the municipality on the condition that a school be established in it, an idiosyncratic local trigger with no counterpart in the population-threshold or parliamentary-placement logic that governed the rijks network.[5]

The HBS curriculum was deliberately weighted toward the natural sciences and mathematics: nearly half of all weekly class hours in the upper grades were devoted to mathematics, physics, chemistry, botany and zoology, mechanics, mineralogy and geology, and cosmography, a sharp departure from the classics-dominated gymnasium (Willink, 1998). That the composition of the curriculum, and not merely the fact of schooling, is what mattered in this era is underscored by Squicciarini (2020): in contemporaneous France, districts that were slow to replace religious instruction with a modern technical curriculum saw weaker industrial development a decade later, precisely because the Second Industrial Revolution had made scientific and technical knowledge economically decisive. Until the Limburg Initiative Act (*initiatiefwet-Limburg*) of 1917, this scientific orientation nonetheless stopped short of direct university access: HBS graduates could not enter a university to study a natural science without first passing a supplementary classical-language examination (*colloquium doctum*), or attending the Polytechnic School in Delft instead (Willink, 1998).[6]

After 1900 the HBS continued to expand, was split in 1921 into humanities- and science-oriented streams, and survived until the 1968 *Mammoetwet* replaced it with the

[4] A parallel girls' secondary school network also emerged during this period; see Appendix A.1 for its timeline.

[5] See Appendix A.1 for further institutional detail, including the Tilburg case.

[6] See Appendix A.1 for the fuller 1867 timetable breakdown and the HBS's laboratory infrastructure.

modern VWO and HAVO tracks (Boekholt and de Booy, 1987; Mandemakers, 1996). Our 1864–1900 window reflects what the rollout permits us to identify cleanly rather than any abolition date. First, our Staatsalmanak-based rollout data naturally end at the turn of the century. Second, the network was already approaching saturation by 1900, weakening the staggered-adoption design thereafter. Third, several institutional changes just after 1900 (girls' admission in 1906 and direct university access in 1917 chief among them) would otherwise have altered the meaning of treatment within the sample.[7]

Local HBS access was, in addition, bounded by who the school served and how far its reach extended. The HBS charged annual tuition (*schoolgeld*), scaled loosely by municipality and offset for less affluent pupils through fee reductions, a practice common across Dutch secondary education in this period; even so, five years of school fees and foregone earnings kept the HBS out of reach for large parts of the working class, consistent with its design as a school for the commercial and industrial middle class rather than the population at large (Mandemakers, 1996). The school was, moreover, boys-only for the entire 1864–1900 window: girls gained no comparable access to a modern-curriculum secondary school until 1906, when the HBS itself opened to girls, after several decades in which a separate girls' network (the *Middelbare Meisjesschool*) offered a more domestic, language-oriented curriculum instead (Appendix A.1). This defines the population at risk in our design and matches the father–son outcome data used throughout the paper. Finally, catchment was geographically narrow: daily attendance required pupils to live within walking or short travel distance of the school, and pupils from villages beyond that radius typically boarded in town for the school term rather than commuting, so that an HBS's practical reach rarely extended much beyond its own municipality and its immediate neighbours (Willink, 1998). This geographic tightness is what makes a municipality-level treatment definition meaningful, and it is also the source of the spillover concern we return to in Section 3.4: a son in a neighbouring, nominally untreated municipality could in principle have boarded near an HBS rather than commuting to one.

2.2 Dutch Scientific Performance

Six of the seven Dutch Nobel laureates in the natural sciences before the Second World War were educated at an HBS (Willink, 1998; Bartels, 1947).[8] Table 2.1 lists these and other prominent Dutch scientists and mathematicians of the period, ten individuals born between 1852 and 1933, eight of them Nobel laureates (the other two, the mathemati-

[7]See Appendix A.1 for an expanded discussion of the 1900 cutoff and the setting's other identifying features.

[8]The seven are van 't Hoff (Chemistry, 1901), Lorentz and Zeeman (Physics, 1902), van der Waals (Physics, 1910), Kamerlingh Onnes (Physics, 1913), Einthoven (Medicine, 1924), and Debye (Chemistry, 1936). All but van der Waals attended an HBS; van der Waals, born in 1837, was already 26 years old when Thorbecke's Act created the school and so could not have attended it regardless of social background. Every count of Dutch Nobel laureates elsewhere in the paper, including the four of the first five laureates cited in the Introduction, refers to this same roster.

cian Brouwer and the astronomer Oort, worked in fields with no Nobel prize category), together with the school each attended. Willem de Sitter, one of the era's most cited astronomers, is included as a deliberate contrast: his father, a judge, sent him to the classical gymnasium rather than the newer HBS, illustrating that the traditional legal and administrative elite continued to favour the classical track even as the commercial and professional middle class embraced the HBS. Family backgrounds among the HBS-educated group are otherwise heterogeneous and skew toward the middle class the school was designed to serve: a market gardener (Lorentz), a village minister (Zeeman), a provincial factory owner (Kamerlingh Onnes), and two schoolteachers (Brouwer, Tinbergen) (Huygens ING, 2026b,c; DBNL, 2026; Huygens ING, 2026a; Erasmus School of Economics, 2026). This is illustrative material, not a systematic sample; the paper's evidence on upper-tail human capital comes from the population-level estimates in Section 4, not from a handful of famous names.

Table 2.1: Notable HBS alumni

Name	Born	Recognition	Secondary school	Source
Hendrik A. Lorentz	1853	Nobel Physics 1902	HBS, Arnhem (1866-69)	Huygens ING
Pieter Zeeman	1865	Nobel Physics 1902 (shared)	HBS, Zierikzee	Huygens ING
Heike Kamerlingh Onnes	1853	Nobel Physics 1913	HBS, Groningen (from 1865)	DBNL
Jacobus H. van 't Hoff	1852	Nobel Chemistry 1901	HBS, Rotterdam	Huygens ING
Willem Einthoven	1860	Nobel Medicine 1924	HBS, Utrecht (diploma 1878)	Huygens ING
Jan Tinbergen	1903	Nobel Economics 1969	HBS, The Hague	Tinbergen Letters (EUR)
Paul Crutzen	1933	Nobel Chemistry 1995	HBS, Amsterdam (1946-51)	NobelPrize.org
Peter Debye	1884	Nobel Chemistry 1936	HBS, Maastricht (to 1901)	Wikipedia (Stedelijk Lyceum Maastricht)
L. E. J. Brouwer	1881	Mathematician (topology)	HBS, Hoorn then Haarlem	Huygens ING
Jan H. Oort	1900	Astronomer	HBS, Leiden (to c.1917)	Universiteit Leiden
Willem de Sitter	1872	Astronomer/cosmologist	No – gymnasium, Arnhem	Huygens ING

Illustrative, non-systematic list of prestigious Dutch scientists born c.1850-1900, compiled to accompany the discussion in Section 2.2; not a sample used in the paper's estimation. Willem de Sitter is included as a contrast case: unlike the other ten, he attended the classical gymnasium track rather than the HBS. Full citations for each Source entry are given in the surrounding text.

Contemporaries already saw this cluster of achievement as remarkable given the country's size: Willink's history of the period is tellingly titled *De tweede Gouden Eeuw* (The Second Golden Age), and Capponi and Frenken (2021) count five Dutch Nobel laureates in the 1901–1913 window alone, the same five laureates as the per-capita comparison below since no additional Dutch science Nobel followed before 1920, arguing, from a core-periphery framework, that HBS science teachers who wrote doctoral theses while working outside the university system, at some remove from established academic hierarchies, produced disproportionately creative contributions. Measured against population rather than against the great powers' output in absolute terms, the achievement looks larger still: tabulating science Nobel laureates (physics, chemistry, and physiology or medicine) by nationality between 1901 and 1920 (NobelPrize.org, 2026) against population circa 1920 (Bolt and van Zanden, 2020), the Netherlands' five laureates per 6.9 million inhabitants place it ahead of every other country in this comparison, and only over the following century did its per-capita standing fall back toward the middle of the international ranking.[9] The HBS's imprint on the wider scientific establishment, not just its most famous graduates, is visible in the professoriate: by 1937–38, 202 of the professors and lectors in Dutch university faculties open to HBS graduates had an HBS background, against only 85 with a gymnasium background (Bartels, 1947).

3 Data and Methodology

3.1 Data sources

The analysis combines data that we construct or harmonise from primary sources, organised by the role each plays in the design: three outcome layers (university enrolment, notable-person production, and broad population occupational attainment), one treatment source (HBS establishment), and two harmonisation layers (occupational coding and municipality-level covariates).

First, for the university-enrolment outcome, we digitise the student registers of all five nineteenth-century Dutch degree-granting institutions (Leiden, Utrecht, Groningen, the Vrije Universiteit Amsterdam, and the Universiteit van Amsterdam), yielding 32,530 student records with birth year, birth place, and faculty. For Leiden, Utrecht, and Groningen these are true *album studiosorum*: matriculation registers that list every enrolled student. The Vrije Universiteit register comes from the *Database VU-studenten 1880–1940*, covering the same matriculation-based logic as the primary registers. The Universiteit van Amsterdam register is collected from the *Album Academicum*, retaining persons recorded as a graduate (*afgestudeerde*, which includes the intermediate *kandidaats/doctoraal* examinations); each person contributes one record, geocoded on city-level birthplace. This underlying record is graduation- rather than enrolment-based; in the baseline analyses, I pool all five registers.[10] These records provide our university-enrolment outcome, the institutional pipeline through which the HBS channelled talent into the knowledge elite.

Second, for notable-person and scientific-elite production, we scrape the biographical dictionary at biografischportaal.nl, which contains approximately 80,000 notable Dutch persons. Each entry includes birth year, birth place, and biographical categories that we classify as upper-tail human capital (business, colonial trade, education and science, healthcare) or not (nobility, fine arts, clergy, military, literature, politics, sport, performing arts).[11] The bioport data provides our measure of extreme-upper-tail human capital:

[9] Laureates per million inhabitants: Netherlands 0.72, Denmark 0.61, Sweden and Switzerland 0.51 each, Germany and France 0.31 each, United Kingdom 0.16, United States 0.02. With five laureates driving the Dutch figure, reclassifying a single laureate's nationality would move the ranking substantially, so this comparison should be read as illustrative rather than as a precise ranking.

[10] In robustness checks, I investigate this difference in the data-generating process, since this last one conditions on degree completion and therefore captures a somewhat more selective margin.

[11] The split follows a single criterion, applied to the category label alone and fixed before any outcome was

the production of historically notable individuals. I supplement bioport with a narrower, discipline-specific companion source: the KNAW's *Digitaal Wetenschapshistorisch Centrum* (dwc.knaw.nl), a structured biographical database of 9,099 persons recognised specifically for scientific achievement: deceased members of the Royal Netherlands Academy of Arts and Sciences (KNAW) and members of regional or disciplinary Dutch learned societies (*geleerde genootschappen*). Of these, 2,178 could be geocoded to an AMCO birth municipality with a usable birth year. For the KNAW subset, the database also links each person's own Academy-proceedings publications, giving a publication-weighted variant of scientific-elite production that bioport does not support.[12]

Third, for the broad-population occupational-attainment outcome, we extract father–son pairs from the genealogical platform genealogieonline.nl, which crowd-sources transcribed Dutch civil-registry and church records and links them into family trees. Our extraction yields 338,880 father–son rows covering sons born between 1828 and 1888, the cohorts who would have been twelve years old in any year between 1840 and 1900, the period over which the HBS rollout we study unfolded.[13] The raw textual occupation field for father and son is preserved.

Fourth, for the treatment source, we construct a municipality-year-level panel of HBS establishments from the editions of the *Staatsalmanak voor het Koningrijk der Nederlanden* (1865-1900), which list every HBS in operation, the school type (rijks or gemeentelijk), and the host municipality. I retrieve the PDFs from the Koninklijke Bibliotheek (Delpher) and transcribe the school lists using a large multimodal language model, and match the resulting municipality names to historical AMCO codes. Of the 1,224 historical municipalities in our analysis, 55 ever received an HBS in the 1864–1898 window; treatment is dated to the first year in which an HBS appeared in the Staatsalmanak.

Fifth, to harmonise the genealogy outcome, we map raw occupation strings into HISCO codes using the Historical Sample of the Netherlands HISCO 2013_01 reference table, and map HISCO codes into the 13-category HISCLASS schema of van Leeuwen and Maas (2011). HISCO/HISCLASS coverage is high: both father and son occupations can be coded for 272,961 of 338,880 pairs (80.5 percent), with the remainder consisting of imprecise occupation strings or strings outside the HISCO classification.

---

tabulated: a category counts as upper-tail if entry into it plausibly reflects acquired knowledge or skill rather than birth, inherited office, or artistic fame. Business, colonial trade, science, and healthcare pass this test; nobility, clergy, military rank, and the arts do not, since standing in those categories derives chiefly from lineage, appointment, or aesthetic judgment rather than a knowledge credential. Politics and literature are the closest calls under this criterion; I conduct a robustness check that admits each to the upper-tail category in turn, alongside a science-only restriction.

[12] Publication linkage is essentially a KNAW-only artifact: a sample check of 80 persons under each filter found linked publications for roughly a fifth to a quarter of KNAW members but zero of the Genootschap members sampled, so we do not construct a publication-weighted measure for Genootschap members.

[13] The dataset contains 295,275 unique father–son pairs; the remaining 12.9 percent of rows are duplicate pairs that surface under multiple URL records on the genealogieonline platform. Before estimation we deduplicate to one row per father–son pair (matched on both names and both birth years), and we cluster the multiplier bootstrap on the son's birth municipality, the level at which treatment is assigned.

Finally, to harmonise covariates across all three outcome layers, we attach municipal-level religious composition from the Historical Database of Dutch Municipalities (HDNG v4), specifically the shares of Roman Catholic, Dutch Reformed (*Nederlands Hervormd*), and *Gereformeerd* inhabitants measured at the census year closest to each son's cohort year. These shares capture plausibly exogenous spatial variation in the cultural environment that may have shaped both HBS placement (through Catholic resistance in the south, as discussed in Section 2) and occupational outcomes independent of treatment.

3.2 Dependent variables

We construct outcomes at three levels of the upper-tail human capital distribution, each measured at a different spatial granularity and drawing on a different primary source.

University enrolment. From the student registers of the five Dutch degree-granting institutions, we aggregate student counts to the (AMCO, 5-year cohort bin) level, where the cohort is dated to the year the student turned twelve (birth year plus twelve), the same clock as the treatment. In each cell we measure enrolment three ways: the raw student count, the rate per $1{,}000$ births, and the rate per $10{,}000$ population. The panel covers the full universe of historical Dutch municipalities, zero-filled for municipalities that send no students, so the outcome captures changes on both the extensive and intensive margins. We also compute a STEM-only version, restricting to students in the natural sciences, mathematics, medicine, dentistry, and pharmacy.

Notable-person production (bioport). From the biografischportaal.nl biographical dictionary, we aggregate notable persons to the same (AMCO, birth-plus-twelve cohort bin) level and compute the raw count of bioport entries and of the upper-tail-category subset (business, colonial trade, science, healthcare), together with the same counts scaled per $10{,}000$ population. I report two population-scaled versions: contemporaneous population (measured in the same year as the cohort bin) and a fixed 1851-baseline population (each municipality's population in 1851, held constant across all of its cohort bins). The 1851 baseline, measured before the earliest HBS opening (1860), avoids controlling for population as an outcome. Scaling notable-person counts by a population predating the outcome window rather than by a contemporaneous or fertility-based denominator follows Bell et al. (2019) and Squicciarini and Voigtländer (2015). This captures the right tail of the human-capital distribution: individuals whose accomplishments were sufficiently significant to warrant a biographical entry in a national reference work. A biographical dictionary's inclusion threshold can itself drift across birth cohorts, as documentation of a person's life becomes more thorough in later archives; cohort fixed effects absorb any such drift that is common nationally, but drift that is correlated with municipality size or urbanity could still load on treatment.[14]

Scientific-elite production (DWC/KNAW). From dwc.knaw.nl we build a narrower companion to the bioport measure, restricted to persons whose notability is specifically scientific recognition. Of the 2,178 geocoded persons, 682 fall within the HBS cohort window (birth years 1810–1910) and populate 372 non-zero (AMCO, cohort-bin) cells. We construct three outcome families at the same (AMCO, cohort-bin) level, each reported as a raw headcount, a rate per $10{,}000$ contemporaneous population, and a rate per $10{,}000$ 1851-baseline population (identical scaling logic to the bioport measure above): KNAW-membership count, the KNAW-subset publication-weighted count (each member's own linked Academy-proceedings publications, summed within the cell), and Genootschap-membership count, a broader, lower-bar notion of scientific recognition reported for comparison in Appendix Table A.19.

Occupational attainment in the broad population (genealogy). From the genealogieonline father–son pairs we construct pair-level outcomes conditional on the father's HISCLASS, so that each estimate answers "given paternal origin, did HBS access shift the son's outcome?". We report the probability that the son reaches HISCLASS 1–3 (corresponding to lower managers/professionals and above),[15] the probability of upward mobility: son's HISCLASS strictly lower, i.e. better, than his father's) and the son's continuous HISCAM-NL occupational-status score (range approximately 40–99). We also estimate the upward-mobility outcome on the low-origin subsample (father HISCLASS > 6). All outcomes are available for the 272,961 pairs in which both father and son occupations are codeable.

These three layers (university enrolment, notable-person production, and broad-population occupational attainment) form a natural hierarchy. The HBS was explicitly designed as a pathway to university and the professions; university enrolment is therefore the most direct institutional mechanism, notable-person production captures the extreme tail of lifetime achievement, and pair-level occupational attainment captures whether the pipeline reached the broader population. Together they trace the causal chain from educational access to upper-tail human capital formation. Table 3.1 reports descriptive statistics for the outcomes, the HBS treatment, and the HDNG municipality covariates, with a Level column indicating the unit (municipality–cohort cell, municipality, or father–son pair) over which each summary is computed.

[14] Bioport aggregates entries from multiple underlying biographical source collections per person, a natural notability weight in the spirit of the "superstar" robustness check of Dittmar and Meisenzahl (2020), who address the analogous concern in the *Deutsche Biographie* by restricting to individuals with longer biographical entries; I conduct the corresponding check, restricting to persons documented in multiple bioport source collections, in Appendix X.

[15] HISCLASS is ordinal from 1 (higher managers and professionals) to 13 (unskilled workers), so a *lower* HISCLASS number denotes *higher* status.

Table 3.1: Descriptive statistics

Variable	Level	Mean	Median	Std. Dev.	Min	Max	N
Panel A: Outcomes							
Students (count)	muni-cohort	1.512	0.000	10.05	0.000	311.00	15,912
Students / 1,000 births	muni-cohort	1.945	0.000	4.595	0.000	133.33	13,660
Students / 10,000 pop	muni-cohort	2.535	0.000	5.576	0.000	157.23	13,914
Bioport / 10k pop	muni-cohort	0.888	0.000	2.590	0.000	47.17	17,232
Bioport upper-tail / 10k pop	muni-cohort	0.140	0.000	0.962	0.000	41.32	17,232
Bioport / 1,000 births	muni-cohort	0.681	0.000	2.493	0.000	166.67	13,660
Bioport upper-tail / 1,000 births	muni-cohort	0.095	0.000	0.848	0.000	66.67	13,660
KNAW membership (count)	muni-cohort	0.019	0.000	0.225	0.000	12.00	19,584
KNAW membership / 10k pop	muni-cohort	0.028	0.000	0.449	0.000	27.62	17,232
KNAW membership / 10k 1851 pop	muni-cohort	0.035	0.000	0.544	0.000	31.75	18,688
KNAW publication-weighted output (count)	muni-cohort	0.035	0.000	0.561	0.000	20.00	19,584
KNAW publication-weighted output / 10k pop	muni-cohort	0.054	0.000	1.572	0.000	124.53	17,232
KNAW publication-weighted output / 10k 1851 pop	muni-cohort	0.074	0.000	2.408	0.000	245.70	18,688
Genootschap membership (count)	muni-cohort	0.017	0.000	0.338	0.000	23.00	19,584
Genootschap membership / 10k pop	muni-cohort	0.025	0.000	0.524	0.000	47.62	17,232
Genootschap membership / 10k 1851 pop	muni-cohort	0.026	0.000	0.529	0.000	40.00	18,688
P(son HISCLASS <= 3)	pair	0.067	0.000	0.249	0.000	1.000	272,961
P(son HISCLASS < father)	pair	0.323	0.000	0.467	0.000	1.000	272,961
Son HISCAM	pair	52.35	51.47	9.649	40.24	99.00	271,859
Panel B: HBS treatment							
Ever-treated municipality (0/1)	muni	0.045	0.000	0.207	0.000	1.000	1,224
Year of first HBS (5-yr bin)	muni	1,867	1,865	7.189	1,860	1,895	55
Son in HBS municipality at age 12 (0/1)	pair	0.061	0.000	0.238	0.000	1.000	272,961
Panel C: Vital statistics							
Total population	muni-cohort	3,985	1,856	16,141	55.00	573,983	13,914
Birth rate (per 1,000)	muni-cohort	32.21	32.13	10.24	0.000	388.24	13,906
Death rate (per 1,000)	muni-cohort	20.64	19.91	8.903	0.000	400.00	13,906
Net migration (per 1,000)	muni-cohort	-3.402	-2.827	20.76	-316.69	1,064	13,914
Panel D: Religious composition							
Catholic share	muni-cohort	0.422	0.265	0.403	0.000	1.000	14,632
Dutch Reformed share	muni-cohort	0.490	0.559	0.356	0.000	1.000	14,632
Gereformeerd share	muni-cohort	0.044	0.000	0.096	0.000	1.000	14,632
Jewish share	muni-cohort	0.004	0.000	0.009	0.000	0.120	14,632
Panel E: Occupational structure							
Agriculture share	muni-cohort	0.401	0.433	0.234	0.001	0.902	4,170
Industry share	muni-cohort	0.303	0.271	0.152	0.029	0.774	4,170
Labor force share	muni-cohort	0.442	0.411	0.149	0.071	3.494	3,975
Panel F: Education, geography, wealth, health							
Literacy rate	muni-cohort	90.09	92.00	5.654	74.00	100.00	1,209
School absenteeism rate	muni-cohort	11.43	10.00	8.003	0.000	52.00	14,599
Population density (per sq. km)	muni-cohort	354.82	91.71	1,701	10.54	39,457	13,859
Wealth-tax share	muni-cohort	1.414	1.270	0.777	0.122	13.56	13,852
Infant mortality (per 1,000)	muni-cohort	54.38	11.24	66.16	0.205	1,169	13,690

Descriptive statistics. The Level column gives the unit over which N is computed: muni-cohort = (municipality x 5-year birth cohort) cells over the full HDNG universe (1,224 municipalities); muni = distinct municipalities; pair = father-son pairs with codeable HISCLASS. Municipality-cohort covariates (Panels C-F) are measured at the HDNG census year closest to each cohort. Treatment is a binary HBS indicator (distance-to-HBS treatment is left to future work); Year of first HBS is the 5-year bin of first establishment among ever-treated municipalities.

3.3 Treatment variable

We define a binary treatment at the AMCO municipality level: a son is treated if his municipality had at least one HBS in operation in the year he turned twelve.[16] The treatment year for a treated municipality is the calendar year in which an HBS first appears in the Staatsalmanak. Of the 55 ever-treated municipalities, treatment years span 1864 to 1898. We bin cohort years (defined as son's birth year plus twelve) into five-year intervals; the same five-year binning is applied to the municipality-level panels, so the pre-treatment event-time grid ($e = -20, -15, -10, -5$) is identical across all three difference-in-differences designs. Table 3.2 summarises the staggered rollout: the number of municipalities whose first HBS opened in each five-year bin, together with the large never-treated reservoir that serves as the control group.

Table 3.2: HBS rollout by cohort

Cohort (5-yr bin of first HBS)	Municipalities treated	Cumulative treated
1860-1864	9	9
1865-1869	31	40
1870-1874	7	47
1875-1879	3	50
1880-1884	2	52
1885-1889	1	53
1890-1894	1	54
1895-1899	1	55
Never treated	1169	
Total	1224	

Number of municipalities whose first HBS opened in each 5-year bin, over the HDNG universe of 1,224 municipalities; 55 were ever treated (first HBS 1864- 1898). Never-treated municipalities form the control reservoir for the difference-in-differences design.

3.4 Estimation and identification

I estimate the effect of HBS access using the staggered difference-in-differences framework of Callaway and Sant'Anna (2021). The building block is the average treatment effect for each pair of treatment cohort (the year a municipality's first HBS opened) and calendar

[16] The choice of age twelve reflects the typical age of HBS entry: the five-year curriculum ran from age twelve to age sixteen or seventeen. Schools entered in the Staatsalmanak as "in oprichting" but not yet operating are not counted as treatments.

cohort (the year a son turned twelve): the estimator compares the evolution of outcomes in municipalities treated in a given year with the contemporaneous evolution in a comparison group of untreated municipalities, estimating each group-time effect by the doubly robust combination of outcome regression and inverse-probability weighting.[17] Because already-treated municipalities never serve as controls, the estimator is immune by construction to the forbidden-comparisons problem that contaminates two-way fixed-effects estimation under staggered adoption with heterogeneous effects (Goodman-Bacon, 2021; de Chaisemartin and D'Haultfœuille, 2023). I aggregate the group-time effects into event-study coefficients in event time (years since the first HBS opening) and summarise the post-treatment path with two scalars: an equal-weighted average of the post-treatment event-time coefficients (labelled "Simple" in the tables) and an average that weights each treatment cohort by its size (labelled "Weighted"). Inference throughout uses a multiplier bootstrap with 999 replications, clustered at the municipality level, the level at which treatment is assigned. For the three municipality-cohort panels this is the estimation unit itself; for the pair-level genealogy panel, where the estimation unit is the father–son pair, we cluster the bootstrap on the son's birth municipality. We report every specification under both a never-treated and a not-yet-treated comparison group.

Two features of the design make standard bootstrap inference fragile and warrant a second, distribution-free check. Only 55 municipalities are ever treated, and several group-time effects are identified off a handful of treated clusters, so the asymptotics underlying the multiplier bootstrap are thin. We therefore complement the bootstrap with randomisation inference: holding the panel fixed, we reassign the 55 observed (municipality, opening-date) adoption events at random to municipalities drawn from the full universe, re-estimate the Simple post-treatment ATT, and repeat this 999 times to obtain a two-sided permutation p-value for each headline outcome (Appendix Table A.10); Appendix Table A.9 reports how many treated municipalities identify each treatment-cohort group. Because we examine several margins per outcome family, we also fix in advance a single *primary* margin for each of the three municipality-level families: for enrolment, the rate per $10{,}000$ population; for notable-person and scientific-elite production, the upper-tail (respectively KNAW) count per $10{,}000$ 1851-baseline population. The remaining scalings (raw counts, the births-scaled rate, and, for the notable-person families, the contemporaneous-population rate) are treated as secondary, descriptive robustness rather than independent confirmations. We do not read significance on a single secondary margin as a finding, and we report both the clustered-bootstrap and the permutation p-value for the primary rows so that the conclusion does not hinge on the bootstrap alone.

Causal interpretation rests on parallel trends and no anticipation, together with co-

[17] The conditioning covariates differ by design, matching what each panel can estimate without inducing a singular regression: the three municipality-cohort panels (Tables 4.1, 4.2, and 4.3) condition on the three HDNG religious shares described in Section 3.1 and estimate every column with the doubly robust estimator. The pair-level specification of Table 4.4 are produced entirely by the outcome-regression-only estimator due to collinearity with the covariates.

variate overlap, which the doubly robust estimator needs to avoid extrapolating off the common support. Each assumption has specific historical content in this setting. Parallel trends requires that, conditional on covariates, treated and comparison municipalities would have followed the same outcome trajectory absent the school: had Leiden not received its rijks-HBS in 1866, the gap in outcomes between Leiden and, say, Maastricht would have evolved as it did before. The threats specific to this period are the differential timing of industrialisation, the rollout of the railway network, differential urbanisation, and the post-1853 Catholic emancipation with its concentrated effect on the south. The religious-share covariates absorb the last of these directly, and the cohort comparisons implicit in the estimator absorb any nationally common trend in the others. The residual concern is that municipal-initiative (*gemeentelijke*) adoption may have been timed to local economic upswings; the state-placed *rijks-HBS*, sited through national parliamentary negotiation rather than local lobbying (Section 2), provide a plausibly exogenous placement margin that I exploit in robustness checks. Because a few pre-treatment placebo coefficients are not exactly zero (Section 4.1), Appendix A.5 subjects the headline estimates to the parallel-trends sensitivity bounds of Rambachan and Roth (2023) and re-estimates them with the alternative staggered-adoption estimators of Sun and Abraham (2021) and Borusyak et al. (2024) and with a specification that conditions on pre-treatment town size directly.

Overlap is a genuine concern because only 55 of 1,224 municipalities are ever treated, and treated municipalities are systematically the larger towns favoured by the ten-thousand-inhabitant threshold (Section 2); the untreated reservoir consists disproportionately of small villages outside the treated group's covariate range. The not-yet-treated comparison group used for the main-text estimates does not remove this concern by construction: it pools eventual adopters that have not yet received their school with *all* never-treated municipalities, so the small-village reservoir enters the main estimates too. Two features of the estimation address it. First, the doubly robust estimator's propensity-score component down-weights comparison municipalities whose covariates are far from the treated towns'; the not-yet-treated adopters, towns that share the scale and trajectory that led them to adopt, are therefore likely to receive a large share of the effective comparison weight, though these weights are not directly observed.[18] Second, Appendix A.3 confronts overlap directly, re-estimating the headline enrolment and notable-person outcomes against control reservoirs restricted to comparable untreated towns: never-treated municipalities with 1851 population above 5,000 and above 10,000, and one-to-one nearest-neighbour matches on 1851 population and on the fitted propensity score of ever-treated status. The headline estimates survive all three restrictions: the enrolment effects shrink modestly, by roughly a tenth to a quarter depending on the margin and the cut, but remain statistically significant, and the notable-person effects are, if anything, slightly larger against compara-

[18]Appendix Table A.1 fits a cross-sectional logit of ever-treated status on pre-treatment (1840) municipality covariates and finds that 38.5 percent of never-treated municipalities' fitted propensity scores fall within the range spanned by the 55 ever-treated municipalities' own scores (Appendix Table A.2 and Figure A.1).

ble towns. The same exercise documents how structurally thin the comparable-untreated margin is: only two never-treated municipalities exceeded ten thousand inhabitants in 1851, so at the Act's own placement margin the design can compare adopters almost only with later adopters.

No anticipation requires that outcomes not respond to treatment before it is realised, and here the relevant question is whether the decisions that determine the outcome predate the school's opening. For pre-treatment cohorts they do: because treatment is dated to the year the son turns twelve, the father's occupation and the family's residence at the son's birth are fixed twelve years before treatment is measured. The announcement-to-opening lag reinforces this: Royal Decrees authorising rijks-HBS schools were typically issued within months of opening, leaving families little advance notice to act on. The one channel that could still violate the assumption is selective migration ahead of an opening. Because every outcome in this paper is assigned by *birthplace*, this channel matters only if it operates before the son's birth: a family that moves toward an HBS town after the son is born, including a move timed to enrol him at twelve, does not contaminate the outcome assignment, since the son's municipality-cohort cell is fixed at birth. Pre-birth residential sorting, by contrast, would contaminate it. Such sorting is testable, however, since it should appear as anticipatory growth in births or population before an HBS opens. As discussed in Section 3.2, births and population in HBS municipalities rise only *after*, not before, treatment. A non-mover restriction, isolating this channel directly using the birthplace information in genealogieonline, is employed in a robustness check.[19]

Finally, a son in a nominally untreated municipality could in principle commute to, or board near, a neighbouring HBS. To the extent this occurs, the binary treatment understates effective treatment intensity in the comparison group and biases the estimates toward zero, so our estimates should be read as lower bounds on the true effect of HBS access; a continuous-treatment specification using distance to the nearest HBS at age twelve is left to future work.[20]

4 Results

We organise the results along three layers of upper-tail human capital, each measured at a different point of the pipeline the HBS was built to feed. The first is university enrolment,

[19] See Appendix X.

[20] This lower-bound argument has a counterweight worth flagging explicitly. Section 3.2 notes that births and population grow in HBS municipalities after treatment, which is why bioport outcomes are scaled by a fixed 1851-baseline population rather than a contemporaneous one. That scaling fixes the *denominator*, but it does not rule out that part of a birthplace-based headcount effect reflects *composition* (different or additional families sorting into HBS municipalities after birth) rather than human-capital production among incumbent families. The two concerns pull in opposite directions on the same estimate: control-side spillovers push it toward zero, post-birth sorting could push it away from zero. The non-mover restriction discussed above is the eventual answer to the sorting concern, just as a continuous-treatment specification is the eventual answer to the spillover concern.

the direct institutional mechanism. The second is the production of biographically notable persons, the extreme right tail. The third is occupational attainment in the broad father–son population. Within the notable-persons layer we report a narrower, corroborating measure alongside the broad biographical dictionary: specific recognition as a scientist, drawn from the KNAW's biographical database. All estimates use the staggered difference-in-differences estimator of Callaway and Sant'Anna (2021) with a not-yet-treated control group.[21]

4.1 University Enrolment

Opening an HBS raised university enrolment in a municipality, and it did so most clearly on the margins least vulnerable to small-municipality noise. In treated municipalities the number of children who went on to university rose by roughly sixteen per five-year cohort, and the population-scaled enrolment rate rose by about 3.7 students per $10{,}000$ inhabitants, both significant at the one-percent level (Table 4.1, Panel A). Against a baseline of about 2.3 per $10{,}000$ among never-treated municipalities, this is more than a doubling of the rate at which a town sent its children to university. The one margin on which the effect is imprecise is the births-scaled rate, which is positive but statistically indistinguishable from zero. We attach little weight to it: the births denominator is volatile in small municipalities, and it grows noisier still once the large, heavily urban Universiteit van Amsterdam graduate register is pooled with the four matriculation registers.[22] We therefore read the population-scaled rate as the more reliable intensive-margin measure.

The response is concentrated in the science and medical faculties, precisely where the HBS's curriculum was directed. Restricting the outcome to the STEM faculties (mathematics, natural sciences, medicine, dentistry, and pharmacy) yields count and population-scaled effects that are highly significant and about half the all-faculty magnitude (Panel B). STEM is also the only margin on which even the births-scaled rate remains significant. That the effect loads on exactly these fields is consistent with the HBS's deliberately beta-heavy programme, which devoted nearly half of all upper-grade class hours to mathematics and the natural sciences (Willink, 1998).

The dynamic path in Figure 4.1 matches what a pipeline mechanism predicts: the post-treatment effect builds gradually over event time rather than appearing all at once. The first treated cohorts, aged twelve when the school opened, do not reach university age until about six years later, and the enrolment effect strengthens as successive cohorts pass through the school. The pre-treatment coefficients are broadly flat, with a few minor exceptions. On the all-faculty count margin the placebo at $e = -10$ is positive and significant

[21] Robustness to the never-treated control and the full dynamic event-study paths are collected in Appendix A.

[22] Appendix A.4 confirms this directly: dropping the Universiteit van Amsterdam register from the enrolment panel leaves the count and population-scaled effects significant at the one-percent level, while the births-scaled rate turns significant once that register is excluded (Table A.5).

(1.73, roughly a tenth to a sixth of the mature post-treatment effect of 12–24 students), the population-scaled rate shows significant placebos at $e = -10$ and (marginally) $e = -5$, and the STEM margins show a marginal placebo at $e = -5$.

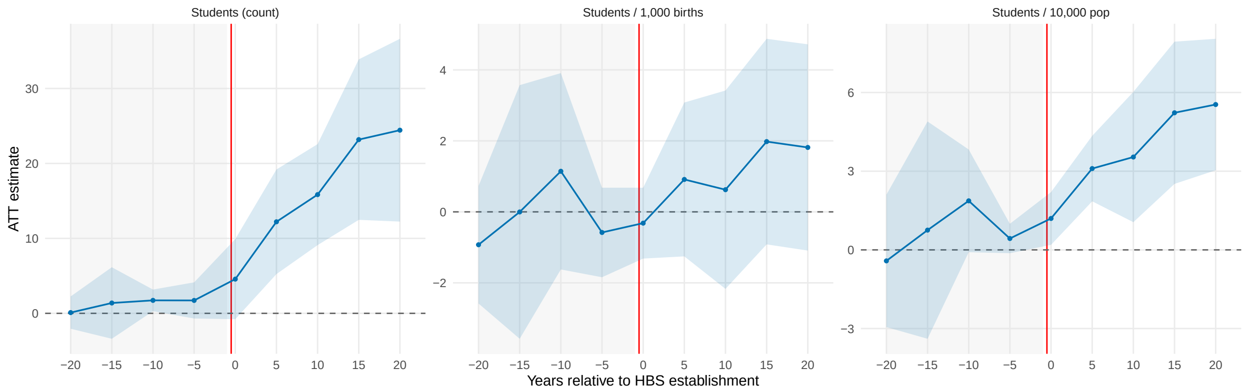
The positive sign of these placebos is exactly what the growing-towns confound of Section 3.4 would produce: towns on a growth path send more children to university before the school arrives and are also more likely to receive one. Two pieces of evidence indicate that this does not drive the headline estimates. First, the significant $e = -10$ count placebo attenuates to insignificance in every trimmed or matched control reservoir of Appendix A.3, consistent with it reflecting the comparison with small villages rather than a differential trend among comparable towns. Second, Appendix A.5 subjects the estimates to the formal sensitivity analysis of Rambachan and Roth (2023), which bounds how large a post-treatment parallel-trends violation the data could hide, and re-estimates the headline margins with the alternative estimators of Sun and Abraham (2021) and Borusyak et al. (2024) and with pre-treatment town size added to the conditioning set; the mature enrolment-count effect survives hypothetical violations up to 1.7 times the worst observed pre-treatment deviation, clearing the natural benchmark of one with room to spare, and doing so equally on the untrimmed panel and on the comparable-town control reservoirs. The population-scaled rate breaks down at smaller multiples, but this is largely a mechanical consequence of carrying the same pre-treatment deviation against a smaller effect; its support rests instead on the point estimates themselves, which remain significant in every trimmed, matched, and town-size-conditioned specification. The alternative estimators reproduce the sign and magnitude of the headline estimates.

Table 4.1: University enrollment DiD estimates

	Students (count)	Students / 1,000 births	Students / 10,000 pop
Panel A: All faculties			
e = -20	0.0995	-0.9270	-0.4198
	(1.0037)	(0.8089)	(1.1554)
e = -15	1.3796	-0.0029	0.7531
	(2.2236)	(1.7457)	(1.8962)
e = -10	1.7310***	1.1432	1.8695**
	(0.6717)	(1.3513)	(0.8961)
e = -5	1.7215	-0.5810	0.4333*
	(1.1236)	(0.6164)	(0.2569)
Post ATT[0,20] (Simple)	16.0457***	1.0027	3.7212***
	(3.9196)	(1.1445)	(0.7241)
Post ATT[0,20] (Weighted)	20.6973***	0.7341	3.4595***
	(4.7383)	(1.4557)	(1.0831)
N observations	13,863	12,749	13,863
N cohorts	8	7	8
Panel B: STEM faculties			
e = -20	0.4008	-0.5333**	0.0055
	(0.7910)	(0.2435)	(1.0174)
e = -15	1.0331	-0.3361	0.6796
	(1.8892)	(2.7192)	(1.7486)
e = -10	0.1696	-0.1452	0.2391
	(0.9543)	(0.8407)	(0.8422)
e = -5	1.4149*	-0.0699	0.4464***
	(0.7372)	(0.5897)	(0.1688)
Post ATT[0,20] (Simple)	7.7408***	0.4852**	1.7784***
	(2.0325)	(0.2275)	(0.2405)
Post ATT[0,20] (Weighted)	8.8686***	0.2843	1.5326***
	(2.0193)	(0.3499)	(0.3540)
N observations	13,863	12,749	13,863
N cohorts	8	7	8

Callaway and Sant'Anna (2021) staggered difference-in-differences on a panel of municipalities by 5-year birth cohort (year the child turned twelve). Unit of observation: municipality-cohort cell. Outcome is the number / rate of children from a municipality enrolling at one of five universities (Leiden, Utrecht, Groningen, VU Amsterdam, Universiteit van Amsterdam; the latter two are graduate/non-matriculation registers). Panel B (STEM) restricts to the mathematics, natural-science, medicine, dentistry and pharmacy faculties. The count is unweighted; the rate per 1,000 births is weighted by births; the rate per 10,000 population is weighted by municipal population. Rows $e < 0$ are pre-treatment placebo ATTs. Post ATT[0,20] (Simple) is the equal-weighted average of dynamic event-time ATTs over e = 0..20; (Weighted) is the treatment-group-size-weighted average. Bootstrapped SE, clustered by municipality (999 reps), in parentheses. Group-time effects are estimated with the doubly robust (outcome regression + inverse-probability weighting) Callaway-Sant'Anna estimator. * p<0.1, ** p<0.05, *** p<0.01. Control group: not-yet-treated municipalities.

Figure 4.1: Event-study estimates: university enrolment, all faculties (not-yet-treated control).



4.2 Notable-person Production

Table 4.2 moves from the institutional pipeline to its rarest output: individuals whose accomplishments earned them an entry in the biografischportaal.nl national biographical dictionary. Here too, treated municipalities produced more. HBS access raised both the number of notable persons a municipality produced per cohort and the notable-person rate per capita, corroborating the enrolment result at the far right tail of the distribution.

Because a municipality's contemporaneous population and birth counts themselves respond to HBS establishment, the cleanest of these measures scales notable-person counts by the municipality's 1851 population, held fixed across cohorts and safely pre-treatment. This follows the way Bell et al. (2019) and Squicciarini and Voigtländer (2015) normalise notable-person and inventor counts by a population predating the outcome window. On this treatment-immune denominator the effect is positive and significant at the one-percent level. The upper-tail-only subset, restricted to the business, science, and healthcare categories, moves in the same direction across all three margins and is significant on the count and 1851-baseline denominators.

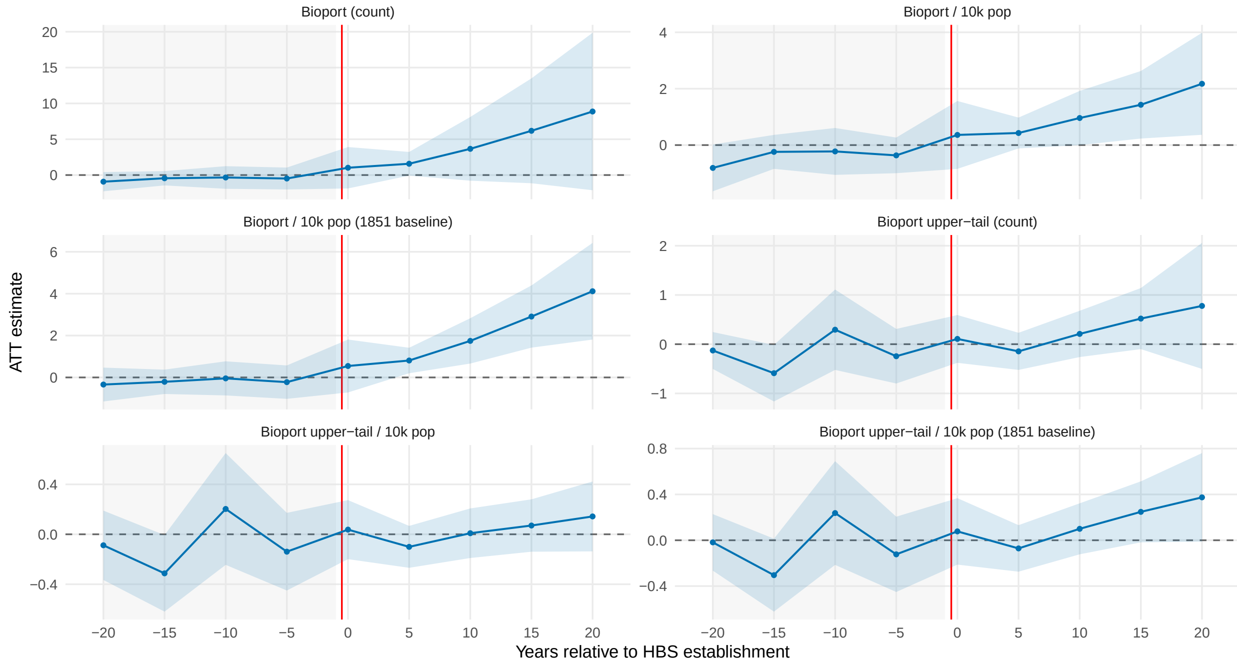
Two qualifications temper this reading. Some pre-treatment coefficients are not perfectly flat, so parallel trends stand on weaker footing here than for enrolment, and the notable-person outcome is by construction the noisiest and most selected of our measures. The extensive margin carries the most weight: under the randomisation-inference test of Appendix A.6 (Table A.10) the upper-tail *count* clears the test decisively, while its noisier population-scaled counterpart points the same way and stays close to conventional significance. We therefore treat the notable-person result as suggestive corroboration of the enrolment finding, firmest on the extensive count margin, rather than as clean identification in its own right.

Table 4.2: Upper-tail human capital DiD estimates

	Bioport (count)	Bioport upper-tail (count)	Bioport / 10k pop	Bioport upper-tail / 10k pop	Bioport / 10k pop (1851 baseline)	Bioport upper-tail / 10k pop (1851 baseline)
e = -20	-0.9231	-0.1276	-0.8101**	-0.0881	-0.3397	-0.0183
	(0.5874)	(0.1535)	(0.3571)	(0.1143)	(0.3384)	(0.0998)
e = -15	-0.4430	-0.5891**	-0.2405	-0.3129**	-0.2087	-0.3045**
	(0.4376)	(0.2352)	(0.2594)	(0.1261)	(0.2436)	(0.1288)
e = -10	-0.3387	0.2942	-0.2250	0.2028	-0.0447	0.2375
	(0.6911)	(0.3344)	(0.3580)	(0.1842)	(0.3414)	(0.1832)
e = -5	-0.4807	-0.2444	-0.3650	-0.1394	-0.2248	-0.1229
	(0.6684)	(0.2270)	(0.2723)	(0.1280)	(0.3343)	(0.1330)
Post ATT[0,20] (Simple)	4.2616*	0.2937	1.0706**	0.0317	2.0247***	0.1457*
	(2.3478)	(0.1889)	(0.4739)	(0.0723)	(0.5514)	(0.0768)
Post ATT[0,20] (Weighted)	7.4976**	0.6035*	0.7980*	-0.0211	3.2899***	0.2734***
	(3.7751)	(0.3524)	(0.4237)	(0.0680)	(0.6860)	(0.0893)
N observations	17,178	17,178	17,178	17,178	16,900	16,900
N cohorts	8	8	8	8	7	7

Callaway and Sant'Anna (2021) staggered difference-in-differences on a panel of municipalities by 5-year birth cohort (year the child turned twelve). Unit of observation: municipality-cohort cell. Outcome is the number of persons born in a municipality who later received a biografischportaal.nl entry ("upper-tail" restricts to business, colonial trade, science and healthcare categories). The count columns are unweighted raw counts; the per-10,000-pop columns scale by contemporaneous municipal population; the 1851-baseline columns scale by each municipality's population in 1851 (the first year of near-universal HDNG coverage, and safely pre-treatment since the earliest HBS opening is 1860), held fixed across all of that municipality's cohorts, so it cannot itself respond to treatment. Scaling notable-person counts by a population measured before the outcome window follows Bell et al. (2019) and Squicciarini and Voigtlander (2015). Rows $e < 0$ are pre-treatment placebo ATTs. Post ATT[0,20] (Simple) is the equal-weighted average of dynamic event-time ATTs over $e = 0..20$; (Weighted) is the treatment-group-size-weighted average. Bootstrapped SE, clustered by municipality (999 reps), in parentheses. Group-time effects are estimated with the doubly robust (outcome regression + inverse-probability weighting) Callaway-Sant'Anna estimator. * $p<0.1$, ** $p<0.05$, *** $p<0.01$. Control group: not-yet-treated municipalities.

Figure 4.2: Event-study estimates: notable-person production (not-yet-treated control).



4.3 Scientific-elite Production

Table 4.3 narrows the notable-person measure to a single, discipline-specific construct: recognition as a scientist, through membership of the KNAW (the Royal Academy), as recorded in the KNAW's Digitaal Wetenschapshistorisch Centrum. The cleanest result is on the publication-weighted margin, which sums KNAW members' own linked Academy-proceedings publications: HBS access raised it significantly, and the treatment-immune 1851-baseline rate confirms the effect. The plain KNAW-membership headcount points the same way but is weaker, reaching significance only in the group-weighted aggregate and on the 1851-baseline rate, and, identified off the fewest treated clusters of any design, it does not clear the randomisation-inference test of Appendix A.6 (Table A.10); we therefore read the KNAW headcount as suggestive rather than established.[23]

Pre-treatment coefficients are generally small and statistically indistinguishable from zero. The one exception is a significant negative placebo on the KNAW headcount at the most distant and thinnest-celled pre-treatment point ($e = -20$); it does not recur on the publication-weighted margin, so we read it as edge-of-panel noise rather than a violation of parallel trends. This panel is built from far fewer in-window persons than the bioport sample, so, like the bioport result, we treat it as corroborating evidence on a specifically scientific margin rather than as a freestanding test. Robustness to the never-treated control, the full disaggregated event-study path, and the corresponding figure are reported in Appendix A.

Table 4.3: Notable-scientist DiD estimates (DWC/KNAW)

	KNAW membership (headcount)	KNAW membership (per 10,000 pop)	KNAW membership (per 10,000 1851 pop)	KNAW publication-weighted output	KNAW publication-weighted output (per 10,000 pop)	KNAW publication-weighted output (per 10,000 1851 pop)
e = -20	-0.2108**	-0.1382**	-0.1142**	-0.2551	-0.1626	-0.1396
	(0.0958)	(0.0594)	(0.0477)	(0.3315)	(0.1885)	(0.1621)
e = -15	-0.0752	-0.0422	-0.0477	-0.0021	-0.0039	-0.0044
	(0.0756)	(0.0475)	(0.0480)	(0.0031)	(0.0194)	(0.0190)
e = -10	0.0420	0.0283	0.0291	-0.0119*	-0.0563	-0.0601
	(0.1111)	(0.0586)	(0.0593)	(0.0071)	(0.0462)	(0.0456)
e = -5	0.0554	0.0154	0.0219	0.2340	0.0469	0.0563
	(0.0887)	(0.0543)	(0.0512)	(0.2356)	(0.1038)	(0.1148)
Post ATT[0,20] (Simple)	0.0618	0.0265	0.0460	0.5251**	0.2561**	0.3302**
	(0.0588)	(0.0263)	(0.0326)	(0.2324)	(0.1068)	(0.1456)
Post ATT[0,20] (Weighted)	0.1629**	0.0200	0.0886***	0.5057**	0.1925*	0.3200**
	(0.0782)	(0.0273)	(0.0302)	(0.2104)	(0.1122)	(0.1265)
N observations	17,178	17,178	16,900	17,178	17,178	16,900
N cohorts	8	8	7	8	8	7

"KNAW membership" counts persons born in a municipality who later appear in the KNAW's Digitaal Wetenschapshistorisch Centrum (dwc.knaw.nl) as a deceased KNAW member; "KNAW publication-weighted output" sums those members' own linked Academy-proceedings publication counts. Callaway and Sant'Anna (2021) staggered difference-in-differences on a panel of municipalities by 5-year birth cohort (year the child turned twelve). Unit of observation: municipality-cohort cell. Headcount columns are unweighted raw counts; per-10,000-pop columns scale by contemporaneous municipal population; per-10,000-1851-pop columns scale by each municipality's 1851 (pre-treatment) population, held fixed across all of that municipality's cohorts so it cannot itself be a treatment effect, following the bioport upper-tail table (Bell et al. 2019; Squicciarini and Voigtlander 2015). Rows $e < 0$ are pre-treatment placebo ATTs. Post ATT[0,20] (Simple) is the equal-weighted average of dynamic event-time ATTs over e = 0..20; (Weighted) is the treatment-group-size- weighted average. Bootstrapped SE, clustered by municipality (999 reps), in parentheses. Group-time effects are estimated with the doubly robust (outcome regression + inverse-probability weighting) Callaway-Sant'Anna estimator. * $p<0.1$, ** $p<0.05$, *** $p<0.01$. Control group: not-yet-treated municipalities.

[23] The DWC database also records membership of the broader Dutch learned societies (*genootschappen*), a lower-bar and noisier notion of scientific recognition than KNAW membership. Estimates for this outcome, reported in Appendix Tables A.19 and A.20, are mostly indistinguishable from zero and, if anything, slightly negative.

4.4 Human capital in the broader population

The genealogieonline father–son pairs let us ask whether the pipeline reached beyond the university-bound and into the broader population's occupational attainment. Table 4.4 reports pair-level estimates that condition on the father's HISCLASS, so each coefficient is the effect of HBS access given paternal origin. The point estimates are consistent in sign with a minor human-capital gain: the son's HISCAM occupational-status score and his probability of upward mobility are positive on average. But they are imprecisely estimated and cannot be distinguished from zero at conventional levels, and the same holds when we restrict to low-origin families (father HISCLASS > 6).[24] The imprecision is not surprising. Identification at the pair level rests on the small number of treated municipalities with genealogical coverage, whereas the municipality-level enrolment and notable-person panels pool the full population of a town rather than a linked subsample and are the better-powered specifications. We read the broad-population occupational effect as positive but not robust, consistent with the HBS operating primarily through the university-bound upper tail rather than through a wholesale shift of the occupational distribution.

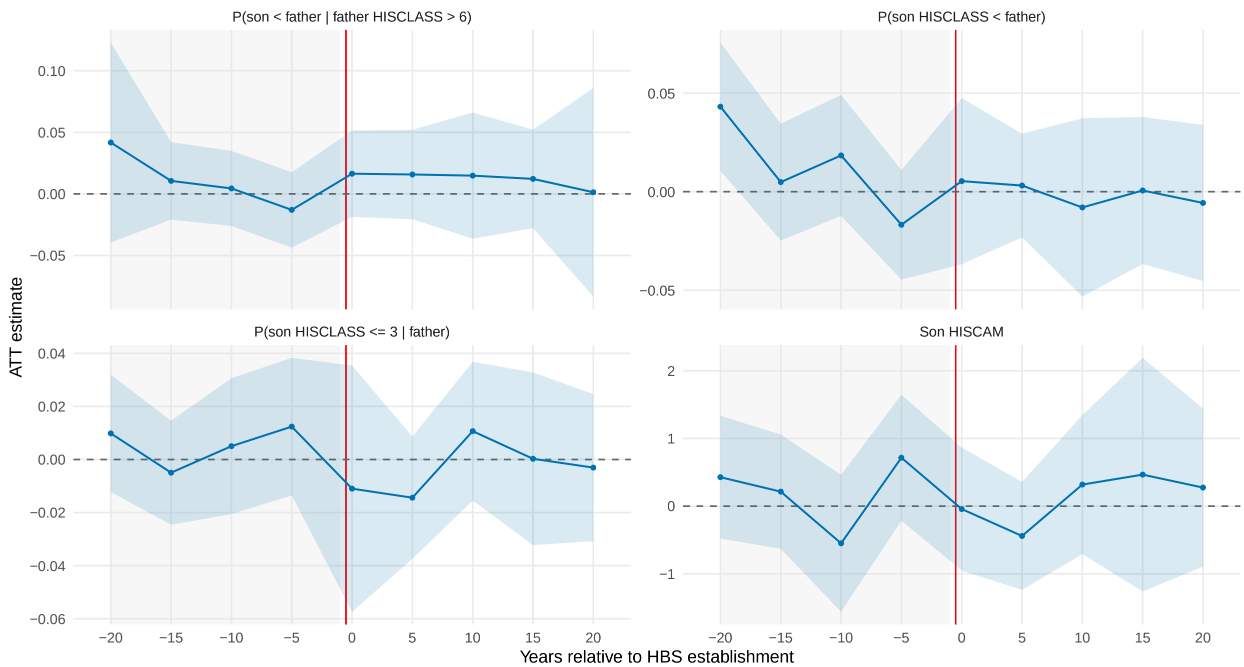
Table 4.4: Human capital in the broader population, DiD estimates

	P(son HISCLASS <= 3 | father)	P(son HISCLASS < father)	Son HISCAM	P(son < father | father HISCLASS > 6)
e = -20	0.0098	0.0431***	0.4279	0.0417
	(0.0082)	(0.0126)	(0.3435)	(0.0321)
e = -15	-0.0050	0.0049	0.2141	0.0106
	(0.0073)	(0.0115)	(0.3197)	(0.0124)
e = -10	0.0050	0.0184	-0.5493	0.0044
	(0.0095)	(0.0119)	(0.3837)	(0.0120)
e = -5	0.0124	-0.0168	0.7155**	-0.0130
	(0.0096)	(0.0108)	(0.3526)	(0.0121)
Post ATT[0,20] (Simple)	-0.0035	-0.0009	0.1150	0.0121
	(0.0069)	(0.0112)	(0.3282)	(0.0135)
Post ATT[0,20] (Weighted)	-0.0059	-0.0080	0.0614	0.0062
	(0.0071)	(0.0109)	(0.3127)	(0.0140)
N observations	237,834	237,834	237,834	207,502
N cohorts	8	8	8	8

Callaway and Sant'Anna (2021) staggered difference-in-differences on genealogieonline father-son pairs, binned to 5-year son cohorts (year the son turned twelve). Unit of observation: father-son pair. All specifications condition on the father's HISCLASS (lower HISCLASS = higher status), so the ATT is the effect of HBS access given paternal origin. The municipality-level HDNG religious shares used elsewhere are not included here: at this conditioning granularity they are collinear with the earliest treatment cohorts and produce a singular outcome regression. Column 4 restricts to low-origin families (father HISCLASS > 6) and therefore has a smaller sample. The sample is deduplicated to one row per father-son pair. Rows e < 0 are pre-treatment placebo ATTs. Post ATT[0,20] (Simple) is the equal-weighted average of dynamic event-time ATTs over e = 0..20; (Weighted) is the treatment-group-size-weighted average. Standard errors from a municipality-clustered multiplier bootstrap (999 reps) in parentheses. Group-time effects are estimated with the outcome-regression-only Callaway-Sant'Anna estimator: the doubly robust estimator is singular under this specification's father-HISCLASS-only covariate. * p<0.1, ** p<0.05, *** p<0.01. Control group: not-yet-treated municipalities.

[24]These pair-level estimates are sensitive to the cohort bin width. Here we bin sons into five-year cohorts so that the pre-treatment event-time grid is identical to the municipality-level Tables 4.1 and 4.2; a finer three-year binning yields a larger and statistically significant Son HISCAM effect. We report the coarser, harmonised specification and read the broad-population occupational effect as positive but not robust.

Figure 4.3: Event-study estimates: broad-population occupational attainment, conditional on father's HISCLASS (not-yet-treated control).



4.5 Interpretation

Taken together, the three layers trace a coherent, if attenuating, causal chain. The HBS, designed as a middle-class pathway to university and the professions, raised the university-enrolment rate in treated municipalities, especially in the STEM fields its curriculum targeted. That institutional pipeline is echoed at the extreme right tail, where treated municipalities produced more biographically notable persons per capita, and the narrower DWC/KNAW measure corroborates this on the scientific margin specifically, most clearly on the publication-weighted outcome. At the broad-population occupational margin the effect is positive in sign but too imprecisely estimated to distinguish from zero, consistent with the HBS operating primarily through the university-bound upper tail rather than through a broad shift of the occupational distribution.

This pattern connects directly to the upper-tail human capital literature. Squicciarini and Voigtländer (2015) show that upper-tail knowledge, not average literacy, predicts industrial growth; our results identify a specific institutional mechanism through which that upper tail is produced, namely secondary schooling with a modern, science-oriented curriculum. Our findings are the mirror image of Squicciarini (2020), who finds that French districts slow to adopt a technical curriculum, held back by religiosity, saw weaker subsequent industrial development: where a science-oriented curriculum was withheld, development lagged; where the HBS supplied one, the upper tail thickened. Dittmar and Meisenzahl (2020) use *Deutsche Biographie* entries as their outcome, and our notable-person estimates provide a comparable link between institutional change and biographical notability. Finally, Mokyr (2016) argues that the "Industrial Enlightenment" required a cadre of educated practitioners below the level of the great inventors; the HBS, with its practical, science-oriented curriculum aimed at the middle class, is a near-textbook example of the institution Mokyr describes.

4.6 Caveats

A few caveats apply. First, our treatment is a binary indicator for the presence of any HBS in the AMCO municipality; a continuous distance-to-nearest-HBS specification, which we plan to implement in future work, would refine the estimates by incorporating access to schools in neighbouring municipalities and would address the SUTVA concern that nominally untreated municipalities partly accessed treatment through commuting and boarding. Second, the genealogieonline sample overrepresents families whose records have been digitised and linked; the stability of the (null) pair-level results across control groups suggests any selection is not correlated with the timing of HBS openings, but the pair-level design is under-powered. Finally, the DWC/KNAW panel is built from only 682 in-window persons, an order of magnitude smaller than the bioport sample; the resulting estimates are noisier and, on the plain KNAW-membership headcount, significant only in the group-weighted aggregate, so we treat the DWC/KNAW result as corroborating evidence on a specifically scientific margin rather than as a fully powered test.

5 Conclusion

This paper provides the first modern causal estimate of the effect of the Hogere Burgerschool, the nineteenth-century Dutch innovation in middle-class secondary education, on the formation of upper-tail human capital. Combining a newly constructed geo-coded panel of HBS establishments with five university enrolment registers, 338,880 father–son occupational pairs from genealogieonline, and the staggered difference-in-differences estimator of Callaway and Sant'Anna (2021), we trace the effect of local HBS access across three layers of the human-capital distribution.

The three layers tell a consistent story of an effect that runs through the top of the distribution and fades as one moves toward its centre. At the institutional-pipeline layer, HBS establishment raised university enrolment in treated municipalities, by roughly sixteen students per cohort and about 3.7 per $10{,}000$ inhabitants, more than doubling the enrolment rate of never-treated towns. The gain is concentrated in the STEM faculties, a direct reflection of the HBS's deliberately beta-heavy curriculum. At the extreme-upper-tail layer, treated municipalities went on to produce more biographically notable persons per capita, and the narrower KNAW measure corroborates this on the scientific margin specifically. At the broad-population occupational layer, by contrast, the father–son estimates are positive in sign but too imprecise to distinguish from zero, indicating that the HBS worked primarily through the university-bound upper tail rather than through a wholesale shift of the occupational distribution. The municipality-level effects build over event time as treated cohorts pass through the school and enter the labour market, exactly as an educational-pipeline mechanism predicts, and the pre-treatment coefficients are broadly, though not perfectly, flat; the mature enrolment-count effect survives formal sensitivity analysis for parallel-trends violations, and alternative staggered-adoption estimators reproduce the sign and magnitude of the headline estimates.

These results connect to four strands of the literature. First, they extend the upper-tail human-capital framework of Squicciarini and Voigtländer (2015) and Dittmar and Meisenzahl (2020) from the pre-industrial and early-modern periods to the formative era of modern educational systems, identifying a specific institutional mechanism, science-oriented secondary schooling, through which upper-tail knowledge is produced. Second, they speak to Mokyr (2016)'s argument that the Industrial Enlightenment required a cadre of educated practitioners situated below the level of the great inventors: the HBS, aimed at the commercial and industrial middle class rather than the classical elite, is precisely such an institution, and the pattern we find, a sharp effect on the university-bound tail alongside a null on broad occupational attainment, is consistent with an institution that thickened this specialised layer rather than lifting the occupational distribution as a whole. Third, they extend the institutional-knowledge-transmission framework of de la Croix et al. (2018), who study clan-based and guild-based apprenticeship, to the modern institutional form of formal secondary schooling. Fourth, they place a plausible causal weight behind the case-study claim of Willink (1998) that the HBS was an engine of late-nineteenth-century Dutch elite formation, and locate the school within the broader rise in Dutch intergenerational mobility documented by Knigge et al. (2014b,a) and Mandemakers (1996).

Several avenues for refinement remain. A continuous distance-to-nearest-HBS treatment specification would address the SUTVA concern that nominally untreated municipalities partly accessed treatment through commuting and boarding, and would likely sharpen the imprecise pair-level occupational estimates. Extending the biografischportaal births series before 1851 would let the births-scaled notable-person outcome be estimated on a fully balanced pre-treatment window. The institutional distinction between rijks and gemeentelijke HBS suggests an instrumental-variable strategy in which rijks placement instruments for total HBS exposure, and the Catholic resistance to HBS openings in Limburg and Noord-Brabant offers a plausibly exogenous source of variation in adoption timing. We see this paper as a first step in a broader research programme that uses the Dutch nineteenth-century institutional landscape to identify the causal determinants of upper-tail human capital formation in a setting where modern administrative data are not available.

References

Bartels, A. (1947) *75 jaar middelbaar onderwijs, 1863–1938*, Groningen: J.B. Wolters.

Bell, Alexander M., Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen (2019) "Who Becomes an Inventor in America? The Importance of Exposure to Innovation," *Quarterly Journal of Economics*, 134 (2), 647–713, 10.1093/qje/qjy028.

Boekholt, P. Th. F. M. and E. P. de Booy (1987) *Geschiedenis van de school in Nederland vanaf de middeleeuwen tot aan de huidige tijd*, Assen/Maastricht: Van Gorcum.

Bolt, Jutta and Jan Luiten van Zanden (2020) "Maddison Style Estimates of the Evolution of the World Economy: A New 2020 Update," technical report, Maddison Project Database.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess (2024) "Revisiting Event-Study Designs: Robust and Efficient Estimation," *The Review of Economic Studies*, 91 (6), 3253–3285, 10.1093/restud/rdae007.

Callaway, Brantly and Pedro H. C. Sant'Anna (2021) "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 225 (2), 200–230, 10.1016/j.jeconom.2020.12.001.

Cantoni, Davide and Noam Yuchtman (2014) "Medieval Universities, Legal Institutions, and the Commercial Revolution," *Quarterly Journal of Economics*, 129 (2), 823–887, 10.1093/qje/qju007.

Cappelli, Gabriele, Leonardo Ridolfi, Michelangelo Vasta, and Johannes Westberg (2023) "Human Capital in Europe, 1830s–1930s: A General Survey," *Journal of Economic Surveys*, 10.1111/joes.12606.

Capponi, G. and K. Frenken (2021) *On the sudden rise of Dutch science at the end of the nineteenth century: a core-periphery approach*, Vol. 28 (9), 10.1080/13662716.2021.1929867.

de Chaisemartin, Clément and Xavier D'Haultfœuille (2023) "Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey," *The Econometrics Journal*, 26 (3), C1–C30, 10.1093/ectj/utac017.

de la Croix, David, Frédéric Docquier, Alice Fabre, and Robert Stelter (2024) "The Academic Market and the Rise of Universities in Medieval and Early Modern Europe (1000–1800)," *Journal of the European Economic Association*, 22 (4), 1541–1589, 10.1093/jeea/jvad061.

de la Croix, David, Matthias Doepke, and Joel Mokyr (2018) "Clans, Guilds, and Markets: Apprenticeship Institutions and Growth in the Preindustrial Economy," *Quarterly Journal of Economics*, 133 (1), 1–70, 10.1093/qje/qjx026.

de la Croix, David and Omar Licandro (2015) "The Longevity of Famous People from Hammurabi to Einstein," *Journal of Economic Growth*, 20 (3), 263–303, 10.1007/s10887-015-9117-0.

DBNL (2026) "Heike Kamerlingh Onnes," https://www.dbnl.org/tekst/delf006heik01_01/delf006heik01_01_0005.php, Digitale Bibliotheek voor de Nederlandse Letteren. Accessed 2026-07-01.

Dittmar, Jeremiah and Ralf R. Meisenzahl (2020) "Public Goods Institutions, Human Capital, and Growth: Evidence from German History," *Review of Economic Studies*, 87 (2), 959–996, 10.1093/restud/rdz002.

Erasmus School of Economics (2026) "Jan Tinbergen (1903-1994)," https://tinbergenletters.eur.nl/tinbergen/, Tinbergen Letters. Accessed 2026-07-01.

Goodman-Bacon, Andrew (2021) "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*, 225 (2), 254–277, 10.1016/j.jeconom.2021.03.014.

Huis van de Nijmeegse Geschiedenis (2026) "Gemeentelijke Hogere Burgerschool voor Meisjes en Middelbare Meisjesschool," https://www.huisvandenijmeegsegeschiedenis.nl/info/Gemeentelijke_Hogere_Burgerschool_voor_Meisjes_en_Middelbare_Meisjesschool, Accessed 2026-05-15.

Huygens ING (2026a) "Brouwer, Luitzen Egbertus Jan (1881-1966)," https://resources.huygens.knaw.nl/bwn1880-2000/lemmata/bwn2/brouwerle, Biografisch Woordenboek van Nederland. Accessed 2026-07-01.

——— (2026b) "Lorentz, Hendrik Antoon (1853-1928)," https://resources.huygens.knaw.nl/bwn1880-2000/lemmata/bwn1/lorentz, Biografisch Woordenboek van Nederland. Accessed 2026-07-01.

——— (2026c) "Zeeman, Pieter (1865-1943)," https://resources.huygens.knaw.nl/bwn1880-2000/lemmata/bwn1/zeeman, Biografisch Woordenboek van Nederland. Accessed 2026-07-01.

Knigge, Antonie, Ineke Maas, and Marco H. D. van Leeuwen (2014a) "Sources of Sibling (Dis)similarity: Total Family Impact on Status Variation in the Netherlands in the Nineteenth Century," *American Journal of Sociology*, 120 (3), 908–948, 10.1086/679104.

Knigge, Antonie, Ineke Maas, Marco H. D. van Leeuwen, and Kees Mandemakers (2014b) "Status Attainment of Siblings during Modernization," *American Sociological Review*, 79 (3), 549–574, 10.1177/0003122414529586.

van Leeuwen, Marco H. D. and Ineke Maas (2011) *HISCLASS: A Historical International Social Class Scheme*, Leuven: Leuven University Press.

Mandemakers, Cees (1996) *Gymnasiaal en middelbaar onderwijs: Ontwikkeling, structuur, sociale achtergrond en schoolprestaties in Nederland, ca. 1800–1968*, Ph.D. dissertation, Erasmus Universiteit Rotterdam.

Meisenzahl, Ralf R. and Joel Mokyr (2016) "The Rate and Direction of Invention in the British Industrial Revolution: Incentives and Institutions," *In The Rate and Direction of Inventive Activity Revisited*, 443–481.

Mokyr, Joel (2009) *The Enlightened Economy: An Economic History of Britain 1700–1850*, New Haven: Yale University Press.

——— (2016) *A Culture of Growth: The Origins of the Modern Economy*, Princeton: Princeton University Press.

NobelPrize.org (2026) "All Nobel Prizes in Physics; All Nobel Prizes in Chemistry; All Nobel Prizes in Physiology or Medicine," https://www.nobelprize.org/prizes/lists/all-nobel-prizes/, Accessed 2026-07-01.

Oreopoulos, Philip, Marianne E. Page, and Ann Huff Stevens (2006) "The Intergenerational Effects of Compulsory Schooling," *Journal of Labor Economics*, 24 (4), 729–760, 10.1086/506484.

Rambachan, Ashesh and Jonathan Roth (2023) "A More Credible Approach to Parallel Trends," *The Review of Economic Studies*, 90 (5), 2555–2591, 10.1093/restud/rdad018.

Schravenlant 150 jaar (2026) "19e eeuw: hbs en gymnasium voor jongens, en de meisjes dan?" https://www.schravenlant150jaar.nl/c-4714202/19e-eeuw-hbs-en-gymnasium-voor-jongens-en-de-meisjes-dan/, Accessed 2026-05-15.

Squicciarini, Mara P. (2020) "Devotion and Development: Religiosity, Education, and Economic Progress in Nineteenth-Century France," *American Economic Review*, 110 (11), 3454–3491, 10.1257/aer.20191054.

Squicciarini, Mara P. and Nico Voigtländer (2015) "Human Capital and Industrialization: Evidence from the Age of Enlightenment," *Quarterly Journal of Economics*, 130 (4), 1825–1883, 10.1093/qje/qjv025.

Sun, Liyang and Sarah Abraham (2021) "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 225 (2), 175–199, 10.1016/j.jeconom.2020.09.006.

Willink, Bastiaan (1998) *De tweede Gouden Eeuw: Nederland en de Nobelprijzen voor natuurwetenschappen 1870–1940*, Amsterdam: Uitgeverij Bert Bakker.

A Appendix: Robustness

A.1 Additional institutional detail

This subsection collects institutional detail on the HBS that Section 2.1 summarises only briefly, to keep the main text focused on the features of the rollout that bear directly on identification.

A parallel girls' secondary school (*Middelbare Meisjesschool*, MMS) network, oriented to a more domestic and language-heavy curriculum, opened alongside the HBS: the first MMS opened in Haarlem in 1867 and the network expanded slowly thereafter, while the HBS itself was not opened to girls until 1906 (Schravenlant 150 jaar, 2026; Huis van de Nijmeegse Geschiedenis, 2026).

Local champions could accelerate an HBS opening well beyond what the population-threshold and national-negotiation forces described in Section 2.1 would predict on their own. The clearest example is Tilburg, where the HBS was housed in King Willem II's former palace: the king's estate transferred the building to the municipality on condition that an HBS be established in it (Willink, 1998).

The HBS's science orientation was not confined to the timetable. An 1867 timetable shows that in the upper grades, mathematics, physics, chemistry, botany and zoology, mechanics, mineralogy and geology, and cosmography together accounted for approximately sixty of the 164 total weekly class hours across the five-year program, nearly half the curriculum (Willink, 1998). The school also invested heavily in laboratory infrastructure: on the basis of inspector Bosscha's reports, Willink (1998) notes that HBS laboratories were often as well-equipped as university laboratories, a material complement to the beta-heavy curriculum discussed in the main text.

Three considerations, beyond the absence of an external abolition date, motivate ending the analysis sample in 1900 rather than continuing to follow the HBS through its twentieth-century history. First, the systematic municipality-level rollout data constructed for this paper, drawn from the editions of the Dutch State Almanac (*Staatsalmanak voor het Koningrijk der Nederlanden*), naturally end at the turn of the century. Second, by 1900 the HBS network had largely reached the saturation level that defines the upper plateau of the rollout: most subsequent expansion added sections or capacity to existing schools rather than opening HBS institutions in previously unserved municipalities, which would weaken the staggered-adoption design. Third, the institutional environment changed materially after 1900: the 1906 opening of the HBS to girls, the 1917 Limburg Initiative Act granting HBS graduates direct university access, and the 1921 split into humanities- and science-oriented streams each altered what HBS attendance meant for a pupil, in ways that would confound the treatment definition used here.

The Dutch nineteenth-century setting is additionally well suited to a municipality-level staggered rollout design because of the scale and richness of the underlying variation: the country contained roughly 1,120 municipalities by 1900, with strong regional variation in industrialisation, religious composition, and economic structure; the HBS itself was a single, well-defined institutional innovation with clearly dated openings, sharply distinguishable from the pre-existing Latin schools and gymnasia; and the combination of digitised genealogical pairs (genealogieonline) with an occupational coding scheme (van Leeuwen and Maas, 2011) permits intergenerational mobility measurement at scale, as described further in Section 3.

The main-text estimates (Tables 4.1–4.4 and 4.3) use the not-yet-treated control group and summarise the post-treatment path in two aggregated average treatment effects. The remainder of this appendix reports two robustness variants for each of the four difference-in-differences tables: (i) re-estimation against the never-treated control group, and (ii) the full disaggregated post-treatment event-study path in place of the two aggregated rows.

A.2 Propensity-score overlap

Section 3.4 notes that only 55 of 1,224 municipalities are ever treated and that they are systematically the larger towns favoured by the ten-thousand-inhabitant threshold. Table A.1 fits a cross-sectional logit of ever-treated status on pre-treatment (1840) municipality covariates; Table A.2 and Figure A.1 summarise the resulting propensity-score distributions of ever-treated and never-treated municipalities and report the share of never-treated municipalities whose fitted score falls inside the ever-treated group's own support. Only a minority of never-treated municipalities are inside that support, so any comparison drawing on the full never-treated reservoir leans on the doubly robust estimator's propensity-score down-weighting (and, for units far outside the support, its model-based extrapolation) rather than on direct comparison. The not-yet-treated comparison used for the main-text estimates faces this problem in attenuated form rather than escaping it: its reservoir contains all never-treated municipalities alongside the later adopters, and only the latter resemble the treated towns in scale. Section A.3 therefore re-estimates the headline outcomes on control reservoirs from which the non-comparable never-treated municipalities are removed outright.

Table A.1: Logit of ever-treated on pre-treatment (1840) municipality covariates

	Coefficient
Intercept	-35.028***
	(5.471)
Log 1851 population	3.865***
	(0.572)
Log population density	0.784***
	(0.170)
Share Catholic	-3.428
	(2.978)
Share Dutch Reformed	-4.286
	(3.134)

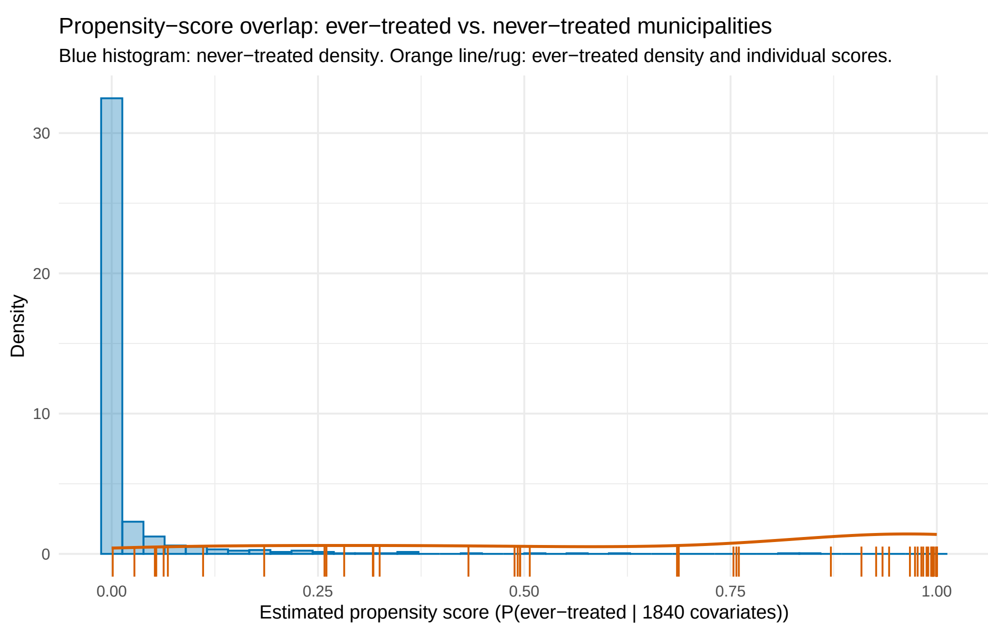
Cross-sectional logit, one row per municipality (N = 901 , of which 51 ever treated). Dependent variable: 1 if the municipality ever received an HBS by 1898. Covariates measured at the 1840 HDNG cross-section, twenty years before the earliest HBS opening (1860), so none can itself be a treatment effect. The Gereformeerd religious share is omitted: the Gereformeerde Kerken did not exist until the 1892 Doleantie split, so the share is uniformly zero in 1840. Standard errors in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A.2: Propensity-score overlap: treated vs. never-treated municipalities

	Value
Treated municipalities (N)	51
Never-treated municipalities (N)	850
Treated p-score range	[0.001, 1.000]
Never-treated inside treated support	327 / 850 (38.5

Propensity scores from the logit in Table A.1. "Inside treated support" is the share of never-treated municipalities whose fitted propensity score falls within the range spanned by the 55 ever-treated municipalities' own fitted scores.

Figure A.1: Propensity-score overlap: ever-treated vs. never-treated municipalities



A.3 Trimmed and matched control reservoirs

Tables A.3 and A.4 re-estimate the headline university-enrolment and notable-person outcomes with the never-treated part of the control reservoir restricted to municipalities comparable to the treated towns on pre-treatment (1851) population, while ever-treated municipalities (as treated units and as not-yet-treated controls) are always retained. Four restrictions are reported alongside the untrimmed reference column: never-treated municipalities with at least 5,000 inhabitants in 1851 (55 municipalities), at least 10,000 (only two municipalities), a one-to-one nearest-neighbour match on 1851 population (54 municipalities), and a one-to-one nearest-neighbour match on the fitted propensity score of ever-treated status from the logit of Table A.1 (51 municipalities). The 10,000 cut, the Act's own placement margin, makes vivid how structurally thin the comparable-untreated reservoir is: essentially every town of that size eventually adopted an HBS, so that column compares adopters almost exclusively with later adopters, the closest feasible approximation to an adopters-only design.

The headline results survive against comparable untreated towns. The enrolment count effect declines from 16.0 students per municipality-cohort in the untrimmed panel to 14.2 under the 5,000 cut and 14.4 and 14.0 in the population- and propensity-score-matched samples, and to 11.8 in the near-adopters-only 10,000 column, remaining significant at the one-percent level throughout; the population-scaled enrolment rate follows the same pattern (3.7 to roughly 2.7–3.2 across the trimmed and matched columns, and 2.3 at the 10,000 cut). Reading the untrimmed and restricted columns together, between a tenth and a quarter of the headline enrolment estimate is attributable to the comparison with small villages, and the remainder stands against towns of comparable scale. The notable-person estimates are, if anything, slightly larger against comparable towns (for example, the 1851-baseline-scaled weighted ATT rises from 0.27 to 0.34 under the 5,000 cut, 0.35 in the population-matched sample, and 0.32 in the propensity-score-matched sample), though they weaken in the thin 10,000 column. The trimming also bears on the pre-treatment path: the one significant enrolment-count placebo in the untrimmed panel ($e = -10$) attenuates and loses significance in every restricted column, consistent with part of that placebo reflecting the small-village comparison rather than a differential trend among comparable towns.

Table A.3: University enrollment: trimmed and matched control reservoirs

	All controls	NT ≥ 5,000	NT ≥ 10,000	NN matched	PS matched
Panel A: Students (count)					
e = -20	0.0995	-0.6097	-1.5749	-0.6316	-0.7639
	(1.0037)	(1.0086)	(1.0127)	(1.0166)	(1.0107)
e = -15	1.3796	0.4223	-0.4336	0.4919	0.4615
	(2.2236)	(2.2282)	(2.1126)	(2.2669)	(2.1854)
e = -10	1.7310***	1.0071	0.4140	1.0177	0.9504
	(0.6717)	(0.6864)	(0.8756)	(0.7163)	(0.7297)
e = -5	1.7215	1.3024	0.6074	1.3164	1.2766
	(1.1236)	(1.1441)	(1.2021)	(1.1880)	(1.1789)
Post ATT[0,20] (Simple)	16.0457***	14.1698***	11.7887***	14.3784***	14.0322***
	(3.9196)	(3.7902)	(4.1579)	(3.8394)	(3.4880)
Post ATT[0,20] (Weighted)	20.6973***	19.0604***	17.6453***	19.2048***	18.8473***
	(4.7383)	(4.7640)	(4.7398)	(4.7460)	(4.6967)
N observations	13,863	1,398	714	1,363	1,337
N cohorts	8	8	8	8	8
N never-treated controls	1,169	55	2	54	51
Panel B: Students / 10,000 pop					
e = -20	-0.4198	-0.5128	-1.0654	-0.5664	-0.8656
	(1.0517)	(1.1736)	(1.2654)	(1.1495)	(1.2755)
e = -15	0.7531	0.3307	0.0470	0.4092	0.4046
	(1.9367)	(1.9204)	(2.0889)	(2.0283)	(2.0312)
e = -10	1.8695**	1.6949*	1.6921*	1.6836*	1.5948*
	(0.9030)	(0.9193)	(0.9078)	(0.8713)	(0.9518)
e = -5	0.4333*	0.2951	-0.3344	0.2768	0.2247
	(0.2589)	(0.3062)	(0.4546)	(0.3018)	(0.2883)
Post ATT[0,20] (Simple)	3.7212***	3.0749***	2.3091*	3.2258***	2.6988***
	(0.6821)	(0.7640)	(1.1991)	(0.7879)	(0.7865)
Post ATT[0,20] (Weighted)	3.4595***	3.0716***	2.9697**	3.1751***	2.7132**
	(1.1604)	(1.0908)	(1.2911)	(1.1177)	(1.1350)
N observations	13,863	1,398	714	1,363	1,337
N cohorts	8	8	8	8	8
N never-treated controls	1,169	55	2	54	51

Callaway and Sant'Anna (2021) staggered difference-in-differences on the municipality x 5-year birth-cohort panel, not-yet-treated control group, doubly robust group-time estimator conditioning on the three HDNG religion shares, as in the main tables. Each column restricts the never-treated part of the control reservoir; ever-treated municipalities (as treated units and as not-yet-treated controls) are always retained. All controls: untrimmed panel, replicating the main design. NT >= 5,000 / >= 10,000: never-treated municipalities kept only if their 1851 (pre-treatment) population is at least 5,000 / 10,000; only two never-treated municipalities clear 10,000, so that column compares adopters almost exclusively with later adopters. NN matched: never-treated municipalities kept only if selected as the 1:1 nearest-neighbour match (greedy, without replacement, on 1851 population) of a treated municipality. PS matched: as NN matched, but matching on the fitted propensity score of ever-treated status from the logit of Table A.1 (log 1851 population, log 1840 population density, Catholic and Dutch Reformed shares). Never-treated municipalities without the relevant matching

Table A.4: Notable-person production: trimmed and matched control reservoirs

	All controls	NT ≥ 5,000	NT ≥ 10,000	NN matched	PS matched
Panel A: Bioport upper-tail (count)					
e = -20	-0.1276	-0.0801	0.0518	-0.0634	-0.0567
	(0.1689)	(0.1807)	(0.2743)	(0.1859)	(0.1864)
e = -15	-0.5891**	-0.5842**	-0.5571**	-0.5630**	-0.5905**
	(0.2310)	(0.2555)	(0.2638)	(0.2321)	(0.2427)
e = -10	0.2942	0.4602	0.7106*	0.4365	0.4516
	(0.3397)	(0.3803)	(0.4230)	(0.3728)	(0.3646)
e = -5	-0.2444	-0.2588	-0.3115	-0.2921	-0.2539
	(0.2290)	(0.2465)	(0.3040)	(0.2212)	(0.2608)
Post ATT[0,20] (Simple)	0.2937	0.3116*	0.1957	0.3183	0.2948
	(0.1871)	(0.1830)	(0.1988)	(0.2089)	(0.1952)
Post ATT[0,20] (Weighted)	0.6035*	0.6201*	0.5130	0.6236*	0.6031*
	(0.3610)	(0.3529)	(0.3517)	(0.3546)	(0.3594)
N observations	17,178	1,714	879	1,677	1,649
N cohorts	8	8	8	8	8
N never-treated controls	1,169	55	2	54	51
Panel B: Bioport upper-tail / 10k pop (1851 baseline)					
e = -20	-0.0183	-0.0311	0.0474	-0.0115	0.0228
	(0.1001)	(0.1387)	(0.2214)	(0.1487)	(0.1389)
e = -15	-0.3045**	-0.3180**	-0.2934*	-0.2819*	-0.3247**
	(0.1247)	(0.1476)	(0.1724)	(0.1466)	(0.1372)
e = -10	0.2375	0.3806*	0.5522**	0.3524*	0.3827*
	(0.1857)	(0.2028)	(0.2371)	(0.2127)	(0.2252)
e = -5	-0.1229	-0.1600	-0.2120	-0.2173	-0.1487
	(0.1372)	(0.1714)	(0.1996)	(0.1597)	(0.1718)
Post ATT[0,20] (Simple)	0.1457*	0.2023**	0.0533	0.2172*	0.1760*
	(0.0830)	(0.0959)	(0.1077)	(0.1128)	(0.0949)
Post ATT[0,20] (Weighted)	0.2734***	0.3444***	0.2124*	0.3546***	0.3159***
	(0.0897)	(0.1066)	(0.1164)	(0.1238)	(0.0951)
N observations	16,900	1,710	875	1,673	1,645
N cohorts	7	7	7	7	7
N never-treated controls	1,169	55	2	54	51

Callaway and Sant'Anna (2021) staggered difference-in-differences on the municipality x 5-year birth-cohort panel, not-yet-treated control group, doubly robust group-time estimator conditioning on the three HDNG religion shares, as in the main tables. Each column restricts the never-treated part of the control reservoir; ever-treated municipalities (as treated units and as not-yet-treated controls) are always retained. All controls: untrimmed panel, replicating the main design. NT >= 5,000 / >= 10,000: never-treated municipalities kept only if their 1851 (pre-treatment) population is at least 5,000 / 10,000; only two never-treated municipalities clear 10,000, so that column compares adopters almost exclusively with later adopters. NN matched: never-treated municipalities kept only if selected as the 1:1 nearest-neighbour match (greedy, without replacement, on 1851 population) of a treated municipality. PS matched: as NN matched, but matching on the fitted propensity score of ever-treated status from the logit of Table A.1 (log 1851 population, log 1840 population density, Catholic and Dutch Reformed shares). Never-treated municipalities without the relevant matching

A.4 University register composition: excluding the Universiteit van Amsterdam

Section 4.1 attributes the imprecise births-scaled enrolment rate in Table 4.1 partly to the births denominator growing noisier once the large, heavily urban Universiteit van Amsterdam graduate register, the one non-matriculation register among the five pooled in the baseline panel, is added to the four true matriculation registers (Leiden, Utrecht, Groningen, VU). Table A.5 tests this directly, re-estimating all three headline enrolment margins on a four-register panel that drops the Universiteit van Amsterdam register outright, alongside the five-register baseline for comparison.

The extensive-margin results are robust to the exclusion: the enrolment count falls from 16.0 to 12.6 (Simple) and the population-scaled rate from 3.7 to 3.0, both remaining significant at the one-percent level. The births-scaled rate, by contrast, firms up exactly as the attribution predicts: the point estimate rises from an insignificant 1.0 in the five-register panel to a significant 1.5 (Simple, five-percent level) and 1.6 (Weighted, ten-percent level) once the Universiteit van Amsterdam register is excluded. This corroborates the main text's reading that the births-scaled null reflects register-composition noise rather than the absence of an intensive-margin effect, while confirming that the paper's headline extensive-margin estimates do not depend on pooling all five registers.

Table A.5: University enrollment: excluding the Universiteit van Amsterdam register

	Five registers	Excl. UvA
Panel: Students (count)		
e = -20	0.0995	0.1608
	(1.0037)	(0.2362)
e = -15	1.3796	0.7298
	(2.2236)	(1.1140)
e = -10	1.7310***	1.0635**
	(0.6717)	(0.5405)
e = -5	1.7215	0.7124**
	(1.1236)	(0.3527)
Post ATT[0,20] (Simple)	16.0457***	12.5744***
	(3.9196)	(2.3890)
Post ATT[0,20] (Weighted)	20.6973***	18.4256***
	(4.7383)	(3.9581)
N observations	13,863	13,863
N cohorts	8	8

Continued on next page

Table A.5: University enrollment: excluding the Universiteit van Amsterdam register (Continued)

	Five registers	Excl. UvA
Panel: Students / 1,000 births		
e = -20	-0.9270	-0.1564
	(0.8089)	(0.2447)
e = -15	-0.0029	0.3033
	(1.7457)	(0.6925)
e = -10	1.1432	0.8882
	(1.3513)	(0.8532)
e = -5	-0.5810	-0.6587*
	(0.6164)	(0.3583)
Post ATT[0,20] (Simple)	1.0027	1.4797**
	(1.1445)	(0.6526)
Post ATT[0,20] (Weighted)	0.7341	1.6289*
	(1.4557)	(0.8551)
N observations	12,749	12,749
N cohorts	7	7
Panel: Students / 10,000 pop		
e = -20	-0.4198	-0.1533
	(1.1554)	(0.3711)
e = -15	0.7531	0.3750
	(1.8962)	(1.1246)
e = -10	1.8695**	1.0992
	(0.8961)	(0.6999)
e = -5	0.4333*	0.1665
	(0.2569)	(0.1028)
Post ATT[0,20] (Simple)	3.7212***	3.0353***
	(0.7241)	(0.9291)
Post ATT[0,20] (Weighted)	3.4595***	3.3124***
	(1.0831)	(1.0504)
N observations	13,863	13,863
N cohorts	8	8

Callaway and Sant'Anna (2021) staggered difference-in-differences on the municipality x 5-year birth-cohort panel, not-yet-treated control group, doubly robust group-time estimator conditioning on the three HDNG religion shares, as in Table 4.1. "Five registers" reproduces the all-faculty baseline (Leiden, Utrecht, Groningen, VU, Universiteit van Amsterdam). "Excl. UvA" drops the Universiteit van Amsterdam graduate register – the one non-matriculation register among the five – and re-estimates on the remaining four matriculation registers. The count is unweighted; the rate per 1,000 births is weighted

A.5 Pre-trend sensitivity and alternative estimators

Section 4.1 documents that the enrolment pre-treatment placebos are not uniformly zero. This subsection quantifies what such deviations could mean for the post-treatment estimates in two complementary ways: formal sensitivity bounds that ask how large a post-treatment violation of parallel trends the data could conceal, and re-estimation of the six headline margins under alternative staggered-adoption estimators and an augmented conditioning set.

Table A.6 applies the sensitivity analysis of Rambachan and Roth (2023) to the dynamic event studies underlying Tables 4.1 and 4.2, re-fit with a universal base period so that every coefficient is measured relative to $e = -5$. The relative-magnitudes restriction allows the post-treatment violation of parallel trends to be up to $\bar{M}$ times the largest observed pre-treatment deviation; the "breakdown" $\bar{M}$ is the largest multiple at which the robust confidence set still excludes zero. Read at the effect's mature horizon ($e = +10$), the paper's headline extensive margin clears the test with room to spare: the all-faculty enrolment count survives violations up to $\bar{M} = 1.7$ times the worst pre-treatment deviation, comfortably above the natural $\bar{M} = 1$ benchmark (the requirement that the effect withstand a post-treatment violation as large as anything actually observed before treatment), and the STEM count breaks down just below that benchmark ($\bar{M} = 0.9$). The population-scaled rates break down earlier ($\bar{M}$ of 0.1–0.3), but the contrast with the counts is largely mechanical rather than evidence of confounding specific to the intensive margin. Relative-magnitude bounds scale with the largest pre-treatment deviation, and the counts and rates carry essentially the same one, the $e = -10$ placebo of Table A.15, so the margin with the smaller effect-to-placebo ratio necessarily breaks down at a smaller multiple. Moreover, what the bound is anchored on matters as much as where it breaks: Section A.3 shows that this anchoring deviation attenuates once the never-treated reservoir is restricted to comparable towns, sharply on the count margin (where it loses significance in every restricted column) and more modestly on the rate margins, so on the count margin, at least, the anchor reflects a small-village comparison artifact rather than a differential trend among the comparisons that identify the effect; Table A.7 below pursues this logic quantitatively. The smoothness restriction, which instead allows the confound to follow an approximately linear differential trend with per-period nonlinearity at most M, is more demanding still: at $e = +10$ the all-faculty count retains significance for nonlinearities up to $M = 0.9$ students per period, the STEM count survives an exactly linear extrapolated trend ($M = 0$) but no additional nonlinearity, and a linearly extrapolated differential trend alone could account for the population-scaled rate effects. The impact-period ($e = 0$) effects are small by construction, since the first treated cohort reaches university age only about six years after the school opens, and are fragile throughout. For the notable-person margins the event-time-specific confidence intervals already include zero, so those rows are reported for completeness but the bounds carry little additional information; this is neutral rather than adverse, since the sensitivity analysis applies to individual event-time coefficients and the significant notable-person results of Table 4.2 are the aggregated post-treatment averages, which the bounds do not address. In sum, the sensitivity analysis supports the enrolment effect as measured by the count of students a municipality sends to university outright, while the intensive-margin rates rest more heavily on the parallel-trends assumption holding exactly. They do not stand or fall with these bounds alone, however: they are independently corroborated by the town-size-conditioned specification of Table A.8 and the comparable-town reservoirs of Table A.3, both of which preserve their significance.

Table A.7 makes the anchoring argument operational: it re-computes the relative-magnitudes bounds at the mature horizon ($e = +10$) for the four enrolment margins on the trimmed and matched control reservoirs of Section A.3, so that the pre-treatment deviations anchoring the bounds are measured among comparable towns only. Two findings emerge. First, the count margin's robustness is not an artifact of the untrimmed comparison: the all-faculty breakdown threshold is essentially unchanged when the never-treated reservoir shrinks from 1,169 municipalities to the 55 towns above 5,000 inhabitants ($\bar{M} = 1.7$) or the 51 propensity-score-matched towns ($\bar{M} = 1.6$), the event-time effect itself barely moves, and the smoothness breakdown moves to $M = 1.0$ under the 5,000 cut and $M = 0.5$ in the matched sample; the STEM count softens ($\bar{M} = 0.9$ to 0.6), as expected when the reservoir shrinks twentyfold and every confidence set widens. Second, the rate margins' breakdown values do not improve on the restricted reservoirs ($\bar{M} = 0.3$ to 0.3–0.2 for the all-faculty rate, 0.1 or below throughout for STEM): the rates' own anchoring placebo attenuates only modestly among comparable towns (from 1.87 to 1.59–1.69, remaining at least marginally significant in Table A.3), and the precision lost with the smaller reservoir offsets what attenuation there is. The trimmed re-run thus sharpens the division of labour in the evidence rather than shifting it: the extensive-margin count clears the relative-magnitudes test on every reservoir, while the intensive-margin rates are not supported by these bounds on any reservoir; their support comes instead from the point estimates themselves, which remain significant in every trimmed and matched column of Table A.3 and under the town-size-conditioned specification of Table A.8.

Table A.8 re-estimates the same six margins with the interaction-weighted estimator of Sun and Abraham (2021), the imputation estimator of Borusyak et al. (2024), and the main Callaway–Sant'Anna specification with log 1851 municipal population added to the conditioning set, so that parallel trends need hold only among towns of similar pre-treatment size, a direct check on the growing-towns confound. No estimator overturns the sign or the order of magnitude of any headline estimate. The all-faculty enrolment count of 16.0 is bracketed by the Sun–Abraham (11.2, significant at the ten-percent level) and imputation (22.2, one-percent level) estimates and barely moves when town size enters the conditioning set (14.7, one-percent level); the STEM count behaves identically. The population-scaled rates are noisier away from the main specification: the Sun–Abraham estimates are larger but marginal, the imputation estimates are close to the headline magnitudes but imprecise, and conditioning on town size attenuates the all-faculty rate from 3.7 to 2.6 while preserving significance at the five-percent level; this is the same one-to-three-tenths attenuation as in the trimmed-reservoir exercise of Section A.3, consistent with a modest town-scale component in the untrimmed comparison. The notable-person estimates remain positive and of similar magnitude throughout.

Table A.6: Pre-trend sensitivity: Rambachan–Roth honest bounds

Outcome	Event time	Original 95% CI	$\bar{M} = 0.5$	$\bar{M} = 1$	$\bar{M} = 1.5$	$\bar{M} = 2$	Breakdown $\bar{M}$	Smooth. breakdown M
Enrolment count (all)	$e = +0$	[0.82, 8.31]	[0.27, 9.38]	[-0.50, 10.76]	[-1.57, 12.29]	[-2.72, 13.82]	0.60	0.47
Enrolment count (all)	$e = +10$	[10.11, 21.57]	[8.37, 24.98]	[5.44, 29.19]	[1.93, 33.75]	[-1.35, 38.32]	1.70	0.93
Enrolment / 10k pop (all)	$e = +0$	[0.43, 1.97]	[-0.60, 2.94]	[-2.08, 4.40]	[-3.62, 5.94]	[-5.15, 7.48]	0.20	0.00
Enrolment / 10k pop (all)	$e = +10$	[1.81, 5.27]	[-1.36, 8.76]	[-5.85, 13.32]	[-10.49, 17.67]	[-15.12, 17.67]	0.30	< 0.0
Enrolment count (STEM)	$e = +0$	[-0.21, 4.89]	[-0.44, 5.71]	[-0.81, 6.75]	[-1.69, 7.95]	[-2.89, 9.20]	< 0.1	< 0.0
Enrolment count (STEM)	$e = +10$	[4.39, 10.51]	[3.03, 13.40]	[-0.53, 17.03]	[-4.47, 21.03]	[-8.47, 25.09]	0.90	0.00
Enrolment / 10k pop (STEM)	$e = +0$	[-0.13, 1.31]	[-0.90, 2.03]	[-2.08, 3.20]	[-3.29, 4.43]	[-4.54, 5.66]	< 0.1	< 0.0
Enrolment / 10k pop (STEM)	$e = +10$	[0.79, 2.23]	[-2.31, 5.39]	[-6.04, 7.36]	[-7.36, 7.36]	[-7.36, 7.36]	0.10	< 0.0
Notable count	$e = +0$	[-0.31, 0.52]	[-0.57, 0.82]	[-0.96, 1.22]	[-1.39, 1.67]	[-1.84, 2.12]	< 0.1	< 0.0
Notable count	$e = +10$	[-0.13, 0.55]	[-1.22, 1.77]	[-2.60, 3.16]	[-3.51, 3.51]	[-3.51, 3.51]	< 0.1	< 0.0
Notable / 10k (1851 base)	$e = +0$	[-0.15, 0.30]	[-0.28, 0.45]	[-0.47, 0.66]	[-0.69, 0.89]	[-0.93, 1.12]	< 0.1	< 0.0
Notable / 10k (1851 base)	$e = +10$	[-0.08, 0.28]	[-0.69, 0.89]	[-1.41, 1.60]	[-1.79, 1.79]	[-1.79, 1.79]	< 0.1	< 0.0

Rambachan and Roth (2023) sensitivity bounds applied to the Callaway and Sant'Anna (2021) dynamic event study of Tables 4.1 and 4.2. Each row is one outcome margin at one post-treatment event time. att_gt is fit exactly as in the main tables (not-yet-treated control, doubly robust group-time estimator conditioning on the three HDNG religion shares, 200-replication multiplier bootstrap) but with a universal base period so that every event-time coefficient is measured relative to $e = -5$, the last pre-treatment period; with 5-year bins this leaves three pre-period coefficients ($e = -20, -15, -10$) and five post ($e = 0, \ldots, +20$). "Original 95% CI" is the conventional confidence interval for the ATT at that event time (fixed-length, from the influence-function covariance). Columns $\bar{M} = 0.5 \ldots 2$ report the relative-magnitudes robust confidence set $\Delta^{RM}(\bar{M})$, which allows the post-treatment violation of parallel trends to be at most $\bar{M}$ times the largest observed pre-treatment violation. "Breakdown $\bar{M}$" is the largest $\bar{M}$ (on a 0.1 grid up to 2) at which the robust confidence set still excludes zero; a larger value means the effect is more robust to pre-trend violations. "Smooth. breakdown M" is the analogous breakdown value for the smoothness restriction $\Delta^{SD}(M)$, which bounds the per-period deviation of the differential trend from linearity. < 0.1 means the robust set already includes zero at the smallest grid value; ≥ 2.0 means it excludes zero throughout the grid. "—" marks a cell HonestDiD could not compute (reported, not silently skipped).

Table A.7: Pre-trend sensitivity on trimmed and matched control reservoirs

Outcome	Reservoir	N never-treated	Original 95% CI	$\bar{M} = 1$	Breakdown $\bar{M}$	Smooth. breakdown M
Enrolment count (all)	All controls	1,169	[10.11, 21.57]	[5.44, 29.19]	1.70	0.93
Enrolment count (all)	NT ≥ 5,000	55	[8.17, 19.63]	[4.74, 26.16]	1.70	1.02
Enrolment count (all)	PS matched	51	[7.95, 19.42]	[4.39, 25.96]	1.60	0.51
Enrolment / 10k pop (all)	All controls	1,169	[1.81, 5.27]	[-5.85, 13.32]	0.30	< 0.0
Enrolment / 10k pop (all)	NT ≥ 5,000	55	[1.33, 5.00]	[-5.24, 11.99]	0.30	< 0.0
Enrolment / 10k pop (all)	PS matched	51	[0.59, 4.53]	[-6.43, 11.78]	0.20	< 0.0
Enrolment count (STEM)	All controls	1,169	[4.39, 10.51]	[-0.53, 17.03]	0.90	0.00
Enrolment count (STEM)	NT ≥ 5,000	55	[3.50, 9.64]	[-2.60, 17.09]	0.60	0.00
Enrolment count (STEM)	PS matched	51	[3.28, 9.43]	[-2.92, 16.98]	0.60	0.00
Enrolment / 10k pop (STEM)	All controls	1,169	[0.79, 2.23]	[-6.04, 7.36]	0.10	< 0.0
Enrolment / 10k pop (STEM)	NT ≥ 5,000	55	[0.70, 2.30]	[-6.24, 8.19]	0.10	< 0.0
Enrolment / 10k pop (STEM)	PS matched	51	[0.09, 1.94]	[-7.04, 9.37]	< 0.1	< 0.0

Rambachan and Roth (2023) sensitivity bounds at the mature horizon ($e = +10$) for the four university-enrolment margins of Table A.6, re-computed on the trimmed and matched control reservoirs of Table A.3. All controls: untrimmed panel, replicating the corresponding rows of Table A.6. NT ≥ 5,000: never-treated municipalities kept only if their 1851 (pre-treatment) population is at least 5,000. PS matched: never-treated municipalities kept only if selected as the 1:1 nearest-neighbour match of a treated municipality on the fitted propensity score of ever-treated status from the logit of Table A.1. Ever-treated municipalities (as treated units and as not-yet-treated controls) are always retained; N never-treated is the number of never-treated municipalities remaining in the reservoir. Estimation is otherwise exactly as in Table A.6: not-yet-treated control, doubly robust group-time estimator conditioning on the three HDNG religion shares, 200-replication multiplier bootstrap, universal base period at $e = -5$. "Original 95% CI" is the conventional confidence interval for the ATT at $e = +10$; "$\bar{M} = 1$" is the relative-magnitudes robust confidence set allowing a post-treatment parallel-trends violation as large as the largest pre-treatment deviation; "Breakdown $\bar{M}$" is the largest multiple (on a 0.1 grid up to 2) at which the robust set still excludes zero, and "Smooth. breakdown M" the analogous value for the smoothness restriction. < 0.1 means the robust set already includes zero at the smallest grid value; ≥ 2.0 means it excludes zero throughout the grid.

Table A.8: Alternative estimators and a differential-trend specification

Outcome	CS (headline)	Sun–Abraham	BJS imputation	CS + log pop 1851
Enrolment count (all)	16.0457***	11.2086*	22.2425***	14.7266***
	(3.7101)	(6.1927)	(4.7076)	(3.5739)
Enrolment / 10k pop (all)	3.7212***	5.2501*	3.7095	2.5729**
	(0.7144)	(2.9213)	(4.3285)	(1.0274)
Enrolment count (STEM)	7.7408***	4.7144*	9.8155***	7.1257***
	(2.1332)	(2.6011)	(2.0693)	(1.9313)
Enrolment / 10k pop (STEM)	1.7784***	2.0882	1.6676	1.5224***
	(0.2490)	(1.2783)	(2.3595)	(0.4231)
Notable count	0.2937	1.1077	0.3166	0.2832
	(0.2060)	(0.8072)	(0.2306)	(0.1901)
Notable / 10k (1851 base)	0.1457*	1.0197***	0.1734	0.1411
	(0.0784)	(0.1849)	(0.5010)	(0.1197)

Post-treatment average treatment effect on the treated for each of the six headline margins of Tables 4.1 and 4.2, under four staggered difference-in-differences estimators. CS (headline): Callaway and Sant'Anna (2021), doubly robust, conditioning on the three HDNG religion shares, not-yet- treated control – the main-text specification; the reported figure is the equal-weighted average of dynamic event-time ATTs over $e = 0, \ldots, +20$. Sun–Abraham: the interaction-weighted event-study estimator of Sun and Abraham (2021) via `fixest::sunab`, reported as the `agg="att"` post-treatment average, with the same religion-share controls and municipality and cohort fixed effects. BJS imputation: the imputation estimator of Borusyak, Jaravel and Spiess (2024) via `didimputation::did_imputation`, overall post-treatment ATT, with the religion shares in the untreated-potential-outcome model. CS + log pop 1851: the CS specification with log 1851-baseline municipal population added to the conditioning set, so that parallel trends are conditioned on pre-treatment town size (the growing-towns confound of Section 3.4); this bounds the contribution of a size-driven differential trend to the estimate. Standard errors, clustered on municipality (amco), in parentheses; Sun–Abraham and BJS use analytic clustered SEs, CS a 200-replication multiplier bootstrap. Weights match the CS runs: count outcomes unweighted, population-scaled rates weighted by contemporaneous (or 1851-baseline for the last row) population. Never-treated municipalities enter as controls (coded 0 for CS and BJS, 100000 for sunab). * p<0.1, ** p<0.05, *** p<0.01.

A.6 Few treated clusters and randomisation inference

Only 55 municipalities are ever treated, and the staggered rollout spreads them thinly across treatment-cohort groups, so the multiplier-bootstrap standard errors reported in the main tables rest on small-sample asymptotics in the number of treated clusters. Table A.9 documents the problem directly: it reports how many treated municipalities identify each treatment-cohort group in each municipality-cohort design, and several groups are identified by only one or two treated towns. To check that the headline conclusions do not depend on the bootstrap's large-sample approximation, Table A.10 reports randomisation-inference p-values for the Simple post-treatment ATT of each headline outcome. Holding the panel fixed, the observed adoption events are reassigned at random to municipalities drawn from the full universe, the Simple ATT is re-estimated, and the exercise is repeated 999 times; the two-sided p-value is the share of permutations whose absolute ATT is at least as large as the observed one. The picture the test paints is sharper than the bootstrap alone. The enrolment effect is unambiguous: both the student count ($p = 0.001$) and the rate per $10{,}000$ population ($p = 0.005$) are more extreme than essentially every random reassignment, so this result is in no way an artefact of the clustered bootstrap. The notable-person effect is firmest on the extensive margin, where the raw count of upper-tail biographical entries clears the test decisively ($p = 0.001$); its population-scaled primary margin points the same way and remains close to conventional significance under this more demanding test ($p = 0.124$), the modest softening being what one expects of the noisier per-capita denominator rather than a sign that the effect is absent. Only the scientific-elite (KNAW) margin, identified off the fewest treated clusters, is clearly indistinguishable from a random placement of the adoption events ($p = 0.239$), consistent with the cautious reading of that outcome in Section 4.3. Randomisation inference thus corroborates the enrolment result and the notable-person count directly and leaves the population-scaled notable-person rate as supportive.

Table A.9: Treated municipalities per treatment-cohort group

design	g=1860	g=1865	g=1870	g=1875	g=1880	g=1885	g=1890	g=1895
Enrolment	9	31	7	3	2	1	1	1
Notable persons	9	31	7	3	2	1	1	1
Scientific elite	9	31	7	3	2	1	1	1

Number of ever-treated municipalities identifying each treatment-cohort group g (the five-year bin in which a municipality's first HBS opened) in each municipality-cohort design. Group-time ATTs for bins identified by only one or two treated municipalities carry little independent information; the randomisation inference in Table A.10 and the municipality-clustered bootstrap in the main tables are the relevant guides to precision for those cells.

Table A.10: Randomisation-inference p-values, headline outcomes

Outcome	Simple ATT	Permutation p	Treated munis
Enrolment / 10k pop (primary)	3.4605	0.005***	55
Enrolment count	20.4498	0.001***	55
Notable upper-tail / 10k pop (1851, primary)	0.2732	0.124	54
Notable upper-tail count	0.6139	0.001***	55
KNAW / 10k pop (1851, primary)	0.0833	0.239	54

Randomisation-inference p-values for the Simple post-treatment ATT of each headline municipality-cohort outcome. Holding the panel fixed, the 55 observed (municipality, opening-date) adoption events are reassigned at random to municipalities drawn from the full universe, the Simple ATT is re-estimated (outcome-regression-only, no covariates, not-yet-treated control), and this is repeated 999 times. The reported two-sided p-value is (1 + number of permutations with |ATT| at least the observed |ATT|) / (1 + number of valid permutations). Unlike the multiplier bootstrap, this test does not rely on large-sample asymptotics in the number of treated clusters. * p<0.1, ** p<0.05, *** p<0.01 by this permutation test.

A.7 Never-treated control group

Tables A.11, A.12, A.13, and A.14 replace the not-yet-treated comparison group with never-treated municipalities. The estimates are close to their main-text counterparts throughout, but this closeness is largely mechanical rather than independently reassuring: the two reservoirs are nearly identical sets, since the 1,169 never-treated municipalities dominate both and the not-yet-treated group merely adds the later adopters to them. The comparison therefore mainly confirms that the later adopters do not receive enough weight to move the point estimates, and carries little information about the overlap concern itself; the informative robustness exercise is the comparable-towns restriction of Section A.3, which changes the composition of the reservoir rather than shuffling units between two nearly coincident versions of it.

A.8 Dynamic event-study path

Tables A.15, A.16, A.17, and A.18 report the disaggregated post-treatment event-study ATTs (event times $e = 0, +5, \ldots, +20$) alongside the pre-treatment placebos. They show the enrolment and notable-person effects building over event time rather than jumping on impact, consistent with the educational-pipeline mechanism discussed in Section 4.5, and confirm that the pre-treatment coefficients are broadly flat, though not uniformly so. The exceptions are the following. On the enrolment margins (Table A.15), the all-faculty count placebo at $e = -10$ is positive and significant (1.73, roughly a tenth to a sixth of the mature post-treatment path), the all-faculty population-scaled rate shows a significant placebo at $e = -10$ (1.87) and a marginal one at $e = -5$, the all-faculty births-scaled rate a negative placebo at $e = -20$, and the STEM count and population-scaled rates marginal placebos at $e = -5$. In addition, there is a distant, small-cell placebo at $e = -20$ on the DWC/KNAW headcount margin discussed in Section 4.3. Section A.3 shows that the significant enrolment-count placebo attenuates to insignificance in every trimmed or matched control reservoir, and Section A.5 quantifies how much of the post-treatment effect such pre-treatment deviations could account for. Figure A.2 plots the corresponding DWC/KNAW event-study path.

Table A.11: University enrollment DiD estimates (never-treated control)

	Students (count)	Students / 1,000 births	Students / 10,000 pop
Panel A: All faculties			
e = -20	0.1443	-0.9153***	-0.3216
	(1.0567)	(0.3146)	(1.2819)
e = -15	1.4208	0.0063	0.8521
	(2.2172)	(1.8045)	(1.8357)
e = -10	1.7685***	1.1307	1.9532**
	(0.6383)	(1.3300)	(0.9437)
e = -5	1.7330	-0.5702	0.4795*
	(1.1499)	(0.6572)	(0.2450)
Post ATT[0,20] (Simple)	16.0651***	1.0107	3.7730***
	(3.5619)	(1.1107)	(0.6877)
Post ATT[0,20] (Weighted)	20.7078***	0.7381	3.4861***
	(4.5504)	(1.5181)	(1.1009)
N observations	13,863	12,749	13,863
N cohorts	8	7	8
Panel B: STEM faculties			
e = -20	0.4343	-0.5286	0.0831
	(0.9362)	(0.6933)	(1.0570)
e = -15	1.0643	-0.3267	0.7612
	(1.8776)	(1.7843)	(1.7608)
e = -10	0.2004	-0.1516	0.3176
	(0.9517)	(0.9279)	(0.8575)
e = -5	1.4222*	-0.0708	0.4773***
	(0.7869)	(0.5930)	(0.1523)
Post ATT[0,20] (Simple)	7.7529***	0.4843**	1.8144***
	(2.0005)	(0.2185)	(0.2322)
Post ATT[0,20] (Weighted)	8.8747***	0.2827	1.5494***
	(1.9855)	(0.3432)	(0.3497)
N observations	13,863	12,749	13,863
N cohorts	8	7	8

Callaway and Sant'Anna (2021) staggered difference-in-differences on a panel of municipalities by 5-year birth cohort (year the child turned twelve). Unit of observation: municipality-cohort cell. Outcome is the number / rate of children from a municipality enrolling at one of five universities (Leiden, Utrecht, Groningen, VU Amsterdam, Universiteit van Amsterdam; the latter two are graduate/non-matriculation registers). Panel B (STEM) restricts to the mathematics, natural-science, medicine, dentistry and pharmacy faculties. The count is unweighted; the rate per 1,000 births is weighted by births; the rate per 10,000 population is weighted by municipal population. Rows $e < 0$ are pre-treatment placebo ATTs. Post ATT[0,20] (Simple) is the equal-weighted average of dynamic event-time ATTs over $e = 0..20$; (Weighted) is the treatment-group-size-weighted average. Bootstrapped SE, clustered by municipality (999 reps), in parentheses. Group-time effects are estimated with the doubly robust (outcome regression + inverse-probability weighting) Callaway-Sant'Anna estimator. * $p<0.1$, ** $p<0.05$, *** $p<0.01$. Control group: never-treated municipalities.

Table A.12: Upper-tail human capital DiD estimates (never-treated control)

	Bioport (count)	Bioport upper-tail (count)	Bioport / 10k pop	Bioport upper-tail / 10k pop	Bioport / 10k pop (1851 baseline)	Bioport upper-tail / 10k pop (1851 baseline)
e = -20	-0.9410*	-0.1350	-0.9741***	-0.1342	-0.4316	-0.0471
	(0.5715)	(0.1572)	(0.3190)	(0.1114)	(0.3439)	(0.0920)
e = -15	-0.4568	-0.5909**	-0.2994	-0.3233**	-0.2369	-0.3090**
	(0.4173)	(0.2424)	(0.2396)	(0.1298)	(0.2217)	(0.1232)
e = -10	-0.3533	0.2830	-0.3214	0.1419	-0.1260	0.1783
	(0.6440)	(0.3444)	(0.3658)	(0.1900)	(0.3401)	(0.1801)
e = -5	-0.4854	-0.2426	-0.3948	-0.1313	-0.2472	-0.1124
	(0.6620)	(0.2183)	(0.2708)	(0.1202)	(0.3283)	(0.1246)
Post ATT[0,20] (Simple)	4.2670*	0.2941	1.0944**	0.0336	2.0569***	0.1484*
	(2.3441)	(0.1851)	(0.4645)	(0.0674)	(0.5517)	(0.0801)
Post ATT[0,20] (Weighted)	7.5010**	0.6038*	0.8119*	-0.0196	3.3093***	0.2755***
	(3.6274)	(0.3590)	(0.4452)	(0.0701)	(0.7212)	(0.0962)
N observations	17,178	17,178	17,178	17,178	16,900	16,900
N cohorts	8	8	8	8	7	7

Callaway and Sant'Anna (2021) staggered difference-in-differences on a panel of municipalities by 5-year birth cohort (year the child turned twelve). Unit of observation: municipality-cohort cell. Outcome is the number of persons born in a municipality who later received a biografischportaal.nl entry ("upper-tail" restricts to business, colonial trade, science and healthcare categories). The count columns are unweighted raw counts; the per-10,000-pop columns scale by contemporaneous municipal population; the 1851-baseline columns scale by each municipality's population in 1851 (the first year of near-universal HDNG coverage, and safely pre-treatment since the earliest HBS opening is 1860), held fixed across all of that municipality's cohorts, so it cannot itself respond to treatment. Scaling notable-person counts by a population measured before the outcome window follows Bell et al. (2019) and Squicciarini and Voigtlander (2015). Rows e < 0 are pre-treatment placebo ATTs. Post ATT[0,20] (Simple) is the equal-weighted average of dynamic event-time ATTs over e = 0..20; (Weighted) is the treatment-group-size-weighted average. Bootstrapped SE, clustered by municipality (999 reps), in parentheses. Group-time effects are estimated with the doubly robust (outcome regression + inverse-probability weighting) Callaway-Sant'Anna estimator. * p<0.1, ** p<0.05, *** p<0.01. Control group: never-treated municipalities.

Table A.13: Human capital in the broader population, DiD estimates (never-treated control)

	P(son HISCLASS <= 3 | father)	P(son HISCLASS < father)	Son HISCAM	P(son < father | father HISCLASS > 6)
e = -20	0.0098	0.0440***	0.4025	0.0431
	(0.0082)	(0.0129)	(0.3266)	(0.0332)
e = -15	-0.0037	0.0058	0.2561	0.0107
	(0.0069)	(0.0110)	(0.3228)	(0.0124)
e = -10	0.0053	0.0177	-0.5287	0.0041
	(0.0102)	(0.0123)	(0.3741)	(0.0127)
e = -5	0.0117	-0.0168	0.6808**	-0.0132
	(0.0088)	(0.0103)	(0.3455)	(0.0131)
Post ATT[0,20] (Simple)	-0.0035	-0.0009	0.1125	0.0121
	(0.0070)	(0.0120)	(0.3127)	(0.0139)
Post ATT[0,20] (Weighted)	-0.0059	-0.0079	0.0600	0.0062
	(0.0074)	(0.0119)	(0.3058)	(0.0151)
N observations	237,834	237,834	237,834	207,502
N cohorts	8	8	8	8

Callaway and Sant'Anna (2021) staggered difference-in-differences on genealogieonline father-son pairs, binned to 5-year son cohorts (year the son turned twelve). Unit of observation: father-son pair. All specifications condition on the father's HISCLASS (lower HISCLASS = higher status), so the ATT is the effect of HBS access given paternal origin. The municipality-level HDNG religious shares used elsewhere are not included here: at this conditioning granularity they are collinear with the earliest treatment cohorts and produce a singular outcome regression. Column 4 restricts to low-origin families (father HISCLASS > 6) and therefore has a smaller sample. The sample is deduplicated to one row per father-son pair. Rows e < 0 are pre-treatment placebo ATTs. Post ATT[0,20] (Simple) is the equal-weighted average of dynamic event-time ATTs over e = 0..20; (Weighted) is the treatment-group-size-weighted average. Standard errors from a municipality-clustered multiplier bootstrap (999 reps) in parentheses. Group-time effects are estimated with the outcome-regression-only Callaway-Sant'Anna estimator: the doubly robust estimator is singular under this specification's father-HISCLASS-only covariate. * p<0.1, ** p<0.05, *** p<0.01. Control group: never-treated municipalities.

Table A.14: Notable-scientist DiD estimates (DWC/KNAW, never-treated control)

	KNAW membership (headcount)	KNAW membership (per 10,000 pop)	KNAW membership (per 10,000 1851 pop)	KNAW publication-weighted output	KNAW publication-weighted output (per 10,000 pop)	KNAW publication-weighted output (per 10,000 1851 pop)
e = -20	-0.2108**	-0.1425**	-0.1165**	-0.2544	-0.1633	-0.1403
	(0.0973)	(0.0621)	(0.0491)	(0.3341)	(0.1962)	(0.1585)
e = -15	-0.0751	-0.0424	-0.0472	-0.0029	-0.0134	-0.0134
	(0.0772)	(0.0444)	(0.0493)	(0.0025)	(0.0209)	(0.0199)
e = -10	0.0416	0.0234	0.0251	-0.0072	-0.0350	-0.0387
	(0.1156)	(0.0585)	(0.0569)	(0.0065)	(0.0495)	(0.0521)
e = -5	0.0552	0.0144	0.0210	0.2338	0.0421	0.0509
	(0.0911)	(0.0475)	(0.0516)	(0.1885)	(0.0994)	(0.1143)
Post ATT[0,20] (Simple)	0.0618	0.0273	0.0468	0.5253**	0.2568**	0.3308**
	(0.0602)	(0.0268)	(0.0330)	(0.2427)	(0.1088)	(0.1294)
Post ATT[0,20] (Weighted)	0.1629**	0.0204	0.0890***	0.5057**	0.1929*	0.3203***
	(0.0783)	(0.0251)	(0.0317)	(0.2324)	(0.1110)	(0.1194)
N observations	17,178	17,178	16,900	17,178	17,178	16,900
N cohorts	8	8	7	8	8	7

"KNAW membership" counts persons born in a municipality who later appear in the KNAW's Digitaal Wetenschapshistorisch Centrum (dwc.knaw.nl) as a deceased KNAW member; "KNAW publication-weighted output" sums those members' own linked Academy-proceedings publication counts. Callaway and Sant'Anna (2021) staggered difference-in-differences on a panel of municipalities by 5-year birth cohort (year the child turned twelve). Unit of observation: municipality-cohort cell. Headcount columns are unweighted raw counts; per-10,000-pop columns scale by contemporaneous municipal population; per-10,000-1851-pop columns scale by each municipality's 1851 (pre-treatment) population, held fixed across all of that municipality's cohorts so it cannot itself be a treatment effect, following the bioport upper-tail table (Bell et al. 2019; Squicciarini and Voigtlander 2015). Rows $e < 0$ are pre-treatment placebo ATTs. Post ATT[0,20] (Simple) is the equal-weighted average of dynamic event-time ATTs over e = 0..20; (Weighted) is the treatment-group-size- weighted average. Bootstrapped SE, clustered by municipality (999 reps), in parentheses. Group-time effects are estimated with the doubly robust (outcome regression + inverse-probability weighting) Callaway-Sant'Anna estimator. * $p<0.1$, ** $p<0.05$, *** $p<0.01$. Control group: never-treated municipalities.

Table A.15: University enrollment DiD estimates, event-study path

	Students (count)	Students / 1,000 births	Students / 10,000 pop
Panel A: All faculties			
e = -20	0.0995	-0.9270	-0.4198
	(1.0037)	(0.8089)	(1.1554)
e = -15	1.3796	-0.0029	0.7531
	(2.2236)	(1.7457)	(1.8962)
e = -10	1.7310***	1.1432	1.8695**
	(0.6717)	(1.3513)	(0.8961)
e = -5	1.7215	-0.5810	0.4333*
	(1.1236)	(0.6164)	(0.2569)
e = +0	4.5660*	-0.3184	1.1976***
	(2.4908)	(0.4878)	(0.4606)
e = +5	12.2130***	0.9130	3.0999***
	(3.2509)	(1.0590)	(0.5712)
e = +10	15.8387***	0.6241	3.5397***
	(3.1388)	(1.3653)	(1.1369)
e = +15	23.1775***	1.9797	5.2271***
	(4.9870)	(1.4149)	(1.2412)
e = +20	24.4330***	1.8149	5.5419***
	(5.6798)	(1.4196)	(1.1471)
N observations	13,863	12,749	13,863
N cohorts	8	7	8
Panel B: STEM faculties			
e = -20	0.4008	-0.5333**	0.0055
	(0.7910)	(0.2435)	(1.0174)
e = -15	1.0331	-0.3361	0.6796
	(1.8892)	(2.7192)	(1.7486)
e = -10	0.1696	-0.1452	0.2391
	(0.9543)	(0.8407)	(0.8422)
e = -5	1.4149*	-0.0699	0.4464***
	(0.7372)	(0.5897)	(0.1688)

Continued on next page

Table A.15: University enrollment DiD estimates, event-study path (Continued)

	Students (count)	Students / 1,000 births	Students / 10,000 pop
e = +0	2.3389	-0.0914	0.5895
	(1.8243)	(0.2446)	(0.4467)
e = +5	5.6789***	0.3602	1.3523***
	(1.6982)	(0.2510)	(0.2586)
e = +10	7.4476***	0.2049	1.5117***
	(1.7639)	(0.2915)	(0.3837)
e = +15	12.1104***	1.1015***	2.7694***
	(2.7734)	(0.4125)	(0.5933)
e = +20	11.1280***	0.8509**	2.6691***
	(2.5456)	(0.3574)	(0.4171)
N observations	13,863	12,749	13,863
N cohorts	8	7	8

Callaway and Sant'Anna (2021) staggered difference-in-differences on a panel of municipalities by 5-year birth cohort (year the child turned twelve). Unit of observation: municipality-cohort cell. Outcome is the number / rate of children from a municipality enrolling at one of five universities (Leiden, Utrecht, Groningen, VU Amsterdam, Universiteit van Amsterdam; the latter two are graduate/non-matriculation registers). Panel B (STEM) restricts to the mathematics, natural-science, medicine, dentistry and pharmacy faculties. The count is unweighted; the rate per 1,000 births is weighted by births; the rate per 10,000 population is weighted by municipal population. Rows e < 0 are pre-treatment placebo ATTs. Post ATT[0,20] (Simple) is the equal-weighted average of dynamic event-time ATTs over e = 0..20; (Weighted) is the treatment-group-size-weighted average. Bootstrapped SE, clustered by municipality (999 reps), in parentheses. Group-time effects are estimated with the doubly robust (outcome regression + inverse-probability weighting) Callaway-Sant'Anna estimator. * p<0.1, ** p<0.05, *** p<0.01. Control group: not-yet-treated municipalities. Rows e >= 0 report the disaggregated post-treatment event-study.

Table A.16: Upper-tail human capital DiD estimates, event-study path

	Bioport (count)	Bioport upper-tail (count)	Bioport / 10k pop	Bioport upper-tail / 10k pop	Bioport / 10k pop (1851 baseline)	Bioport upper-tail / 10k pop (1851 baseline)
e = -20	-0.9231	-0.1276	-0.8101**	-0.0881	-0.3397	-0.0183
	(0.5874)	(0.1535)	(0.3571)	(0.1143)	(0.3384)	(0.0998)
e = -15	-0.4430	-0.5891**	-0.2405	-0.3129**	-0.2087	-0.3045**
	(0.4376)	(0.2352)	(0.2594)	(0.1261)	(0.2436)	(0.1288)
e = -10	-0.3387	0.2942	-0.2250	0.2028	-0.0447	0.2375
	(0.6911)	(0.3344)	(0.3580)	(0.1842)	(0.3414)	(0.1832)
e = -5	-0.4807	-0.2444	-0.3650	-0.1394	-0.2248	-0.1229
	(0.6684)	(0.2270)	(0.2723)	(0.1280)	(0.3343)	(0.1330)
e = +0	1.0226	0.1071	0.3602	0.0374	0.5458	0.0774
	(1.2640)	(0.1990)	(0.5183)	(0.0973)	(0.5293)	(0.1174)
e = +5	1.5774**	-0.1453	0.4284*	-0.1007	0.8083***	-0.0711
	(0.7225)	(0.1545)	(0.2358)	(0.0691)	(0.2552)	(0.0822)
e = +10	3.6610*	0.2091	0.9601**	0.0086	1.7453***	0.0998
	(1.9433)	(0.1926)	(0.4137)	(0.0817)	(0.4534)	(0.0901)
e = +15	6.1765*	0.5202**	1.4300***	0.0701	2.9078***	0.2478**
	(3.1988)	(0.2542)	(0.5155)	(0.0863)	(0.6223)	(0.1078)
e = +20	8.8708*	0.7772	2.1746***	0.1429	4.1163***	0.3744**
	(4.8035)	(0.5243)	(0.7789)	(0.1151)	(0.9668)	(0.1564)
N observations	17,178	17,178	17,178	17,178	16,900	16,900
N cohorts	8	8	8	8	7	7

Callaway and Sant'Anna (2021) staggered difference-in-differences on a panel of municipalities by 5-year birth cohort (year the child turned twelve). Unit of observation: municipality-cohort cell. Outcome is the number of persons born in a municipality who later received a biografischportaal.nl entry ("upper-tail" restricts to business, colonial trade, science and healthcare categories). The count columns are unweighted raw counts; the per-10,000-pop columns scale by contemporaneous municipal population; the 1851-baseline columns scale by each municipality's population in 1851 (the first year of near-universal HDNG coverage, and safely pre-treatment since the earliest HBS opening is 1860), held fixed across all of that municipality's cohorts, so it cannot itself respond to treatment. Scaling notable-person counts by a population measured before the outcome window follows Bell et al. (2019) and Squicciarini and Voigtlander (2015). Rows e < 0 are pre-treatment placebo ATTs. Post ATT[0,20] (Simple) is the equal-weighted average of dynamic event-time ATTs over e = 0..20; (Weighted) is the treatment-group-size-weighted average. Bootstrapped SE, clustered by municipality (999 reps), in parentheses. Group-time effects are estimated with the doubly robust (outcome regression

Table A.17: Human capital in the broader population, event-study path

	P(son HISCLASS <= 3 $\mid$ father)	P(son HISCLASS < father)	Son HISCAM	P(son < father $\mid$ father HISCLASS > 6)
e = -20	0.0098	0.0431***	0.4279	0.0417
	(0.0082)	(0.0126)	(0.3435)	(0.0321)
e = -15	-0.0050	0.0049	0.2141	0.0106
	(0.0073)	(0.0115)	(0.3197)	(0.0124)
e = -10	0.0050	0.0184	-0.5493	0.0044
	(0.0095)	(0.0119)	(0.3837)	(0.0120)
e = -5	0.0124	-0.0168	0.7155**	-0.0130
	(0.0096)	(0.0108)	(0.3526)	(0.0121)
e = +0	-0.0110	0.0054	-0.0432	0.0164
	(0.0172)	(0.0164)	(0.3441)	(0.0139)
e = +5	-0.0144*	0.0031	-0.4406	0.0158
	(0.0085)	(0.0102)	(0.3024)	(0.0143)
e = +10	0.0107	-0.0080	0.3189	0.0148
	(0.0097)	(0.0176)	(0.3889)	(0.0203)
e = +15	0.0003	0.0006	0.4656	0.0122
	(0.0121)	(0.0145)	(0.6539)	(0.0158)
e = +20	-0.0031	-0.0057	0.2744	0.0013
	(0.0103)	(0.0154)	(0.4422)	(0.0335)
N observations	237,834	237,834	237,834	207,502
N cohorts	8	8	8	8

Callaway and Sant'Anna (2021) staggered difference-in-differences on genealogieonline father-son pairs, binned to 5-year son cohorts (year the son turned twelve). Unit of observation: father-son pair. All specifications condition on the father's HISCLASS (lower HISCLASS = higher status), so the ATT is the effect of HBS access given paternal origin. The municipality-level HDNG religious shares used elsewhere are not included here: at this conditioning granularity they are collinear with the earliest treatment cohorts and produce a singular outcome regression. Column 4 restricts to low-origin families (father HISCLASS > 6) and therefore has a smaller sample. The sample is deduplicated to one row per father-son pair. Rows e < 0 are pre-treatment placebo ATTs. Post ATT[0,20] (Simple) is the equal-weighted average of dynamic event-time ATTs over e = 0..20; (Weighted) is the treatment-group-size-weighted average. Standard errors from a municipality-clustered multiplier bootstrap (999 reps) in parentheses. Group-time effects are estimated with the outcome-regression-only Callaway-Sant'Anna estimator: the doubly robust estimator is singular under this specification's father-HISCLASS-only covariate. * p<0.1, ** p<0.05, *** p<0.01. Control group: not-yet-treated municipalities. Rows e >= 0 report the disaggregated post-treatment event-study.

Table A.18: Notable-scientist DiD estimates (DWC/KNAW), event-study path

	KNAW membership (headcount)	KNAW membership (per 10,000 pop)	KNAW membership (per 10,000 1851 pop)	KNAW publication-weighted output	KNAW publication-weighted output (per 10,000 pop)	KNAW publication-weighted output (per 10,000 1851 pop)
e = -20	-0.2108**	-0.1382**	-0.1142**	-0.2551	-0.1626	-0.1396
	(0.0958)	(0.0594)	(0.0477)	(0.3315)	(0.1885)	(0.1621)
e = -15	-0.0752	-0.0422	-0.0477	-0.0021	-0.0039	-0.0044
	(0.0756)	(0.0475)	(0.0480)	(0.0031)	(0.0194)	(0.0190)
e = -10	0.0420	0.0283	0.0291	-0.0119*	-0.0563	-0.0601
	(0.1111)	(0.0586)	(0.0593)	(0.0071)	(0.0462)	(0.0456)
e = -5	0.0554	0.0154	0.0219	0.2340	0.0469	0.0563
	(0.0887)	(0.0543)	(0.0512)	(0.2356)	(0.1038)	(0.1148)
e = +0	0.2265*	0.1014**	0.1237**	0.3435	0.2035	0.2373
	(0.1190)	(0.0465)	(0.0571)	(0.3771)	(0.1873)	(0.2102)
e = +5	0.0152	0.0079	0.0188	0.9606**	0.4227**	0.5139***
	(0.0816)	(0.0352)	(0.0441)	(0.3857)	(0.1658)	(0.1948)
e = +10	-0.0589	-0.0264	-0.0224	0.1663	0.0778	0.0975
	(0.0783)	(0.0321)	(0.0386)	(0.2556)	(0.1340)	(0.1684)
e = +15	0.0913	0.0394	0.0731*	0.6247	0.3175	0.4331
	(0.0741)	(0.0303)	(0.0379)	(0.5432)	(0.1950)	(0.2768)
e = +20	0.0346	0.0100	0.0365	0.5304	0.2592	0.3693
	(0.0829)	(0.0396)	(0.0506)	(0.5184)	(0.1980)	(0.2652)
N observations	17,178	17,178	16,900	17,178	17,178	16,900
N cohorts	8	8	7	8	8	7

"KNAW membership" counts persons born in a municipality who later appear in the KNAW's Digitaal Wetenschapshistorisch Centrum (dwc.knaw.nl) as a deceased KNAW member; "KNAW publication-weighted output" sums those members' own linked Academy-proceedings publication counts. Callaway and Sant'Anna (2021) staggered difference-in-differences on a panel of municipalities by 5-year birth cohort (year the child turned twelve). Unit of observation: municipality-cohort cell. Headcount columns are unweighted raw counts; per-10,000-pop columns scale by contemporaneous municipal population; per-10,000-1851-pop columns scale by each municipality's 1851 (pre-treatment) population, held fixed across all of that municipality's cohorts so it cannot itself be a treatment effect, following the bioport upper-tail table (Bell et al. 2019; Squicciarini and Voigtlander 2015). Rows e < 0 are pre-treatment placebo ATTs. Post ATT[0,20] (Simple) is the equal-weighted average of dynamic event-time ATTs over e = 0..20; (Weighted) is the treatment-group-size- weighted average. Bootstrapped SE,

Figure A.2: Event-study estimates: scientific-elite production, DWC/KNAW (not-yet-treated control).

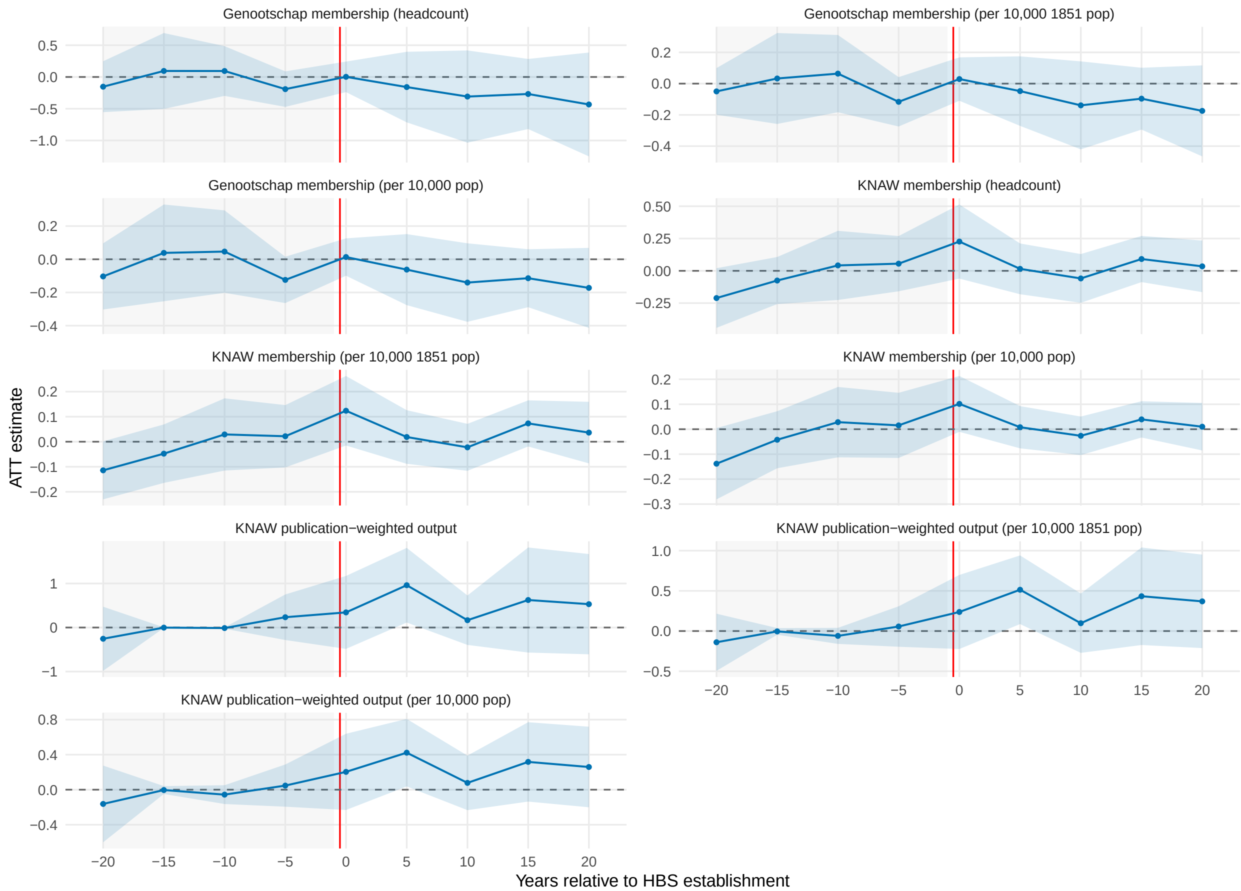


Table A.19: Learned-society (Genootschap) membership DiD estimates (DWC)

	Genootschap membership (headcount)	Genootschap membership (per 10,000 pop)	Genootschap membership (per 10,000 1851 pop)
e = -20	-0.1532	-0.1033	-0.0497
	(0.1839)	(0.0898)	(0.0639)
e = -15	0.0946	0.0384	0.0329
	(0.2740)	(0.1314)	(0.1246)
e = -10	0.0940	0.0463	0.0640
	(0.1797)	(0.1119)	(0.1059)
e = -5	-0.1914	-0.1243**	-0.1163*
	(0.1283)	(0.0629)	(0.0677)
Post ATT[0,20] (Simple)	-0.2331	-0.0951	-0.0856
	(0.2433)	(0.0734)	(0.0828)
Post ATT[0,20] (Weighted)	-0.4167	-0.1600*	-0.1660
	(0.3390)	(0.0893)	(0.1045)
N observations	17,178	17,178	16,900
N cohorts	8	8	7

"Genootschap membership" counts persons born in a municipality who later appear in the KNAW's Digitaal Wetenschapshistorisch Centrum (dwc.knaw.nl) as a member of a Dutch learned society (a broader, lower-bar elite than KNAW membership), reported for comparison with the main-text KNAW table – publication linkage is essentially absent for this group, so no publication-weighted analogue is reported. Callaway and Sant'Anna (2021) staggered difference-in-differences on a panel of municipalities by 5-year birth cohort (year the child turned twelve). Unit of observation: municipality-cohort cell. Headcount columns are unweighted raw counts; per-10,000-pop columns scale by contemporaneous municipal population; per-10,000-1851-pop columns scale by each municipality's 1851 (pre-treatment) population, held fixed across all of that municipality's cohorts so it cannot itself be a treatment effect, following the bioport upper-tail table (Bell et al. 2019; Squicciarini and Voigtlander 2015). Rows e < 0 are pre-treatment placebo ATTs. Post ATT[0,20] (Simple) is the equal-weighted average of dynamic event-time ATTs over e = 0..20; (Weighted) is the treatment-group-size- weighted average. Bootstrapped SE, clustered by municipality (999 reps), in parentheses. Group-time effects are estimated with the doubly robust (outcome regression + inverse-probability weighting) Callaway-Sant'Anna estimator. * p<0.1, ** p<0.05, *** p<0.01. Control group: not-yet-treated municipalities.

Table A.20: Learned-society (Genootschap) membership DiD estimates (DWC, never-treated control)

	Genootschap membership (headcount)	Genootschap membership (per 10,000 pop)	Genootschap membership (per 10,000 1851 pop)
e = -20	-0.1545	-0.1190	-0.0615
	(0.1916)	(0.0864)	(0.0728)
e = -15	0.0948	0.0435	0.0401
	(0.2605)	(0.1190)	(0.1217)
e = -10	0.0927	0.0380	0.0569
	(0.1564)	(0.1124)	(0.0953)
e = -5	-0.1916	-0.1283**	-0.1191*
	(0.1251)	(0.0637)	(0.0627)
Post ATT[0,20] (Simple)	-0.2334	-0.0959	-0.0866
	(0.2464)	(0.0722)	(0.0794)
Post ATT[0,20] (Weighted)	-0.4169	-0.1603*	-0.1664*
	(0.3456)	(0.0943)	(0.1007)
N observations	17,178	17,178	16,900
N cohorts	8	8	7

"Genootschap membership" counts persons born in a municipality who later appear in the KNAW's Digitaal Wetenschapshistorisch Centrum (dwc.knaw.nl) as a member of a Dutch learned society (a broader, lower-bar elite than KNAW membership), reported for comparison with the main-text KNAW table – publication linkage is essentially absent for this group, so no publication-weighted analogue is reported. Callaway and Sant'Anna (2021) staggered difference-in-differences on a panel of municipalities by 5-year birth cohort (year the child turned twelve). Unit of observation: municipality-cohort cell. Headcount columns are unweighted raw counts; per-10,000-pop columns scale by contemporaneous municipal population; per-10,000-1851-pop columns scale by each municipality's 1851 (pre-treatment) population, held fixed across all of that municipality's cohorts so it cannot itself be a treatment effect, following the bioport upper-tail table (Bell et al. 2019; Squicciarini and Voigtlander 2015). Rows e < 0 are pre-treatment placebo ATTs. Post ATT[0,20] (Simple) is the equal-weighted average of dynamic event-time ATTs over e = 0..20; (Weighted) is the treatment-group-size- weighted average. Bootstrapped SE, clustered by municipality (999 reps), in parentheses. Group-time effects are estimated with the doubly robust (outcome regression + inverse-probability weighting) Callaway-Sant'Anna estimator. * p<0.1, ** p<0.05, *** p<0.01. Control group: never-treated municipalities.