INNOVATION &
RESEARCH
CAUCUS

UK DOCTORAL GRADUATES' CONTRIBUTION TO INNOVATION (UK DGCI)

IRC Report No: 041

REPORT PREPARED BY

An Yu Chen
Manchester Institute of Innovation Research

Professor Cornelia Lawson
Manchester Institute of Innovation Research

Dr Xin Deng
Manchester Institute of Innovation Research

Dr Catalina Martinez
Spanish National Research Council (CSIC),
Institute of Public Goods and Policies (IPP)

Dr Alberto Corsini
Spanish National Research Council (CSIC),
Institute of Public Goods and Policies (IPP)

Dr Liangping Ding
Manchester Institute of Innovation Research

Delivered with
ESRC and
Innovate UK

Authors

The core members of the research team for this project were as follows:

- An Yu Chen – Manchester Institute of Innovation Research
- Professor Cornelia Lawson (PI) - Manchester Institute of Innovation Research
- Dr Xin Deng - Manchester Institute of Innovation Research
- Dr Catalina Martinez - Spanish National Research Council (CSIC), Institute of Public Goods and Policies (IPP)
- Dr Alberto Corsini - Spanish National Research Council (CSIC), Institute of Public Goods and Policies (IPP)
- Dr Liangping Ding - Manchester Institute of Innovation Research

This document relates to IRC Project FFOpen001: UK Doctoral Graduates' contributions to innovation

Acknowledgements

This work was supported by Economic and Social Research Council (ESRC) grant ES/X010759/1 to the Innovation and Research Caucus (IRC).

We are very grateful to the project sponsors at UK Research & Innovation (UKRI) for their input into this research. The interpretations and opinions within this report are those of the authors and may not reflect the policy positions of UKRI or its constituent councils. We would also like to acknowledge and appreciate the efforts of the IRC Project Administration Team for their support in preparing this report.

The project also received additional support from Alliance Manchester Business School via a small research grant and the European Patent Office via the DOC-TRACK UK project extension.

We are very grateful to members of the DOC-TRACK consortium for their input into this research, foremost Dr Johannes Koenig, and to Dr Tiantian Shen, Sanya Panda, Bahaar Dahiya, Dr Seunghyun Lee and Emily Johnson for excellent research assistance.

About the Innovation and Research Caucus

The Innovation and Research Caucus supports the use of robust evidence and insights in UKRI's strategies and investments, as well as undertaking a co-produced programme of research. Our members are leading academics from across the social sciences, other disciplines and sectors, who are engaged in different aspects of innovation and research systems. We connect academic experts, UKRI, IUK and the (ESRC), by providing research insights to inform policy and practice. Professor Tim Vorley and Professor Stephen Roper are Co-Directors. The IRC is funded by UKRI via the ESRC and IUK, grant number ES/X010759/1. The support of the funders is acknowledged. The views expressed in this piece are those of the authors and do not necessarily represent those of the funders.

Contact

You are also welcome to email us if you have any questions about this report or the work of the IRC generally: info@ircaucus.ac.uk

Cite as: Chen. AY, Lawson. C, Deng. X, Martinez. C, Corsini. A, and Ding. L. February 2026. *UK Doctoral Graduates' Contribution to Innovation (UK DGCI)*. Oxford, UK: Innovation and Research Caucus

Contents

Executive summary

This report draws upon a comprehensive dataset of UK doctoral graduates to uncover their contribution to science and innovation. The data relate to 347,838 graduates listed in the British Library's EThOS database of UK doctoral theses, who completed their studies between 2000 and 2020, covering all disciplines in all UK Higher Education Institutions. The graduates were matched to publications listed on Scopus and OpenAlex and to patents that cite these publications via Reliance on Science, following the DOC-TRACK (https://doc-track.eu/) methodology. The resulting data provides a **comprehensive and unprecedented opportunity to examine the contributions of UK doctoral graduates over time.**

Graduate population. The total number of UK dissertations grew significantly between 2000 and 2020, from approx. 12,000 to more than 20,000 graduates per year. Dissertations in STEMM (Science, Technology, Engineering, Mathematics and Medicine) fields consistently and substantially outnumber non-STEMM fields across the whole period. Still, social science graduates account for the largest share, with 24% of all UK dissertations. This is followed by medicine (excluding doctors of medicine) and engineering with 20% and 18% respectively.

Publishing trends. Publishing has become an increasingly essential part of the doctoral experience, with the share of PhD graduates who publish during their studies rising by 70% between 2000 and 2020. Publishing is more common in STEMM fields, particularly in physical sciences and medicine, with more than 65% of graduates having at least one publication during the PhD. This share is lower post-PhD, especially in engineering where only 42.6% of graduates continue to publish. In non-STEMM fields publishing propensity during the PhD has also increased and is now above 30%. An equivalent share of non-STEMM graduates continue to publish after the PhD indicating continued engagement in research. The share of graduates' publishing post-PhD has remained relatively stable over time for both STEMM and non-STEMM graduates.

The number of publications per PhD graduate has increased, with more recent cohorts of PhDs significantly more productive than earlier cohorts across all fields. For instance, in STEMM the number of publications more than doubled from 1.6 in 2000 to 3.5 in recent cohorts. This likely reflects a greater push towards journal publication for doctoral students in all domains. It is also consistent with expansion of publication opportunities, such as the founding of new journals and increased journal indexing in databases such as Scopus during our observed time period.

Inputs to innovation. UK PhD graduates have a substantial impact on patenting. Up to 40% of STEMM graduates who publish are being subsequently cited by EPO or USPTO patents. This represents a significant contribution to innovation by PhDs in the UK and highlights the innovation potential of doctoral research. The highest shares are observed in medicine and life sciences, followed by engineering. This finding does however contrast sharply with UK PhD graduates' low participation in patenting compared to other countries, which we reported elsewhere.

Gender patterns. Assigning gender as male or female based on first names[1], we document a persistent gender disparity in the number of PhD graduates in STEMM fields, where male graduates account for approximately 60% of dissertation authors, contrasting sharply with the near gender parity in non-STEMM fields. In terms of publications, there is only a small difference between male and female graduates in their propensity to publish during or after the PhD, with both almost equally as likely to share their research in publications. However, we observe a significant and growing gap in the number of papers published by male and female graduates, both during and after the PhD. We also observe a small but noticeable difference in male and female STEMM graduates' likelihood to inform future invention, with women's research less likely to be directly cited by a patent.

Regional patterns. There are substantial regional differences in STEMM vs. non-STEMM education and in the share of women in STEMM PhDs, with Northern Ireland and Scotland for instance showing a higher participation of women in STEMM compared to England and Wales. There is also a substantial difference in PhD graduates' publication propensity, which is highest in London, East and South-East England and in Scotland, and lowest in Northern Ireland and Wales. The highest shares of STEMM graduates' being cited in patents can be found in areas of key economic activity, with lowest shares in more rural areas of the UK. The majority of USPTO and EPO patents citing UK PhD graduates' research are however not filed by UK but by foremost US and European companies, with UK-based patent applicants accounting for just 6.6% of citing patents. This highlights the role of UK science as a foundational source for technological developments across multiple major innovation systems.

[1] This methodology has many limitations, and does not take account of the gender identity of the individual.

Implications. Our findings show that publishing has become an increasingly central feature of the doctoral experience, which suggests a need for policies that support equitable access to publishing opportunities across disciplines. A substantial share of STEMM graduates' work being later cited by patents highlights the value of strengthening pathways from academic research to innovation, including measures that promote gender-equitable access to commercialisation opportunities. Regional disparities in publishing propensity and innovation potential indicate that targeted investment in research infrastructure and doctoral training may be needed to support under-represented areas.

Contribution: This project responded to a need for more extensive, longitudinal data on PhD graduates in the UK. We mobilised already available data sources to demonstrate the contributions of UK doctoral graduates across time and regions. The data and methodology can become an important tool for policy, enabling not only the monitoring of doctoral performance but also pathways post PhD. The analyses presented here already draw attention to the significant share of graduates that contribute to science and innovation after the conclusion of their PhD, thus alleviating concerns regarding the touted exodus of PhDs from research. The database compiled for this project will in future enable analyses to understand where these researchers end up and work is underway to link this data to career databases to uncover non-academic pathways. The data can moreover shed light on different aspects of inequality. Here we presented some initial findings on gender and regional disparities. In an ***Annex B*** we demonstrated further uses for the data, with analyses of disparities in AI research and doctoral funding. All these analyses and findings shed light on the innovation potential of UK PhD graduates and help inform policies aimed at increasing it.

Further readings: The data collected as part of this study formed the basis for follow-on research in conjuncture with DOC-TRACK, funded by the European Patent Office. For a comparison of UK STEMM graduates to graduates of six European countries (Austria, Germany, France, Italy, Netherlands & Spain), for an analysis of STEMM graduates' patenting and distance to the technological frontier, and future data releases see https://doc-track.eu/.

1. Introduction

PhD holders are argued to make substantial economic, social and cultural contributions, and to present an important channel of knowledge transfer from science to industry, often beginning to do so during their doctoral studies (Buenstorf and Heinisch 2020; Corsini et al. 2022). With most programmes aimed towards training the next generation of academics (OJEU 2023) and PhDs continuing to express a preference for an academic career (Lawson and Lopes-Bento 2024), there is an increasing emphasis on research outputs with all the pathways towards knowledge transfer this entails. Moreover, with an increasing share of doctorate holders leaving academia for other sectors (Kwon 2025), there is added expectation that doctoral careers can build bridges between sectors (Marti & Peneoasu 2025). This is especially the case for the UK, where more than 70% of PhDs take up non-academic employment (Hancock 2023). Yet, whether these developments translate into innovation remains an open question as the evidence base regarding UK PhDs is currently underdeveloped and there is specific need for richer data on PhD holders, and more extensive, longitudinal studies of their careers and scientific and inventive activity (Hancock 2021; Grove 2025).

To address this, this project created a longitudinal database of UK PhDs utilising open databases that have become available in recent years. We adopt the methodology developed by DOC-TRACK (https://doc-track.eu/), a European-wide effort to investigate the performance of PhDs and create a new open access PhD database funded by the European Patent Office Academic Research Programme (EPO-ARP). At its core, this methodology relies on an Electronic Theses & Dissertations (ETD) repository, matched to publication and patent data (Corsini et al., 2025). For the UK, we build on metadata from EThOS (Electronic Theses Online Service), the British Library's national database for UK doctoral theses. EThOS is a free online service, last updated in November 2023, and covers roughly 98% of all PhDs awarded by UK higher education institutions since 1787. The metadata database is available under creative commons license and the British Library actively encourages the use of the data. These dissertation authors are matched to scientific publications and patents, both during and after their doctoral studies. The final database enables us to measure the contributions of UK doctoral students to science and innovation. To the best of our knowledge this represents the most comprehensive record of UK PhD graduates' activity.

2. Main Results of the Project

This chapter gives a summary of the main results of the project, focussing on doctoral dissertations published between 2000 and 2020 in the UK, a total of 347,838 graduates.[2] Detailed information on the methodology can be found in *Annex A: UK DGCI Methodology*. Complementary analyses using the project data are reported in *Annex B: Complementary analyses by UK-DGCI.*

2.1 UK dissertation authors

The annual number of UK PhD dissertations recorded in EThOS increased from just over 12,000 dissertations in 2000 to more than 20,000 dissertations for years since 2016 (see Figure 1). This upward trajectory indicates a significant expansion of doctoral education and research activity within the UK over nearly two decades, driven by an expansion of the higher education sector and the growing importance of doctoral students as indicators of research capacity within national research assessment exercises (Bogle, 2015; Halse and Mowbray 2011; UKRI 2024).

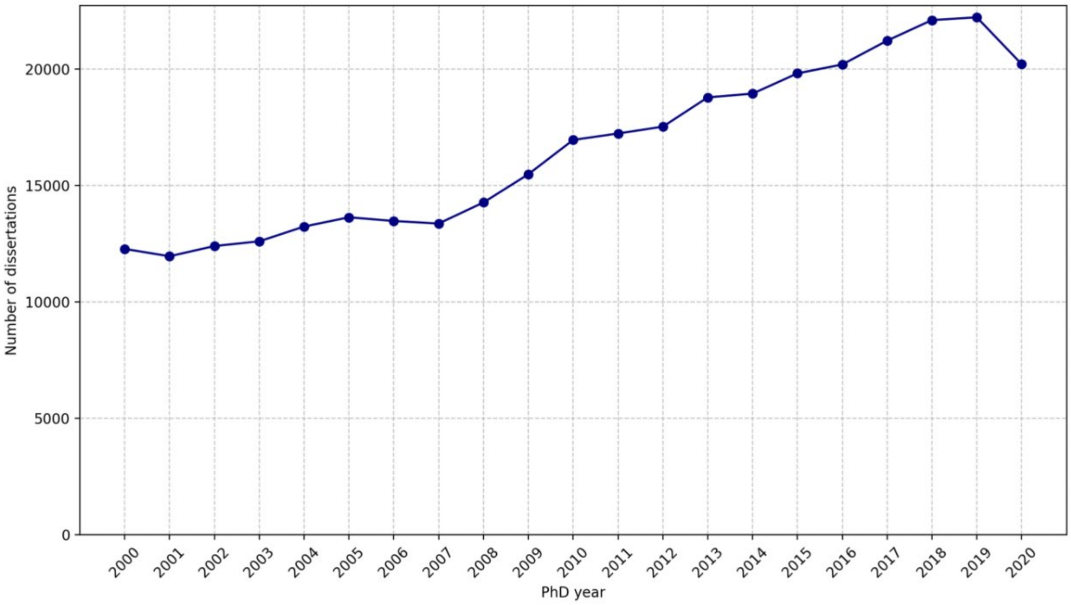

Figure 1: Number of UK doctoral theses over time.
Source: Own analysis based on EThOS metadata

[2] EThOS contains details on academic and professional doctorates and other postgraduate research degrees obtained in the UK. We retain only PhD, DPhil and EngD dissertations and discount any Master research dissertations or professional doctorates such as MD or DBA for the purpose of this analysis.

EThOS records the discipline of dissertations, differentiating between 19 fields. We grouped these into STEMM (Science, Technology, Engineering, Mathematics and Medicine) and non-STEMM domains. Figure 2 reports the yearly number of PhD dissertations differentiating between STEMM and non-STEMM disciplines.

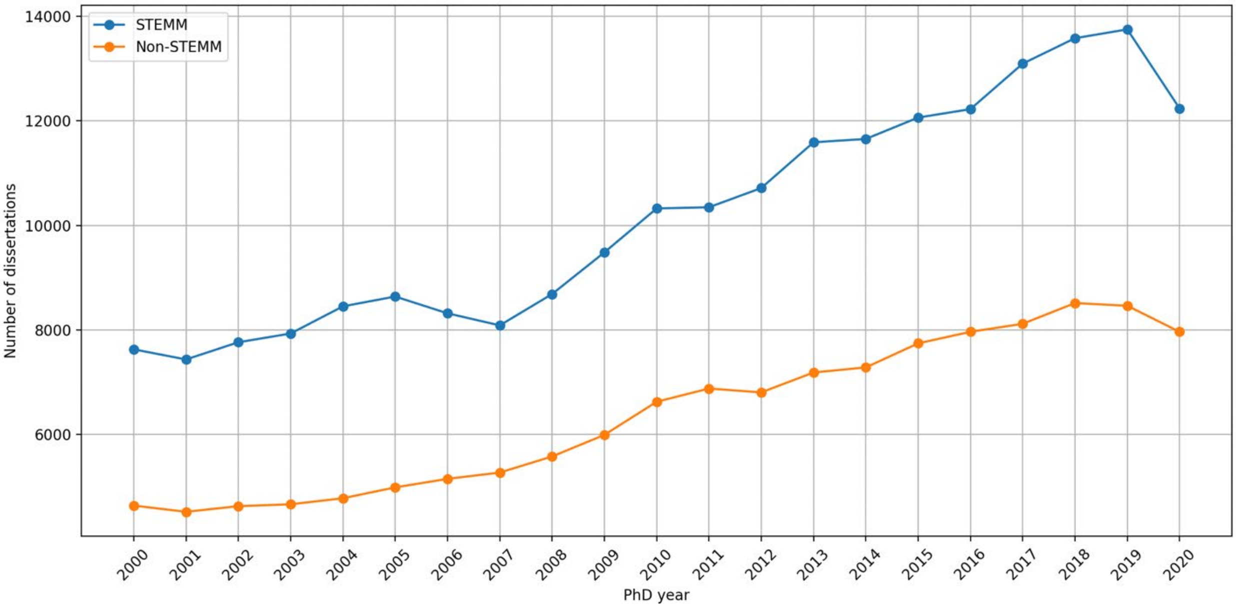

Figure 2: Number of UK theses over time, comparing STEMM and non-STEMM disciplines.
Source: Own analysis based on EThOS metadata

Throughout the entire period, the number of STEMM dissertations was consistently and substantially higher than that of non-STEMM dissertations. The growth in dissertations is seen in both domains. The number of STEMM graduates rises from approximately 7,600 in 2000 to a peak of nearly 14,000 in 2019. Non-STEMM dissertations grew from around 4,800 and peaking near 8,700 in 2018. While we see increases by about 80% in both domains, the gap between STEMM and non-STEMM has widened significantly, particularly after 2008, highlighting the increasing dominance of STEMM training within UK higher education.

While we observe an overall growth in PhD completions over the time period, this growth is interrupted around 2006–2007, particularly in STEMM field (see Figure 2). This temporary stagnation may be due to structural changes to doctoral training in the early 2000s rather than declining demand. Shifts in policy priorities in the early 2000s, altered both the structure and duration of doctoral study, introducing the 1+3 integrated PhD model, applied doctorates (e.g. EngD), and Doctoral Training Centres (DTCs) (Park, 2005; Bogle, 2015). These changes lengthened time-to-completion and temporarily reduced annual thesis submissions, and may thus explain the short-lived stagnation before numbers rebounded once cohorts progressed through the new system.

We further see a dip in graduate numbers in 2020, which is especially pronounced in STEMM fields (see Figure 2) and may reflect graduation delays during the COVID-19 pandemic; e.g. HESA gradate data show a return to 2019 numbers in 2023.[3] Still, the decline, preceded by stagnation in 2019, may also signal an end to the growth in PhD graduates. The announced reduction in PhD support by UKRI and Welcome Trust, and in overall university budgets, may indicate that we should expect a drop in the take-up of doctoral training and decrease in graduates moving forward (Grove, 2024).

To shed further light on the structure of the graduate population, Figure 3 breaks dissertations down by seven broad subject categories[4], 4 in STEMM and 3 in non-STEMM, which are used in the same order throughout the report. While STEMM accounts for a larger share of graduates than non-STEMM, the single largest subject domain is Social Sciences, accounting for 23.8% of dissertations, followed by Medicine with 19.7% and Engineering at 18.1%. Physical and Life Sciences have the smallest shares within the STEMM category, with proportions of 13.6% and 10.2%, respectively. Arts and Humanities account for about 11.0% of dissertation.

[3] See: https://www.hesa.ac.uk/data-and-analysis/students/outcomes (accessed: 1 December 2025)

[4] The seven broad subject categories are derived from the 19 EThOS fields and are abbreviated as ENGI, MEDI, PHYS, LIFE, SOCIAL, ARTS, and PSYCHO in relevant figures throughout the report. *Arts* refers to Arts & Humanities, which includes Language & Literature, History & Archaeology, Creative Arts & Design, Music. The label *Psycho* covers Philosophy, Psychology, and Religious Studies a grouping kept unchanged from the EThOS subject classification due to its ambiguity. *Social* includes Social, Economic & Political Studies, Education, Law, Business & Administrative Studies and Sport & Recreation. *Medi* covers Medicine and Health-related fields, while *Life* includes Biological Sciences and Agricultural & Veterinary Sciences. *Phys* corresponds to Physical Sciences, comprising subjects such as Physics, Earth Sciences, Mathematics & Statistics. *Eng* brings together Engineering & Technology, Architecture, Building & Planning, Computer Science and Librarianship & Information Science, covering disciplines from applied sciences and engineering fields. EThOS does not list Chemistry as a field. Manually reviewing theses on topics related to chemistry we found that they were mostly grouped within the *Eng*. These groupings derived from EThOS are intended to provide clarity and consistency in presenting subject fields.

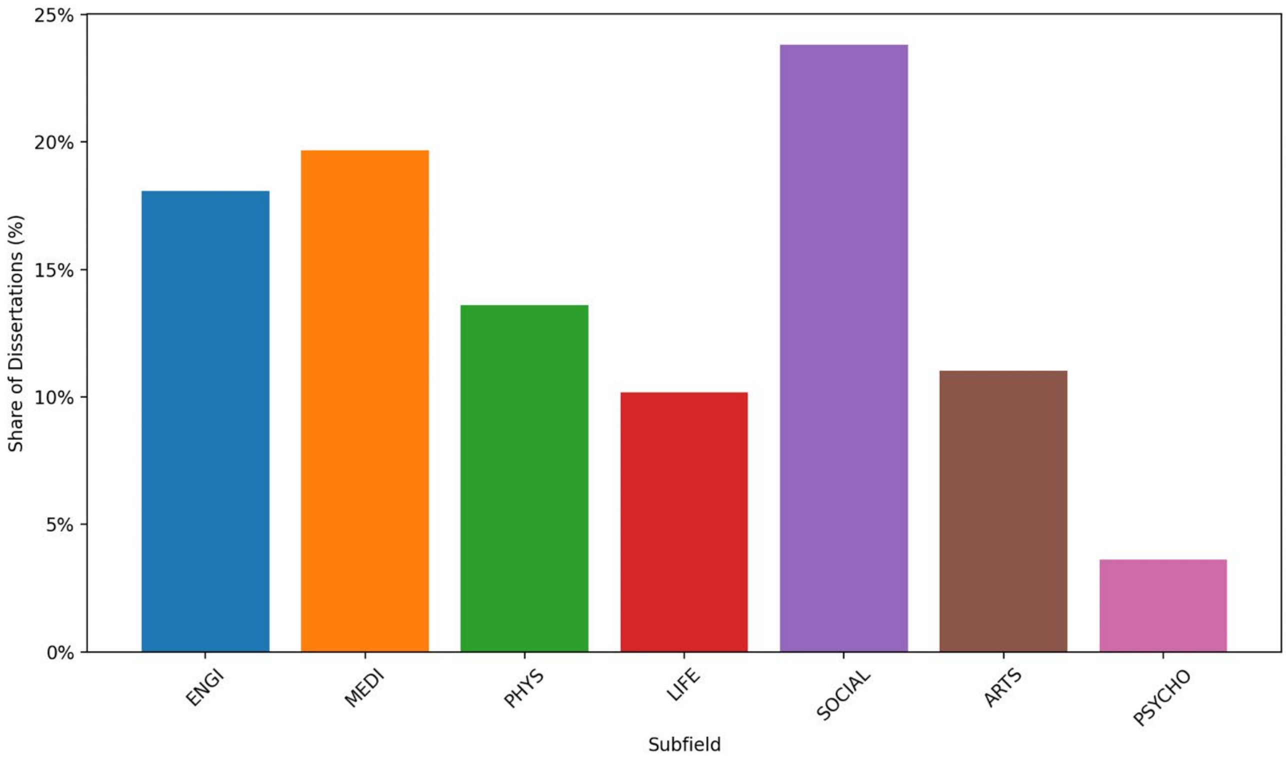

Figure 3: Share of UK doctoral dissertations (2000–2020) across seven subfields.

Note: STEMM subfields include ENGI, MEDI, PHYS, LIFE; non-STEMM subfields include SOCIAL, ARTS, and PSYCHO. Psychology in EThOS is defined as Philosophy, Psychology & Religious Studies. Source: Own analysis based on EThOS metadata.

Figure 4 summarises the gender distribution across STEMM and non-STEMM domains. Gender information is not included in EThOS but is inferred using name dictionaries by WIPO (Lax-Martinez et al., 2016, 2021). We were able to assign gender as male or female with high confidence to 86.1% of dissertation authors (299,649 out of 347,838 theses for the period 2000–2020).[5] All gender results in this report exclude observations with missing or ambiguous gender predictions. The data indicates a significant gender disparity in STEMM disciplines, with male graduates accounting for 59.7% of all PhD graduates, compared to 40.3% female graduates. While the share of female graduates increased between 2000 and 2005, it has since stagnated around that mark, but with renewed increase since 2017 to 42%.

[5] 14.8% of authors on STEMM theses and 12.4% of authors on non-STEMM theses could not be assigned a gender. Gender assignment was not possible for three reasons: (1) Missing first names (8% of authors); (2) Ambiguous or gender neutral first names (e.g. Taylor); (3) Low gender assignment and prediction confidence rates for graduates with non-European names, especially Chinese names where more than 50% could not be assigned a gender. This last point may introduce biases in the reported gender shares.

Non-STEMM fields, instead, exhibit gender parity with male and female graduates accounting for 49.9% and 50.1% of PhD dissertations, respectively, during the full period. The share of women has increased steadily and in 2020 was at about 53%.[6] It is important to note that the graphs should be interpreted as showing the overall trend for STEMM and non-STEMM fields. Minor fluctuations across years may be due to differences in data records, coverage, and methods used for gender prediction, rather than departures from longer term trends.

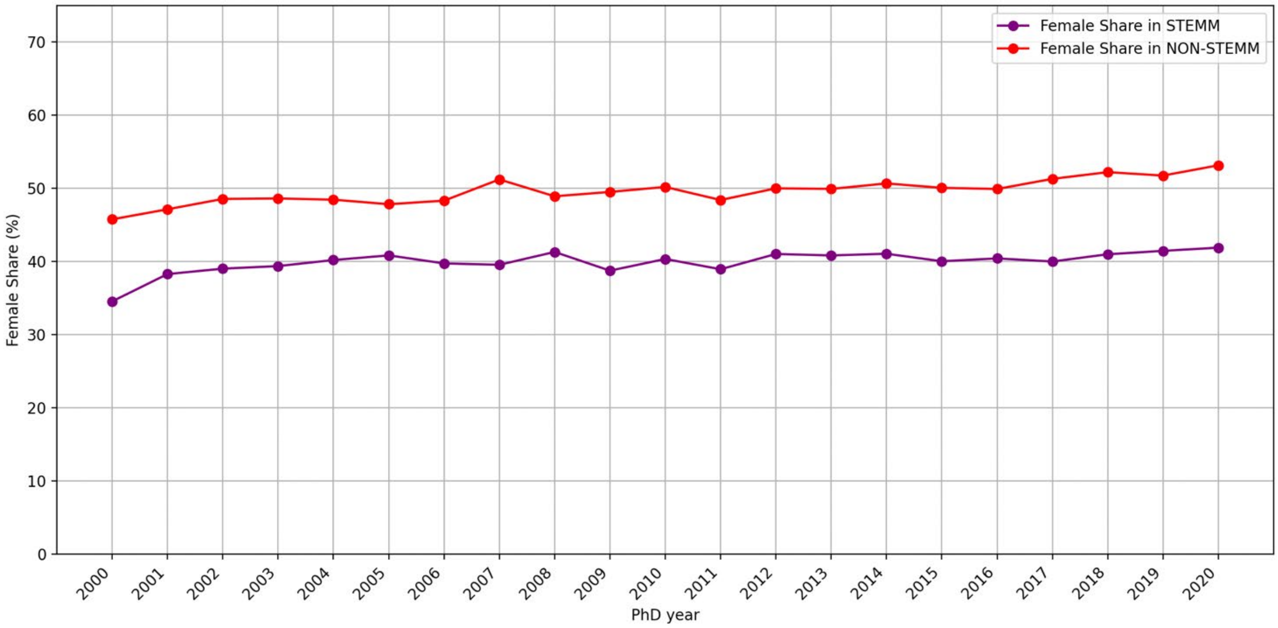

Figure 4: Disciplines Comparison of UK doctoral theses, share of female PhD by gender within STEMM and non-STEMM disciplines (2000–2020)
Source: Own analysis based on EThOS metadata; gender prediction using WIPO WGND

2.2 Publication propensity during and after PhD

Publications in academic journals are the foremost means of knowledge dissemination for doctoral students and a requirement in some fields. We follow the methodology developed by Corsini et al. (2025) and match dissertation authors to Scopus author profiles. Scopus records come with unique author IDs that are more accurate than other bibliographic databases and contain fewer false positives. These publications are then matched to OpenAlex records for processing and in the remainder, we report results using OpenAlex. We were able to identify

[6] HESA enrolment data for 2014-2020 confirms a modest increase in the share of women in STEMM disciplines and more prominent rise in non-STEMM; https://www.hesa.ac.uk/data-and-analysis/students/whos-in-he (accessed: 1 December 2025).

publications for 61% of PhD graduates. This share increases from just over 55% for PhDs who graduated in 2000 to approximately 63% for 2013 graduates, before stabilising.

2.2.1 Publishing during the PhD

The share of UK PhD graduates who published already during their studies, defined as the time from three years before the dissertation publication year and one year after, has seen a steep increase (see Figure 5). Starting at around 37% in 2000, publishing during the PhD reaches a peak of approximately 56% in 2020. This upward trend suggests that publishing has become an increasingly integral part of the doctoral experience over the past two decades.

The increase in publishing is seen in STEMM and non-STEMM fields, though, as we would expect, the average publication propensity is significantly higher in STEMM fields (65.27% vs 28.74%), reflecting different disciplinary publishing cultures. Figure 5 shows that publication propensity in STEMM fields grew from approximately 49% for PhDs who graduate in 2000 to a peak of nearly 70% for 2020 graduates. In the case of non-STEMM, we see a growth from about 19% in 2000 to almost 36% in 2020, which corresponds to a 70% increase in publishing propensity. This indicates that publishing during the PhD has increased across all fields, but remains most pronounced in STEMM.

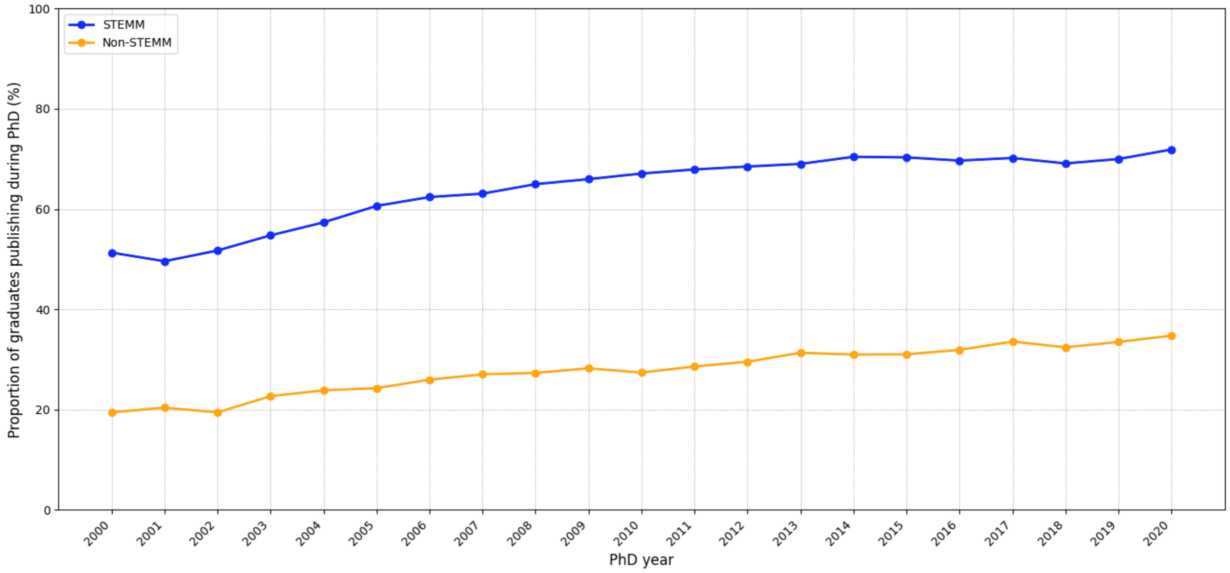

Figure 5: Trend of publications among UK PhD graduates (2000–2020) within STEMM and non-STEMM disciplines.
Note: Share of dissertations matched with at least one publication during the PhD period (t-3 to t+1). Source: Own analysis based on EThOS metadata matched with OpenAlex publication information

Figure 6 provides a breakdown by the seven subject domains. The propensity to publish during the PhD is highest in Physical Sciences with 65.3%, followed by Medicine and Life Sciences (all above 60%) and is lowest in the Arts and Humanities with fewer than 20% of graduates publishing during the PhD.

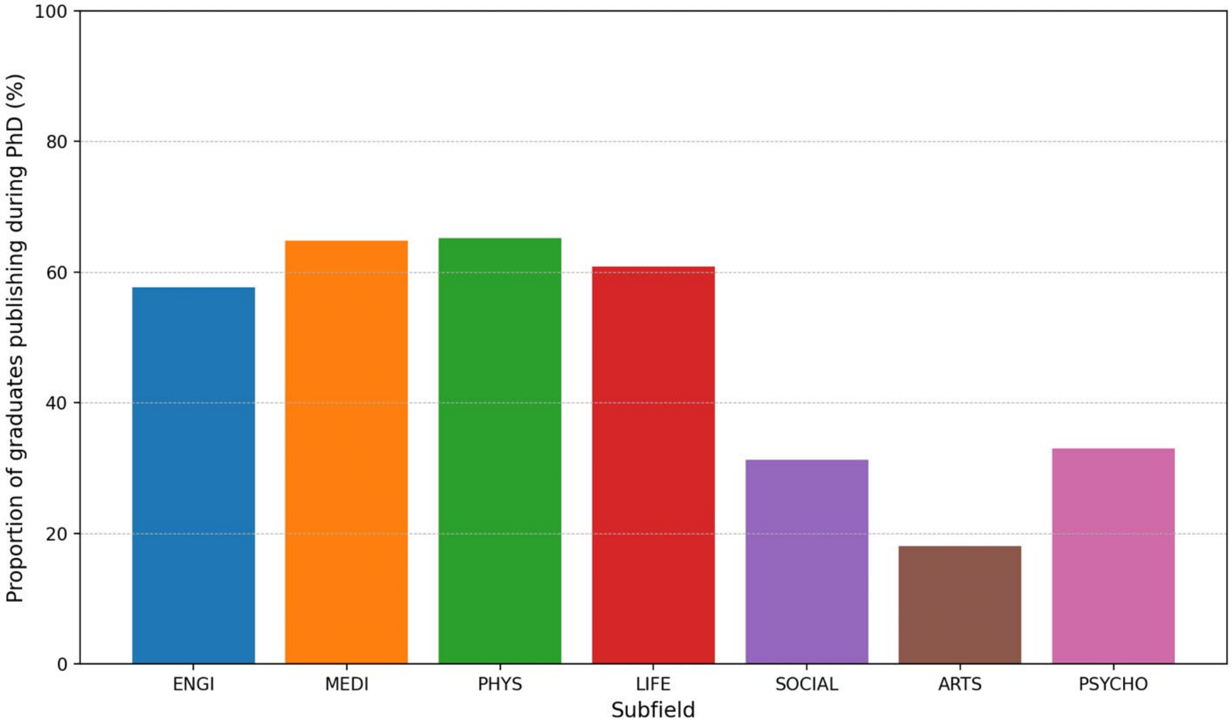

Figure 6: Proportion of UK PhD Graduates Publishing During PhD Period by Subfield (2000–2020).
Note: PhD Graduates Matched with Publication Authors Who Have At Least One Publication During PhD Period (t-3 to t+1) By Discipline. STEMM subfields include ENGI, MEDI, PHYS, LIFE; non-STEMM subfields include SOCIAL, ARTS, and PSYCHO. Psychology in EThOS is defined as Philosophy, Psychology & Religious Studies.
Source: Own analysis based on EThOS metadata matched with OpenAlex publication information

Prior literature has highlighted the relative lower publication performance of women academics (Bentley, 2012; Muric et al., 2021; Aksnes et al., 2025). In Figure 7, we show that in STEMM disciplines, men have a slightly higher propensity to publish during the PhD compared to women, with 64.0% of men publishing compared to 62.4% of women when considering the full period. This is in contrast to non-STEMM disciplines, where women account for a slightly higher share than men, with 28.8% of women publishing compared to 27.7% of men. Overall, the time trends are similar and the difference very small.

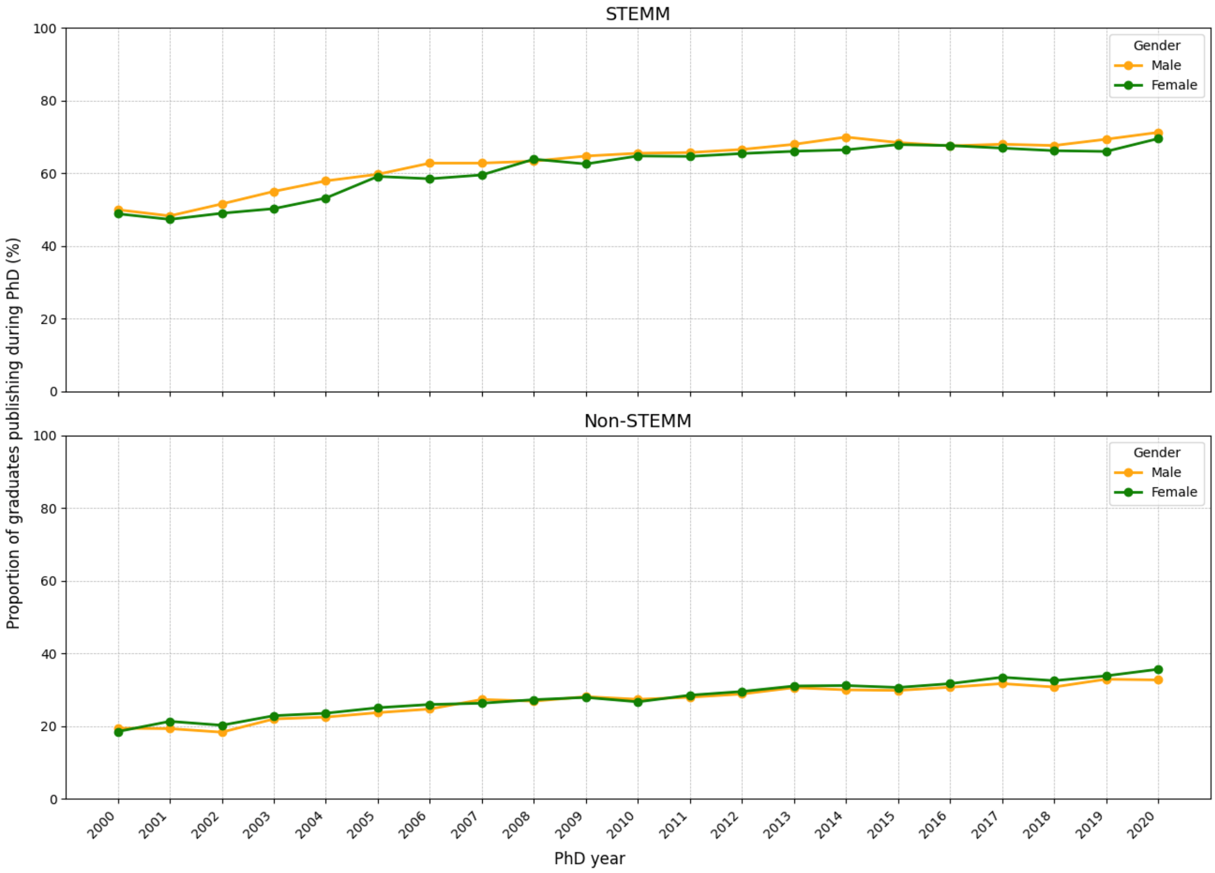

Figure 7: Share of UK PhD graduates publishing during PhD period by gender (2000–2020).
Source: Own analysis based on EThOS metadata matched with OpenAlex publication information; gender prediction using WIPO WGND.

2.2.2 Publishing after the PhD

The propensity to publish after the PhD, that is from 2 years after the dissertation publication date, has remained relatively stable at around 45% to 50%. Figure 8 reports post-PhD publication propensity by subject domain, showing that amongst STEMM graduates between 50% and 58% publish after the PhD, while amongst non-STEMM graduates this rate is between 30% and 35%. The decline in later years and especially after 2017 is due to data truncation issues, with more recent cohorts of graduates having fewer years to accumulate publications. The higher propensity amongst STEMM graduates to publish post-PhD indicates that a larger proportion successfully transition into a research career that requires sustained publication output, whether in academia or industry.

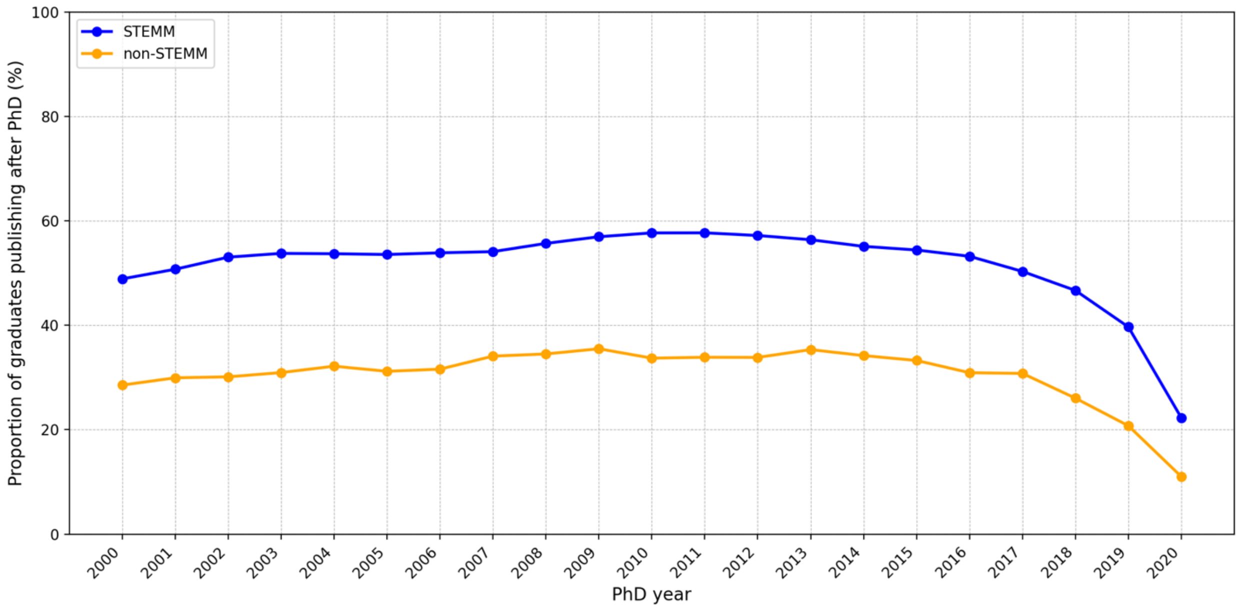

Figure 8: Share of UK PhD graduates publishing from after PhD by STEMM and non-STEMM disciplines (2000–2020).

Note: Dissertation authors matched with at least one publication after PhD (t+2 onwards). Source: Own analysis based on EThOS metadata matched with OpenAlex publication information

Figure 9 reports the breakdown by subfield. Medical and Life Sciences have the highest publishing rates, at 56.4% and 55.3%, respectively, followed by Physical Sciences at 51.5%, demonstrating strong research continuity after the PhD. This is followed by Engineering, where around 42.6% graduates continue to publish. Social Sciences and Psychology show moderate engagement at around 34% (34.2% and 34.1%, respectively), while Arts & Humanities has the lowest rate at 20.6%, confirming the lowest engagement with journal-article publication across disciplines but with rates comparable to publishing figures during the PhD (see Figure 6).

Figure 10 reports publishing rates by gender. In STEMM disciplines, men account for a slightly higher share of publications after the PhD compared to women, with 54.7% of men publishing compared to 52.2% of women, a gap that is closing over time. In non-STEMM disciplines, women and men publish at similar rates (31.9% of women compared to 31.5% of men).

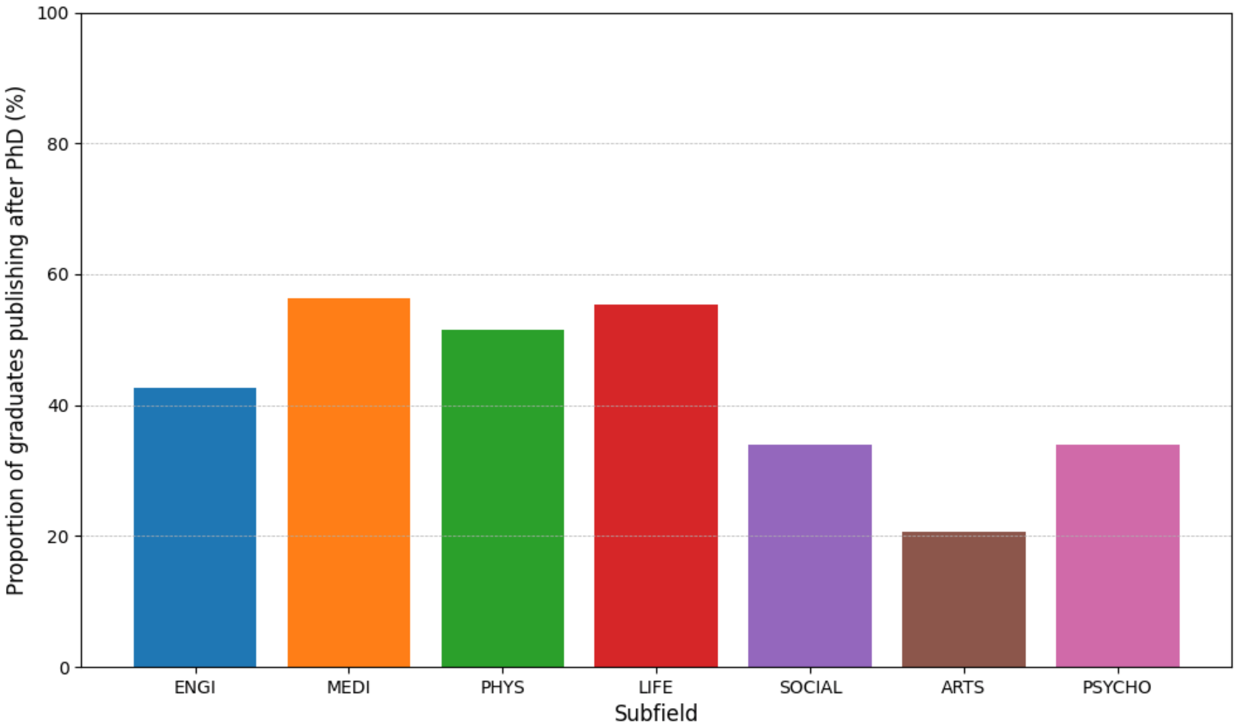

Figure 9: UK PhD graduates publishing from 2 years after PhD by subfield (2000–2020).

Note: PhD graduates' theses matched with publication authors who have at least one publication from 2 years after PhD, by subfields. STEMM subfields include ENGI, MEDI, PHYS, LIFE; non-STEMM subfields include SOCIAL, ARTS, and PSYCHO. Psychology in EThOS is defined as Philosophy, Psychology & Religious Studies.

Source: Own analysis based on EThOS metadata matched with OpenAlex publication information

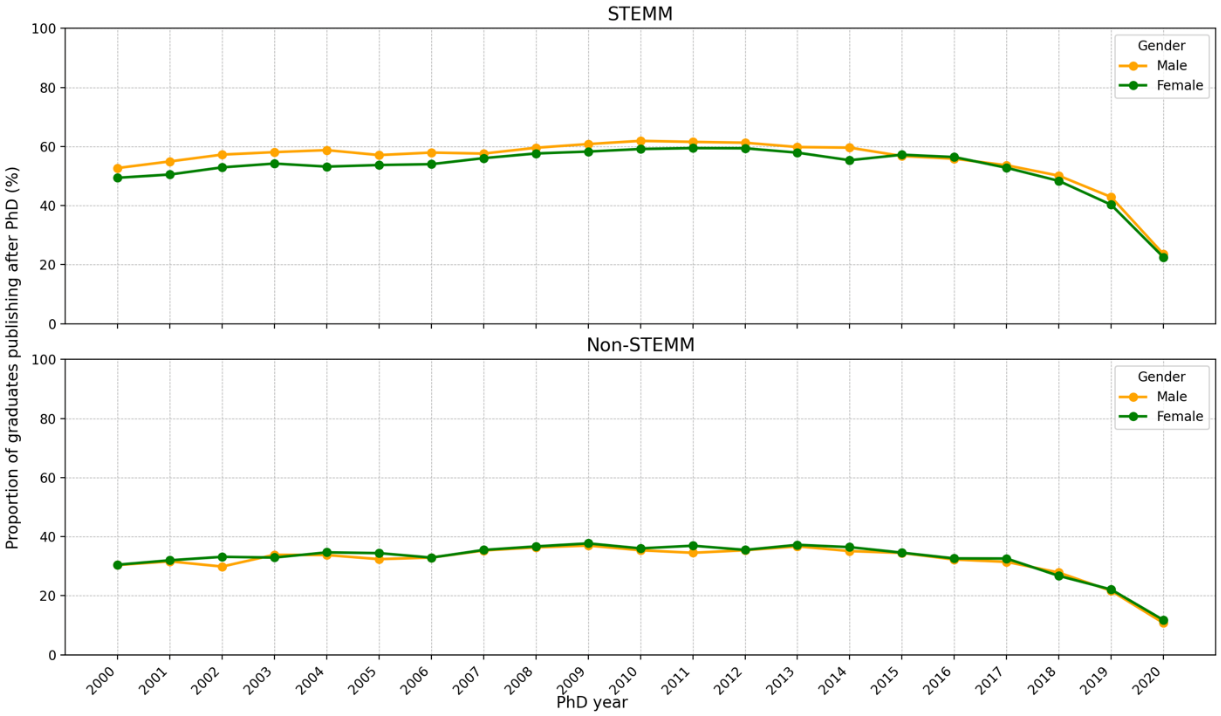

Figure 10: Share of UK PhD graduates publishing after PhD period by gender (2000–2020).

Source: Own analysis based on EThOS metadata matched with OpenAlex publication information; gender prediction using WIPO WGND

2.3 Publication numbers during and after PhD

The previous section showed the publishing propensity of PhD graduates, but now we shift our focus to the number of publications per PhD graduate. We focus on all PhD graduates, not only those who published, i.e. allowing for the number of publications to equal zero.

2.3.1 Publishing during the PhD

Figure 11 presents a clear upward trend in the average number of publications per UK PhD graduate during their PhD period. In STEMM disciplines, the average increases from around 1.6 publications to more than 3.5 in recent cohorts. Non-STEMM disciplines also show a steady rise, from approximately 0.5 to more than one publication on average.

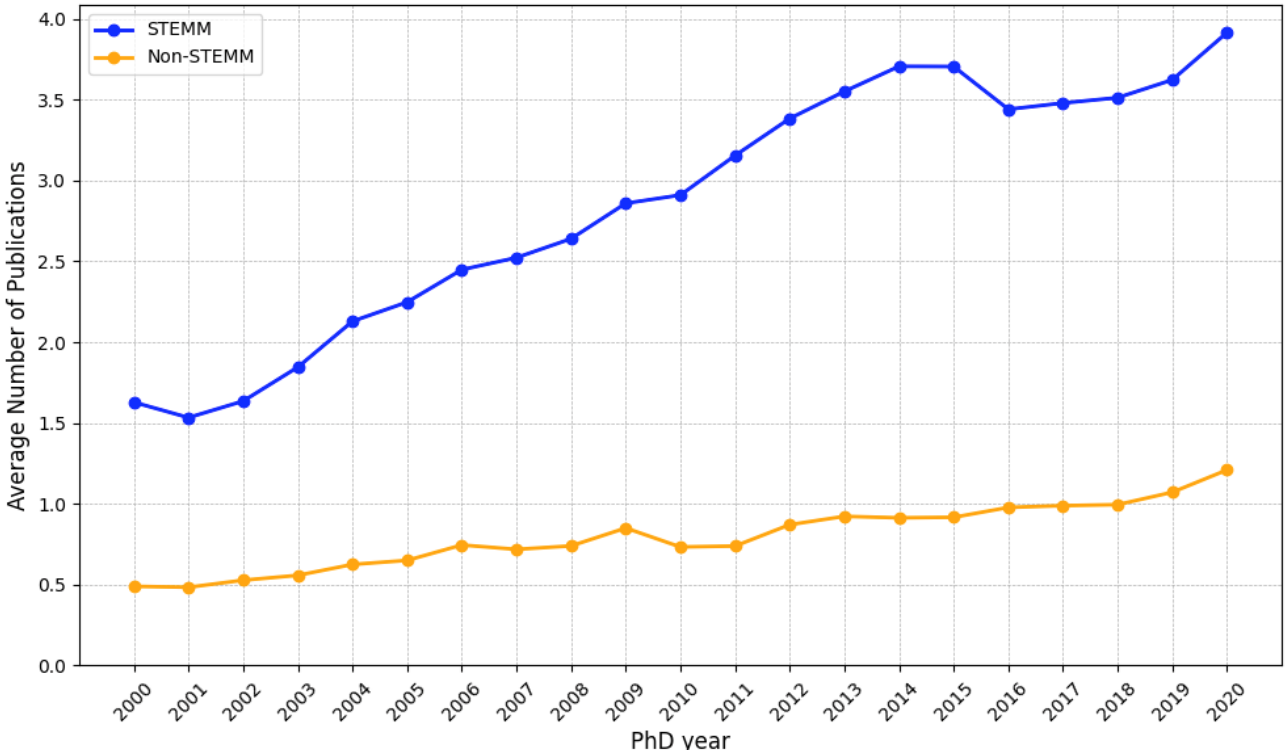

Figure 11: Average number of publications per UK PhD graduate during PhD, STEMM vs non-STEMM. Source: Own analysis based on EThOS metadata matched with OpenAlex publication information

It is worth reflecting on why this upward trend is not strictly monotonic, and why there may be a dip or plateau in output during certain time periods. One of the great drivers of publication behaviour in the UK research system has been research evaluation (Research Excellence Framework (REF) and its predecessor Research Assessment Exercise (RAE)), which has been shown to create short-term incentives affecting the timing, volume and focus of publications (e.g. Groen-Xu et al., 2023; Marcella and Lockerbie, 2016; Hughes et al., 2019), with surges immediately preceding submission deadlines and slower output growth immediately afterwards. This dynamic provides a plausible explanation, in the context of Figure 11, for the flattening or slight decline in average publications visible for PhD cohorts whose training roughly overlapped with the post-RAE-2008 and REF-2014 periods. It may be that, as university researchers shifted attention away from rapid output generation following the RAE/REF submission, publication rates temporarily stagnated or dropped before resuming their longer-term upward trend.

Figure 12 shows the distribution of publications across the seven academic subfields. We present distributions, rather than means, due to the large variation in publication numbers within fields. Graduates in Physical Sciences are most productive, exhibiting the highest median publication count at 2, with quarter publishing more than 4. other STEMM fields show

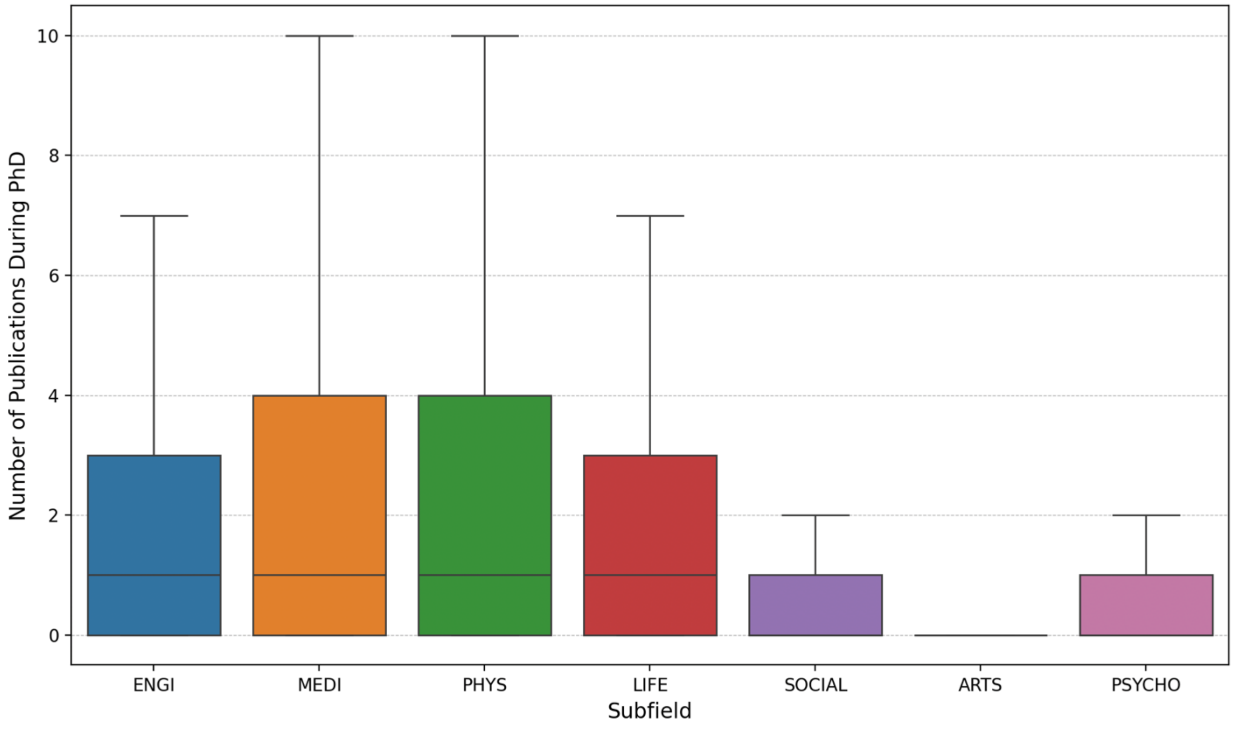

Figure 12: Distribution of publication outputs across seven academic subfields per PhD During PhD Period (t-3 to t+1).

Note**:** Each box-and-whisker plot summarizes the variation in publication counts within a field. The horizontal line inside each box represents the median number of publications, while the box boundaries indicate the interquartile range (IQR), capturing the middle 50 percent of PhD graduates. The top whiskers indicate the highest publication counts that fall within 1.5 times the interquartile range (IQR) above the 75th percentile; values beyond this threshold are treated as outliers and are not shown. STEMM subfields include ENGI, MEDI, PHYS, LIFE; non-STEMM subfields include SOCIAL, ARTS, and PSYCHO. Psychology in EThOS is defined as Philosophy, Psychology & Religious Studies. Source: Own analysis based on EThOS metadata matched with OpenAlex publication information

a median publication count of 1, indicating that most PhD students published just one paper. However, in Medicine the 75th percentile is equivalent to Physical Sciences, suggesting a highly skewed publication distribution with some very productive PhDs. PhDs in non-STEMM fields have much lower numbers of publications, and a median of zero, indicating that a majority of PhD students in these fields did not publish during their PhD. Arts & Humanities shows the lowest numbers with almost all graduates not publishing during their PhD, which likely reflects differences in subject norms.

Figure 13 reports the mean publication count during the PhD period for men and women, differentiating by STEMM and non-STEMM fields and cohorts. While we can see an increase in the number of publications for male and female graduates over time, we also see a persistent gender gap in the number of publications. The difference is particularly pronounced in STEMM disciplines, where male graduates had an average of 3.3 publications compared with 2.4 for female graduates, with the gap widening over time and being particularly pronounced for cohorts after 2012. In non-STEMM fields, publication rates are lower overall, with men averaging 0.9 publications and women 0.8 during their PhD years.

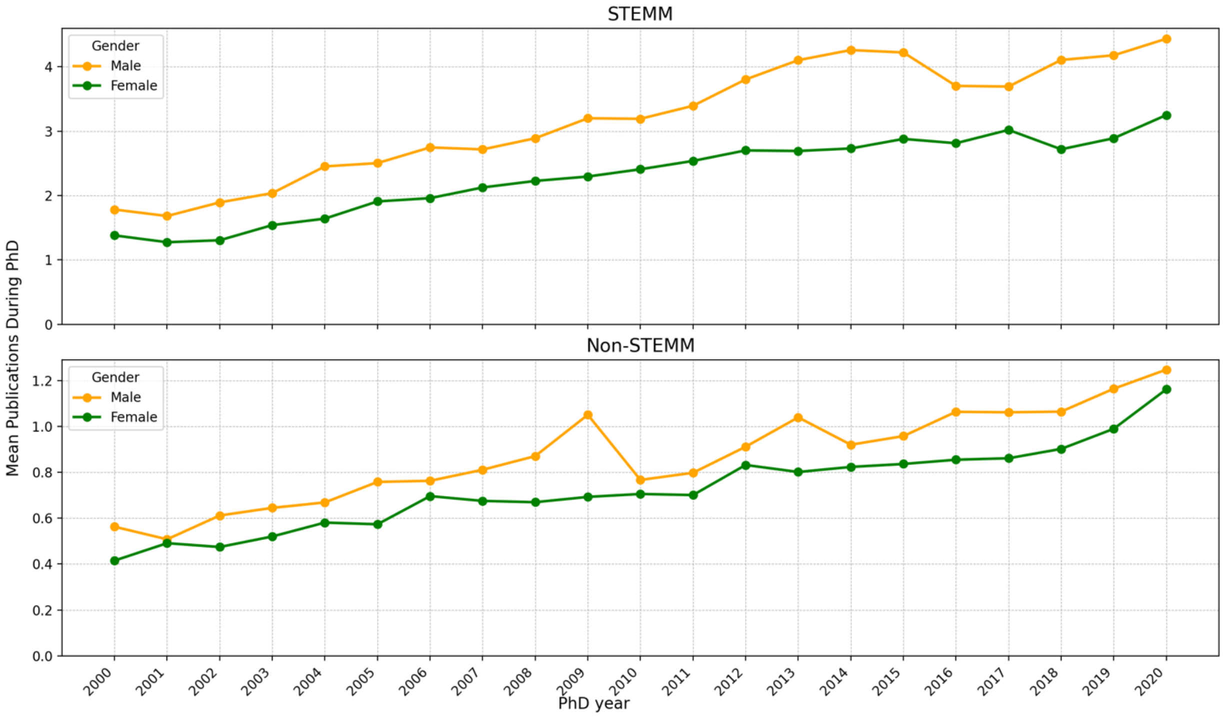

Figure 13: Mean number of publications produced during the PhD period (−3 to +1 years) by gender, 2000–2020.

Source: Own analysis based on EThOS metadata matched with OpenAlex publication information; gender prediction using WIPO WGND

2.3.2 Publishing after the PhD

Figure 14 reports the average number of publications per UK PhD graduate within a five-year window following completion of the PhD (that is, from 2 years after the PhD to account for publication lags; e.g., publications during 2017-2021 for the 2015 cohort). Note that for most recent cohorts we are unable to observe the full 5-year window and therefore report publication numbers for PhD holders who graduated up to 2017 (allowing at least three full years from 2019 until 2021, the last year for which complete publication data is available). The data reveal a steady upward trend in post-PhD research output, starting at around 3.5 publications for 2000 graduates in STEMM fields and reaching more than 5.5 publications for the 2010 to 2014 cohorts. The decline for recent cohorts is partially due to right-censoring, with incomplete publication data for 2016 and 2017 cohorts, but may also be exacerbated by the impact of COVID-19 pandemic, which delayed research and publication for most recent cohorts across subject domains (Gao et al. 2021).

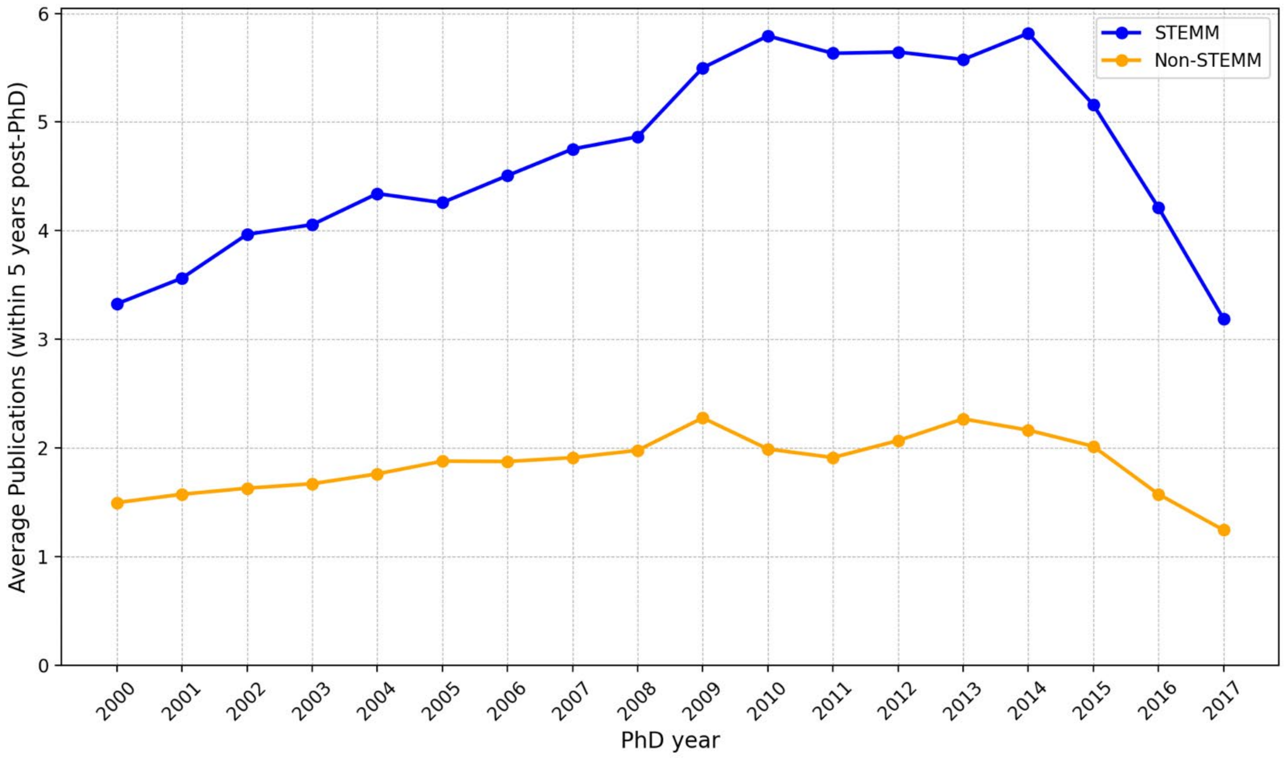

Figure 14: Mean number of publications per UK PhD graduate with 5 year-window post-PhD (+2 to +7 years), cut-off at 2017.

Source: Own analysis based on EThOS metadata matched with OpenAlex publication information. Publication data complete until 2021.

Figure 15 shows the number of post-PhD publications broken down by the seven subject categories. Medicine, Physical Sciences, and Life Sciences all exhibit a median of one post-PhD publication, indicating comparable publication activity across these STEMM fields. However, Medicine displays a wider upper distribution, with the 75th percentile reaching 6 publications, compared to 5 in both Physical and Life Sciences. This suggests that while typical post-PhD output is similar across these domains, Medicine includes a larger subset of highly productive graduates. Post-graduation publication numbers in other domains remain relatively low, with medians of zero in Engineering, Social Sciences, and Psychology, suggesting that many PhDs in these disciplines do not publish during the immediate post-PhD period. Still, with values of 2 and above at the 75th percentile, we can see a strong core of graduates that continue publishing a large number of works. Arts & Humanities again shows the lowest numbers with almost all graduates not publishing, which partially reflects different subject norms rather than a lack of research activity.

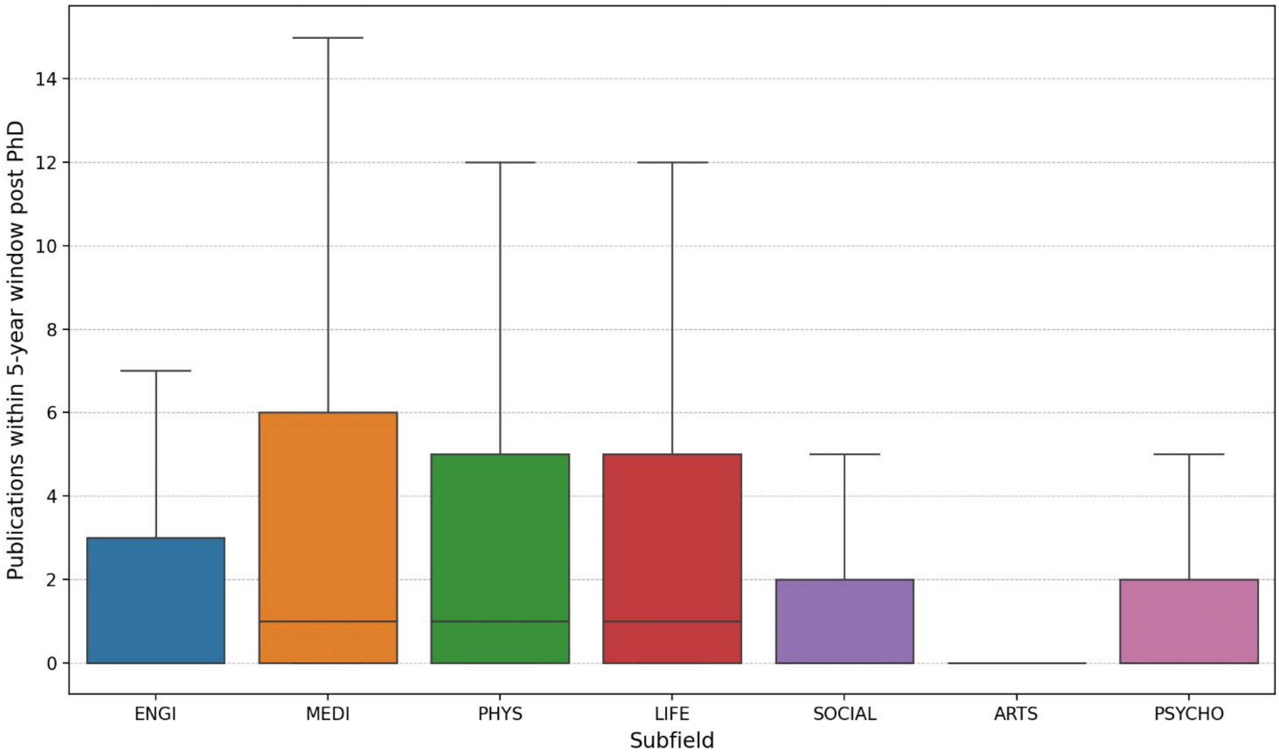

Figure 15: Distribution of Publications per UK PhD graduate with 5 year-window post-PhD (+2 to +7 years) by subfield, cut-off in 2017.

Note**:** Each box-and-whisker plot summarizes the variation in publication counts within a field. The horizontal line inside each box represents the median number of publications, while the box boundaries indicate the interquartile range (IQR), capturing the middle 50 percent of PhD graduates. STEMM subfields include ENGI, MEDI, PHYS, LIFE; non-STEMM subfields include SOCIAL, ARTS, and PSYCHO. Psychology in EThOS is defined as Philosophy, Psychology & Religious Studies. Source: Own analysis based on EThOS metadata matched with OpenAlex publication information

Figure 12, which showed distributions during the PhD can provide context for the post-PhD outcomes presented in Figure 15. Fields that are highly productive during the PhD, such as Physical Sciences and Medicine, tend to show a continuation of publication activity after graduation, although the median counts adjust slightly. Fields with lower PhD-period outputs, particularly non-STEMM subfields, largely maintain low publication activity post-PhD, highlighting a persistence of early publication behaviour and of disciplinary norms also in subsequent research output trajectories.

Figure 16 shows the number of publications post-PhD for men and women graduates. The output for male STEMM graduates averages 5.3 publications and for female STEMM graduates 3.7, thus showing a persistent gender gap also post-PhD, which has widened over time, but closed slightly in most recent years. The difference is smaller in non-STEMM where male graduates produced 2.0 publications on average, while their female counterparts averaged 1.7. Fluctuations across years should be read with caution and in the context of fluctuations in data records, publication data coverage, and methods used for gender prediction. The overall gender trend is similar for STEMM and non-STEMM fields.

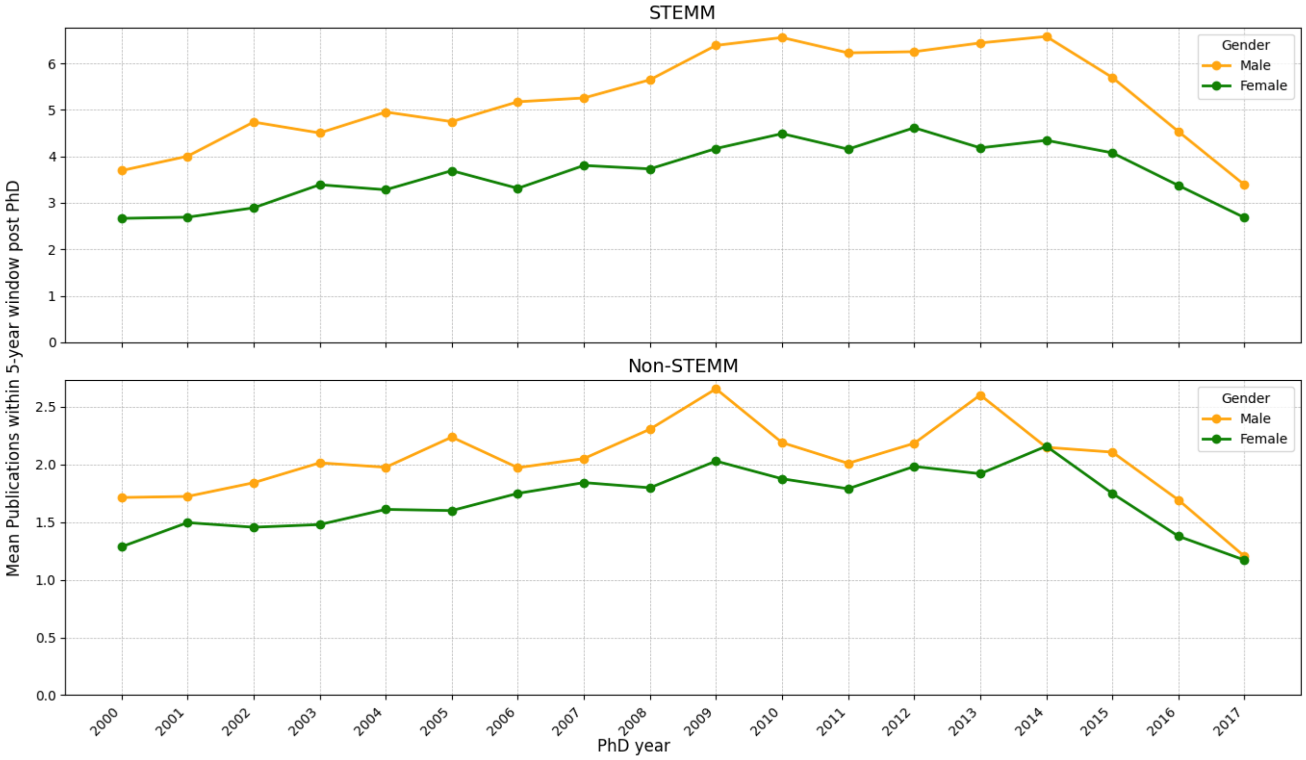

Figure 16: Mean number of publications produced with 5 year-window post-PhD (+2 to +7 years) by gender, cut-off at 2017.

Source: Own analysis based on EThOS metadata matched with OpenAlex publication information; gender prediction using WIPO WGND

2.4 STEMM Doctoral graduates' publications as knowledge inputs for patents

The previous sections established that a significant percentage of graduates publish during the PhD and continue to publish once they graduated. This section explores whether these publications serve as input to innovation, i.e. whether they are cited by patents. In this section, we focus on STEMM doctoral graduates as these are not only more likely to publish but also more likely to conduct research that has the potential to be cited in patents. We utilise Reliance on Science (Marx and Fuegi 2020a, 2020b), a database of USPTO and EPO patents[7] and their citations to scientific publications, including both in-text and front-page references.

Figure 17 reports the share of graduates who have at least one of their PhD period publications (that is from three years before to one year after the year of PhD award) cited directly by a

[7] USPTO and EPO are often considered the most relevant for academic studies due to their wide reach and ease of use (Kim and Lee, 2015). Reliance on Science data has reduced coverage for other patent offices.

patent by STEMM field. We only consider graduates with at least one PhD-period publication. The overall share of publishing PhDs being cited by a patent is around 32%, with Medical Sciences accounting for the highest proportion with 41.5% of graduates cited in a patent, followed by Life Sciences at 34.2% and Engineering at 26.4%. Physical Sciences account for the smallest proportion at 25%.

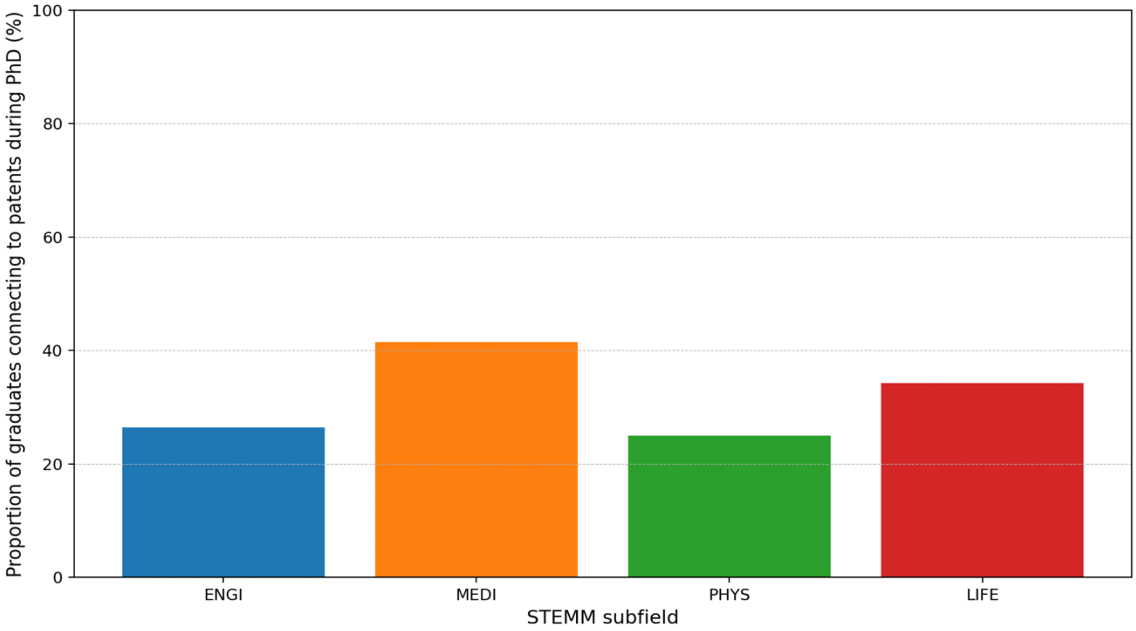

Figure 17: Share of UK STEMM PhD graduates with at least one publication during PhD period (from t-3 to t+1) connected to a patent, by STEMM field.

Source: Own analysis using EThOS metadata, publication matches from OpenAlex, patent linkages from Reliance on Science

Figure 18 reports the share of graduates that are close to the technological frontier by cohort and gender. Female graduates are slightly less likely to be cited in a patent (30.5% vs. 33.1% of men considering all cohorts). This difference is small and substantially lower than gender differences in patenting reported in the extant literature (Ding et al., 2006; Hunt et al., 2012). This suggests that research by women is just as likely to have commercial potential as that of men. The share of graduates being cited is about 40% for early cohorts and then declines. The decline is due to right censoring with truncation, with more recent cohorts having had less time to accumulate citations to their PhD work, rather than a genuine decrease in research–innovation linkage. There is a substantial time-lag between the publication of a scientific article and its citation in patents, as evidenced by Marx and Fuegi (2022b), who find that articles cited in patents are referenced approximately 14 to 17 years after publication, reflecting the slow and cumulative nature of knowledge diffusion from science to technological application. The decline in propensity to be cited thus does not imply a reduced innovation potential.

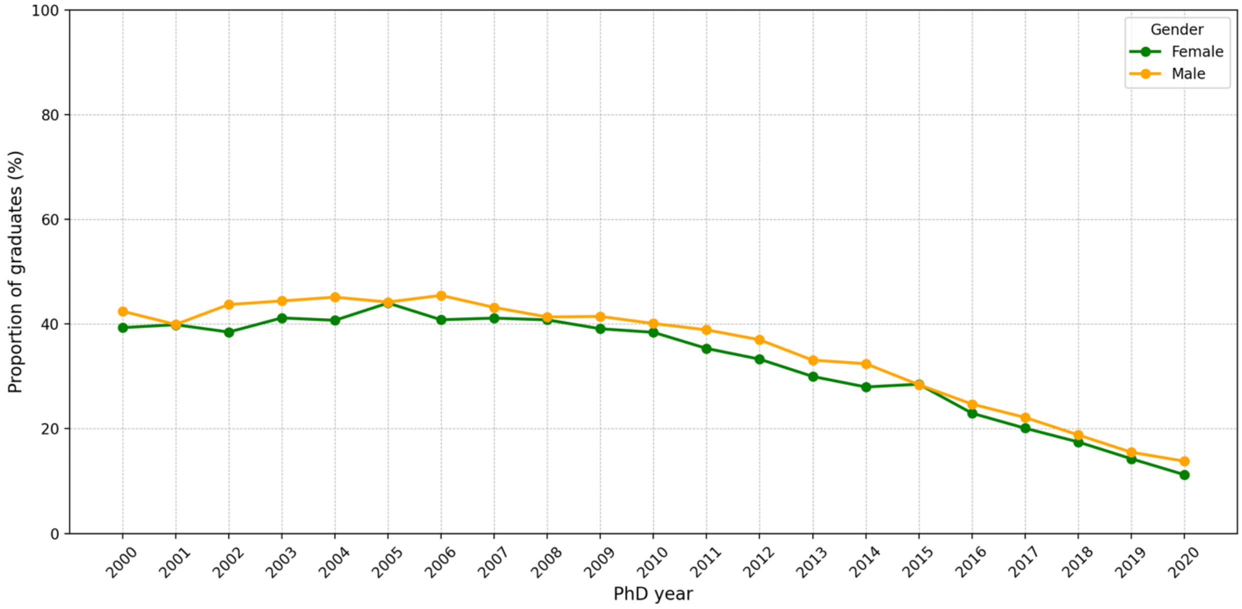

Figure 18: Share of UK STEMM PhD Graduates with at least one publication connected to a patent during PhD period (from t-3 to t+1).

Source: Own analysis using EThOS metadata, publication matches from OpenAlex, patent linkages from Reliance on Science and gender prediction using WIPO WGND

Figure 19 illustrates the share of UK STEMM PhD graduates who produced at least one post-PhD publication that is linked to a patent by STEMM field. Medical Sciences remains the most important with a citation propensity at 35.6%, followed by Life Sciences at 31.5%. Physical Sciences and Engineering remain low at 20-21%. Comparing these post-PhD outcomes to Figure 17, which reported patent citation of articles published during the PhD, we find that the relative ordering of subfields is consistent, with Medical and Life Sciences maintaining the highest citation propensity. The decline in proportions compared with PhD period publications can again partially be explained by right-censoring due to citation delays rather than a reduction in technology relevant works.

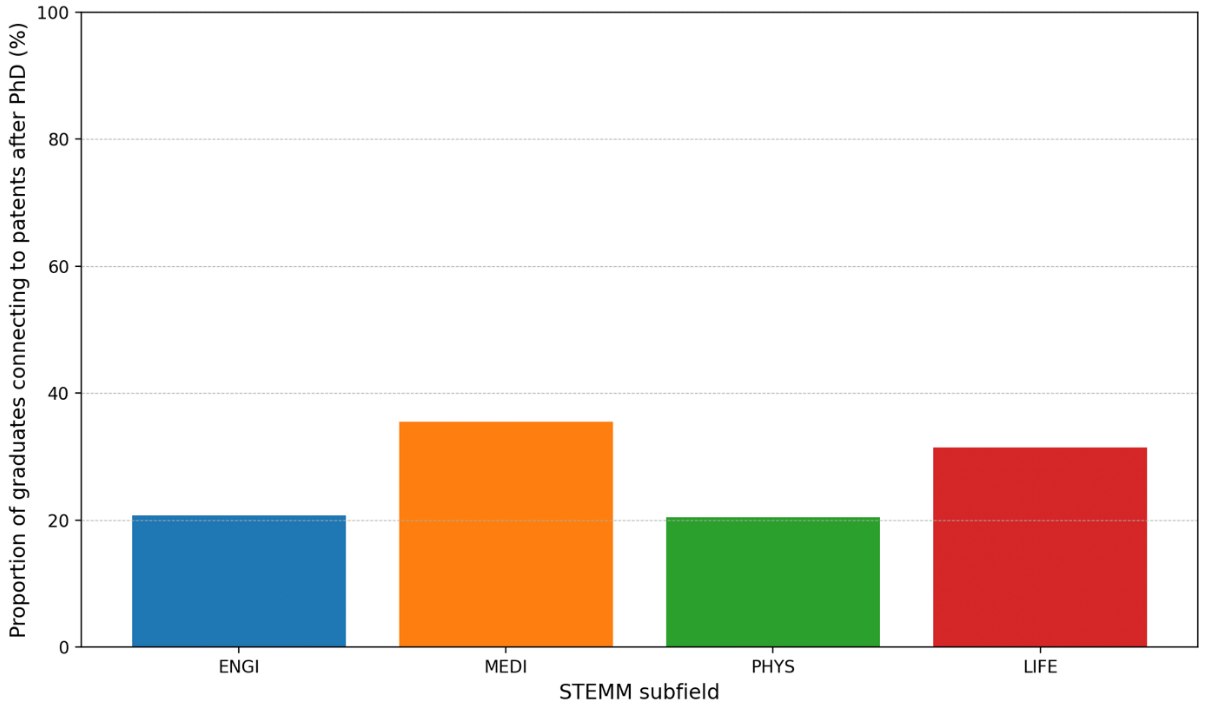

Figure 19: Share of UK STEMM PhD graduates with at least one publication connected to a patent after PhD, measured 2 years after the PhD until 2022, by STEMM field.

Source: Own analysis using EThOS metadata, publication matches from OpenAlex, patent linkages from Reliance on Science

Figure 20 reports the share of graduates whose post-PhD work is cited in a patent by cohort and gender. There is again a lower proportion of women to get cited with 25.1%, compared to 28.3% for male graduates, when considering all cohorts. The share of publishing PhDs connected to a patent is about 40% for early cohorts, comparable to Figure 18, and again declines for more recent cohorts, consistent with right-censoring due to truncation in the patent linkage outcomes. The drop is more severe compared to Figure 18, due not only to citation lags but also the larger number of publications observed for early cohorts, and thus an increased likelihood that at least one of these papers is connected to a patent. Overall, the results indicate that UK PhD graduates continue to produce research that has technology relevance, with women almost as likely as men to see their work informing invention.

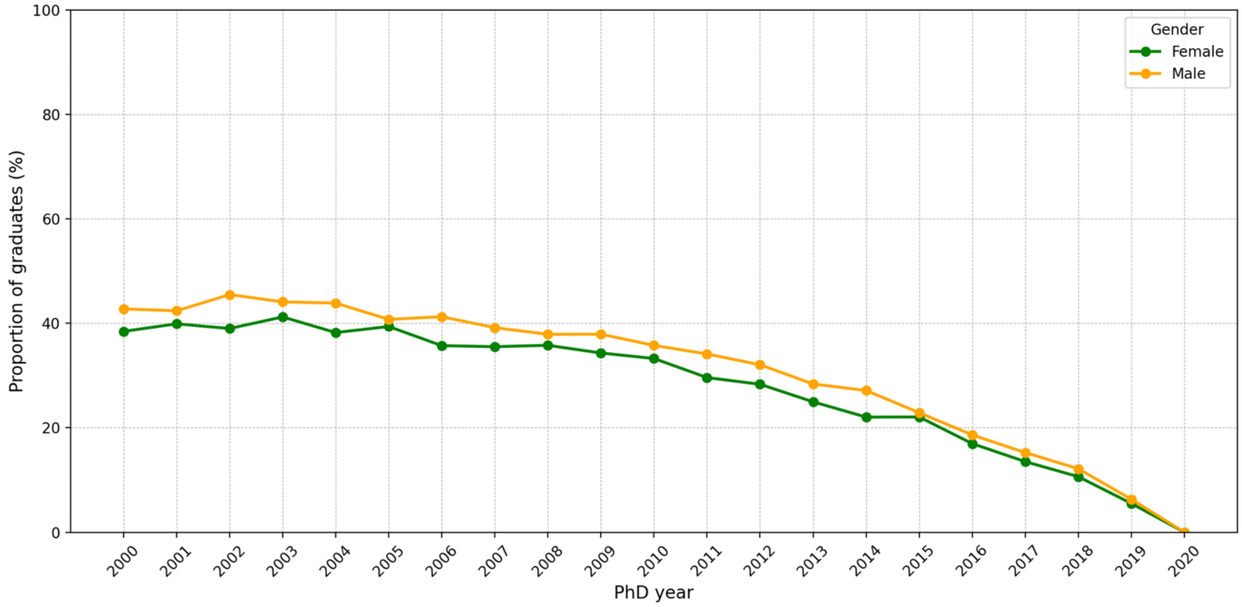

Figure 20: Share of UK STEMM PhD Graduates with at least one publication connected to a patent after PhD, measured from 2 years after the PhD until 2022.

Source: Own analysis using EThOS metadata, publication matches from OpenAlex, and patent linkages from Reliance on Science and gender prediction using WIPO WGND

2.5 Exploratory regression analysis

To interpret the science and innovation potential of doctoral graduates, it is of importance to understand how different characteristics of the PhD population interact. This section presents some exploratory regressions to uncover whether correlations observed in the previous sections hold in a multivariate setting.

We estimate two logistical regression models: the propensity to publish after the PhD and the propensity to be cited by a patent. Both models account for productivity during the PhD, gender, disciplines and graduation year effects. We report dot plots of the marginal effects with error bars to represent the regression models for ease of reading.

2.5.1 Propensity to publish after the PhD

Figure 21 reports the results for post-PhD publications activity. We find that that male graduates are slightly more likely to publish after the PhD than women, but the effect is very small indicating minimal gender differences in post-PhD publishing once we control for other factors.

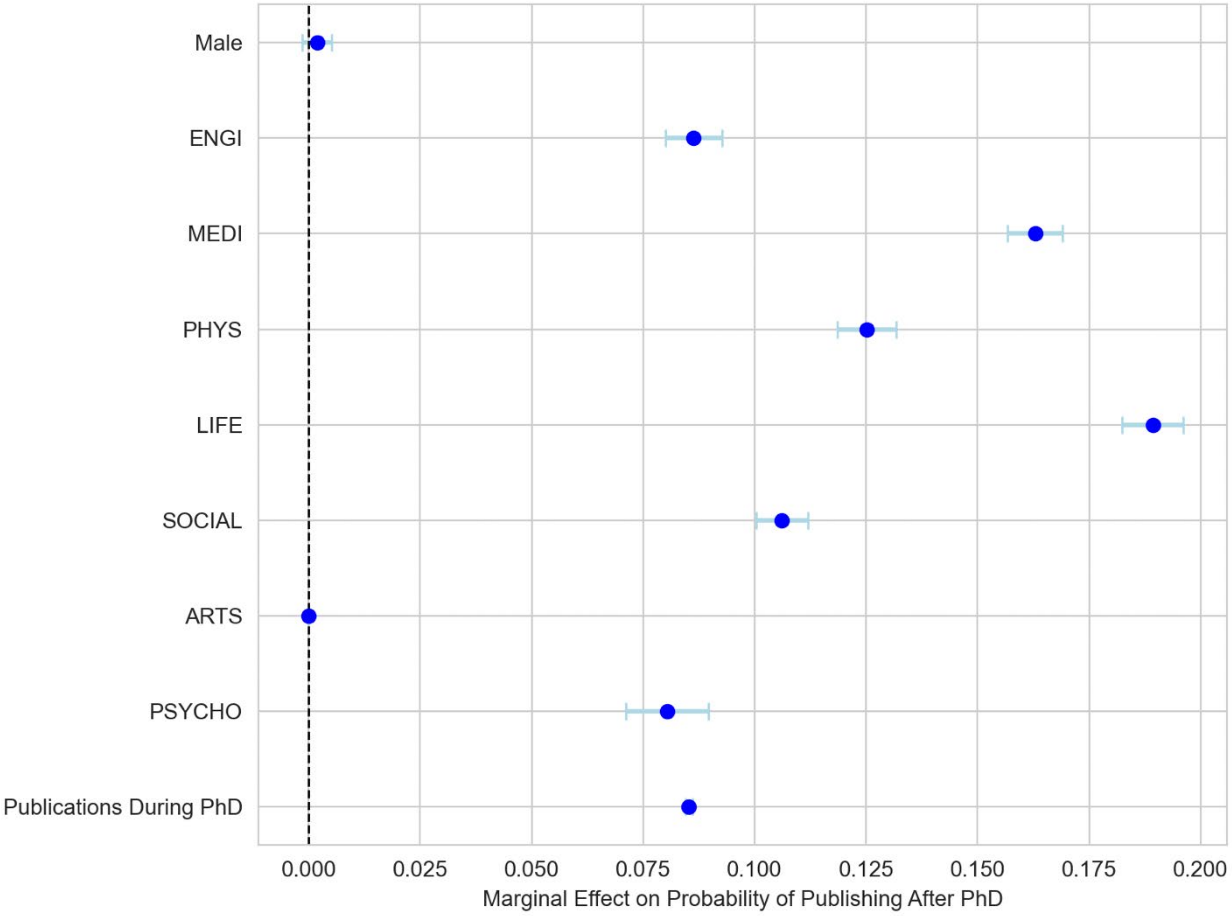

Figure 21: Marginal effects of PhD publications, gender, and academic discipline on the probability of publishing after the PhD.

Note: The estimates are derived from a logistic regression model fitted using the statsmodels Python package, with marginal effects computed and plotted with matplotlib and seaborn. The figure visualises the estimated change in probability associated with each variable, with error bars representing 95 per cent confidence intervals. ARTS (Arts & Humanities) is the reference category for academic discipline, so its marginal effect is zero. N = 292,163 observations. Source: Own analysis using EThOS metadata, publication matches from OpenAlex and gender prediction using WIPO WGND

Discipline is an important determinant of post-PhD publishing. Life Sciences are the most likely to continue publishing, with a marginal effect of 18.9 percentage points compared to the reference category (Arts & Humanities). Medical Sciences follow with a marginal effect of 16.3 percentage points. Physical Sciences graduates show a positive effect of 12.5 percentage points, Social Sciences 10.6 percentage points, Engineering 8.6 percentage points, and Psychology 8.0 percentage points. This demonstrates that Life and Medical Sciences exhibit the highest likelihood of post-PhD publishing, while PhDs in applied disciplines such as Psychology and Engineering are less likely to continue publishing, though the effect remains positive when compared to Arts & Humanities graduates. Productivity during the PhD has a strong influence on subsequent publishing. Each additional publication during the PhD increases the probability of publishing after the PhD by approximately 8.5 percentage points, holding all other factors constant. This indicates a strong persistence effect, showing that early research activity indicates continued engagement in scientific publishing. This is consistent with the extant literature, which suggests that publication experience during doctoral studies serves as a strong predictor of continued participation in research careers and sustained scientific productivity (Aksnes and Sivertsen, 2004; Larivière, 2012; Horta and Santos, 2016).

2.5.2 Propensity to be cited in a patent

Figure 22 reports the results for graduates' likelihood to be cited by a patent.

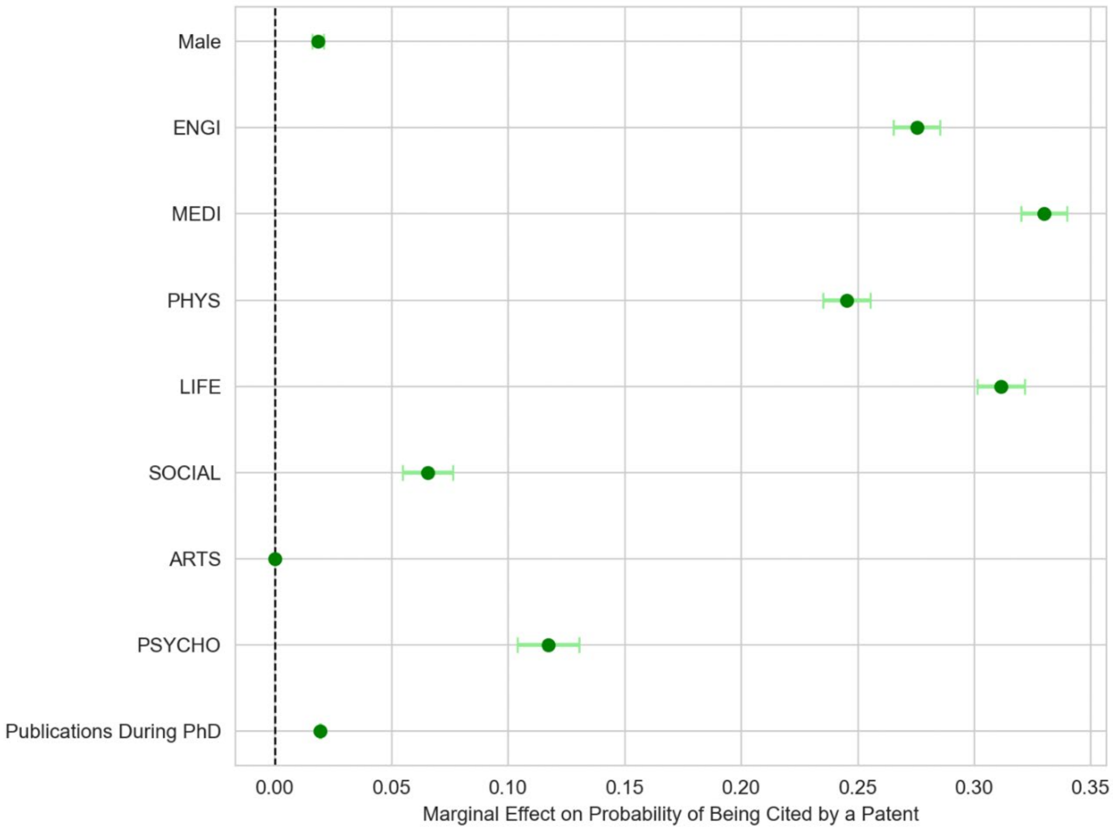

Figure 22: Marginal effects of PhD publications, gender, and academic discipline on the probability of publishing after the PhD the likelihood to be connect to a patent (with scientific publications cited by patents).

Note: The estimates are derived from a logistic regression model fitted using the statsmodels Python package, with marginal effects computed and plotted with matplotlib and seaborn. The figure visualises the estimated change in probability associated with each variable, with error bars representing 95 per cent confidence intervals. ARTS (Arts & Humanities) is the reference category for academic discipline, so its marginal effect is zero. N =

292,163 observations. Source: Own analysis using EThOS metadata, publication matches from OpenAlex, and patent linkages from Reliance on Science and gender prediction using WIPO WGND

Specifically, we measure the propensity to be cited in a patent as a binary indicator that is equal to 1 if the PhD graduate's publications have ever been cited in a patent and 0 otherwise. The results indicate that publishing during the PhD modestly increases the probability of a patent citation, with each additional publication raising the likelihood by approximately 1.9 percentage points, holding all else constant. Graduates in Medical Sciences are the most likely to be cited in patents, with an increase of around 33.0 percentage points compared to the reference category (Arts & Humanities), followed by Life Sciences (31.1%), Engineering (27.5%), Physical Sciences (24.5%), Psychology and Social Sciences display lower probabilities, at approximately 11.7% and 6.5%, respectively, demonstrating substantial variation across fields but also indicating that even Social Sciences provide relevant inputs into technology. Gender exerts a small but statistically significant effect, with male graduates being roughly 1.83 percentage points more likely than female or unclassified graduates to receive a patent citation. These results show that while there is little to no difference in publishing, male graduates are slightly more likely to produce research that is close to the technological frontier.

3. Regional Variation in UK PhD graduates

The science and innovation potential of the UK PhD population exhibits substantial regional differences. In this section we report key statistics by UK ITL2/NUTS2 locations of universities that awarded PhD dissertations.[8] The left hand panel of Figure 23 reports the share of STEMM PhD dissertations across UK regions. The colourings show a degree of heterogeneity, with the proportion of STEMM PhDs ranging from 58% to 64% of all doctorates. The areas with the highest concentration of STEMM PhDs (darkest blue, 62% to 64%) are located in Northern Ireland, Scotland, East Midlands and East England. By contrast, the lowest shares of STEMM, and hence highest shares of non-STEMM doctorates, (lightest purple, <58%) are concentrated in the West Midlands and Oxford. This suggests a regional

[8] Regional boundaries in maps are based on vector polygons of UK administrative subdivisions provided by the Natural Earth package and giscoR package in R.

divide, with the constituent countries of the UK (Scotland and Northern Ireland) and Eastern areas of England showing a stronger focus on scientific and technical doctoral education.

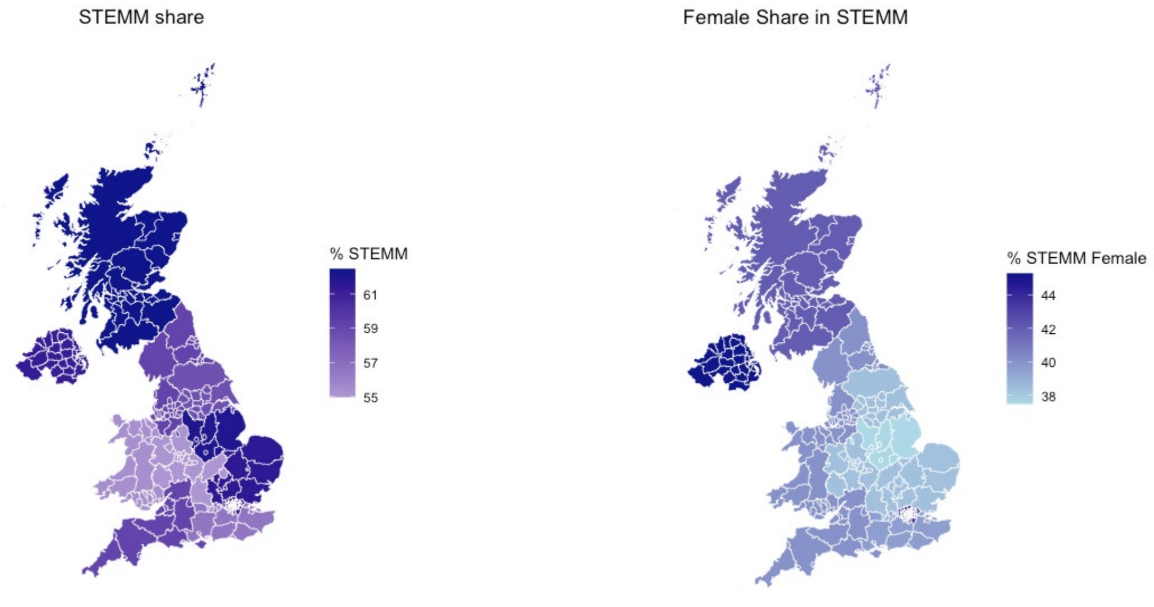

Figure 23: Regional variation in UK PhD graduates (2000–2020).

Left: Share of total PhD theses in STEMM across the UK, calculated as the number of STEMM theses in each region divided by the total number of PhD theses in that region, expressed as a percentage. Right: Share of women among STEMM PhD dissertations by region, including gender predictions derived from the WIPO WGND dataset. Data are based on analysis of EThOS metadata; map polygons sourced from rnaturalearth and visualized using the R sf and ggplot2 packages

The right-hand panel of Figure 23 maps the share of women in UK PhD dissertations in STEMM fields across regions. The underrepresentation of women in STEMM identified in section 2 also reflects in regional variation. Northern Ireland shows the highest proportion of female STEMM PhD students, with some areas exceeding 44%. Scotland and parts of the South West also exhibit relatively high shares, around 42%. By contrast, much of England, particularly the Midlands and East of England, has lower representation, with some regions as low as 38%.

Figure 24, left side panel, illustrates the geographic distribution of STEMM doctoral graduates who published during their PhD period across the UK. We focus on STEMM PhD graduates as these are more likely to publish and particularly critical to innovation. The highest publication rates appear in Scotland and parts of East England and the London area, falling within a range of 71% to 75%, suggesting a higher propensity for publication in these areas. By contrast, the lowest publication shares, around 69–71%, are in Wales and Northern

Ireland, indicating comparatively lower proportions of STEMM doctoral graduates who published during their studies.

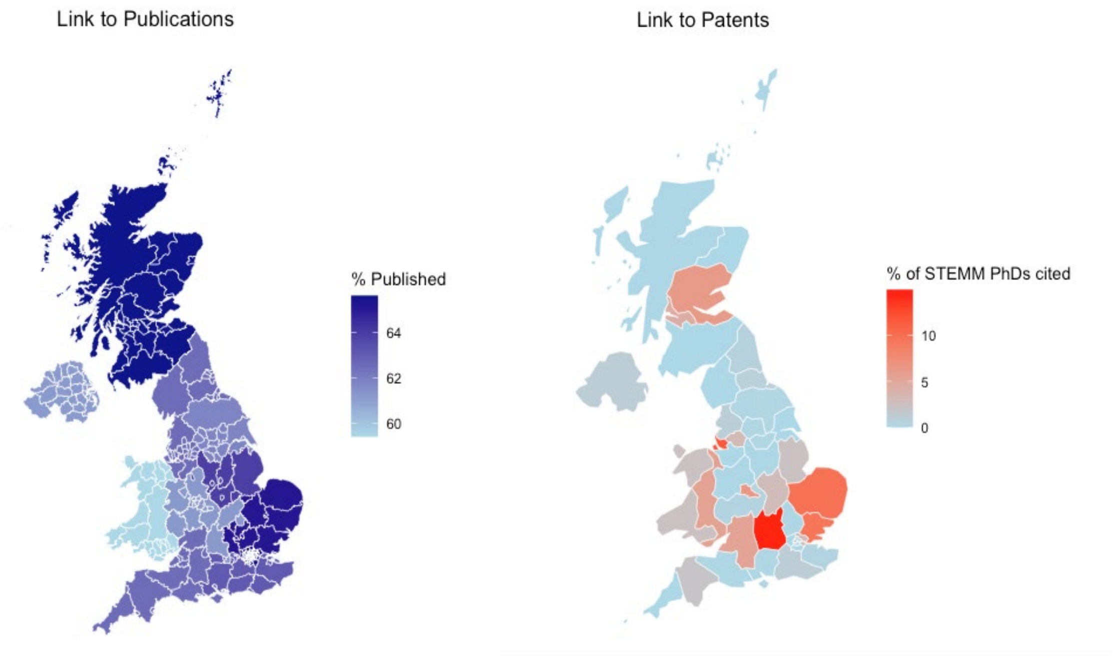

Figure 24: Regional variation in Science and Innovation Contribution of UK PhD graduates (2000–2020). Left: Share of total PhD theses in STEMM across the UK, calculated as the number of STEMM theses in each region divided by the total number of PhD theses in that region, expressed as a percentage. Right: Share of women among STEMM PhD dissertations by region, including gender predictions derived from the WIPO WGND dataset. Data are based on analysis of EThOS metadata; map polygons sourced from rnaturalearth and visualized using the R sf and ggplot2 packages

The potential of these STEMM PhDs to contribute towards innovation likely differs by region too, due to underlying economic differences and the distinct characteristics of regional innovation systems and universities (Cooke, 2001; Asheim and Coenen, 2005; Huggins and Johnston, 2009; Faggian et al., 2019). On the right-hand side of Figure 24 we report the share of graduates whose publications are being cited in patents by UK regions. Areas of high citation rates, marked in darker red, correspond to regions with established research clusters, such as those around Oxfordshire, Warwickshire, and Northamptonshire. By contrast, many parts of the UK, particularly in Northern Scotland, Wales, and Northern Ireland, exhibit lower rates of publication citations in patents, generally ranging from 0% to 4%. These findings highlight the uneven geographical distribution of STEMM-related innovation and suggest that innovation spillovers are more difficult to achieve by graduates' from less centrally positioned universities.

In addition, we found that majority of patents citing UK graduates’ publications are filed by applicants in the United States, over 70% of all patents citing UK science originate from US applicants, reflecting the global prominence of US organisations in translating UK academic knowledge into technological applications. European countries, foremost Germany and France, make up 9.2% of all patents citing UK scientific output, further highlighting the cross-border flow of knowledge and the role of UK science as a foundational source for cross-national innovation networks. UK applicants also feature prominently, representing around 6.6% of total patent filings citing UK PhD’s research, indicating that domestic innovators frequently seek protection at the USPTO and EPO to capitalise on their research. Together, these patterns demonstrate that, while some regions in the UK contribute more directly to innovation locally, the influence of UK PhD research extends internationally, shaping technological developments across multiple major innovation systems.

4. Conclusions

The evidence presented in this report highlights the substantial and growing contributions made by UK PhD graduates to both scientific knowledge production and downstream innovation. Across disciplines, doctoral candidates are becoming increasingly research-productive, with publishing now firmly embedded as a core component of the doctoral experience. This shift has important implications for doctoral training policy. While rising publication expectations may enhance research visibility and career prospects, they also risk reinforcing disciplinary inequalities, given persistent differences in publishing norms and output structures between STEMM and non-STEMM fields. Policy interventions that expand access to publishing support and funding could help ensure that heightened expectations do not disproportionately disadvantage groups of students. Our results also highlight the potential of UK PhDs to contribute to innovation. The technological benefits of their research are currently realised elsewhere and it is thus important to continue strengthening institutional and policy frameworks that support pathways from doctoral research to application, including knowledge exchange offices, translational funding, and structured engagement with local industry.

Our evidence also points to unequal participation in science and innovation pathways. Gender disparities persist, not so much in the likelihood of publishing, but in the volume of publications and in the likelihood that research informs patented inventions. This suggests that policy efforts should move beyond participation targets and address differences in access to resources, networks, and recognition that shape research intensity and visibility during and after the PhD. Importantly, we see that women in STEMM produce research that has the same potential for commercialisation, and there is thus substantial value in promoting pathways towards commercialisation for this group.

The uneven regional distribution of research participation and outcomes identified in this study suggests a need for place-based investment in research capacity and doctoral training infrastructure. PhD participation, publishing propensity, and innovation-related outcomes are unevenly distributed across the UK, with London, the East and South-East of England, and Scotland consistently outperforming other regions, while Northern Ireland and Wales lag behind on several indicators. These differences reflect broader inequalities in research infrastructure, institutional capacity, and proximity to innovation-intensive industries. Addressing such imbalances will require targeted, place-based policy interventions. These may include sustained investment in doctoral training and research infrastructure in under-represented regions, support for inter-regional doctoral networks, and incentives for collaboration between institutions with differing resource endowments. Strengthening regional research ecosystems could not only improve equity but also enhance the UK's overall innovation capacity.

While UK doctoral training produces knowledge of high international value, the domestic capture of innovation benefits appears limited. This does not necessarily indicate policy failure, but it does suggest scope for better alignment between doctoral training, national innovation strategies, and industrial policy. Enhancing opportunities for PhD graduates to engage with UK-based firms, supporting mobility between academia and industry, and embedding doctoral training more explicitly within regional and sectoral innovation strategies may help retain a greater share of downstream benefits within the UK.

Overall, UK PhD programmes should focus not only on increasing numbers or outputs, but also on improving the equity, inclusiveness, and societal embeddedness of doctoral training. Doing so would help ensure that the expanding contributions of PhD graduates translate into broad-based scientific, economic, and regional benefits.

Now that you have read our report, we would love to know if our research has provided you with new insights, improved your processes, or inspired innovative solutions.

Please let us know how our research is making a difference by completing our short feedback form via this link.

You are also welcome to email us if you have any questions about this report or the work of the IRC generally: info@ircaucus.ac.uk

Thank you

The Innovation & Research Caucus

References

Ahmadpoor, M., & Jones, B. F. (2017). The dual frontier: Patented inventions and prior scientific advance. Science, 357(6351), 583-587.

Aksnes, D. W., Nygaard, L. P., & Reiling, R. B. (2025). A matter of time? How absence from work affects gender gaps in research productivity. Higher Education, 1-16.

Aksnes, D. W., & Sivertsen, G. (2004). The effect of highly cited papers on national citation indicators. Scientometrics, 59(2), 213-224.

Asheim, B. T., & Coenen, L. (2005). Knowledge bases and regional innovation systems: Comparing Nordic clusters. Research Policy, 34(8), 1173-1190.

Bentley, P. (2012). Gender differences and factors affecting publication productivity among Australian university academics. Journal of Sociology, 48(1), 85-103.

Bogle, I. (2018, September). 100 Years of the PhD in the UK. Careers Research and Advisory Centre (CRAC) Limited.

Buenstorf, G. & Heinisch, D.P. (2020). When do firms get ideas from hiring PhDs? Research Policy, Volume 49, Issue 3, 103913, ISSN 0048-7333

Cooke, P. (2001). Regional innovation systems, clusters, and the knowledge economy. Industrial and corporate change, 10(4), 945-974.

Corsini, A., Pezzoni, M. & Visentin, F. (2022). What makes a productive Ph.D. student? Research Policy, 51(10), 104561.

Corsini, A., Koenig, J., Özgun, B., Romanyuk, A., Buenstorf, G., Lissoni, F., Miguelez, E., Pezzoni, M. & Martinez, C., (2025) Research careers in Europe: New evidence from the Doc-Track database. NBER Conference https://conference.nber.org/conf_papers/f230121.pdf

Ding, W. W., Murray, F., & Stuart, T. E. (2006). Gender differences in patenting in the academic life sciences. science, 313(5787), 665-667.

Faggian, A., Modrego, F., & McCann, P. (2019). Human capital and regional development. Handbook of regional growth and development theories, 149-171.

Gao, J., Yin, Y., Myers, K. R., Lakhani, K. R., & Wang, D. (2021). Potentially long-lasting effects of the pandemic on scientists. Nature communications, 12(1), 6188.

Grove, J. (2024, January 10). PhDs: Is doctoral education in trouble in the UK? Times Higher Education. https://www.timeshighereducation.com/depth/phds-doctoral-education-trouble-uk

Groen-Xu, M., Bös, G., Teixeira, P. A., Voigt, T., & Knapp, B. (2023). Short-term incentives of research evaluations: Evidence from the UK Research Excellence Framework. Research Policy, 52(6), 104729.

Grove, J. (2025, February 12) Where do UK PhD graduates go? Careers observatory may hold answer. Times Higher Education. https://www.timeshighereducation.com/news/where-do-uk-phd-graduates-go-careers-observatory-may-hold-answer

Halse, C., & Mowbray, S. (2011). The impact of the doctorate. Studies in higher education, 36(5), 513-525.

Hancock, S. (2021). What is known about doctoral employment? Reflections from a UK study and directions for future research. Journal of Higher Education Policy and Management, 43(5), 520-536.

Hancock, S. (2023). Knowledge or science-based economy? The employment of UK PhD graduates in research roles beyond academia. Studies in Higher Education, 48(10), 1523-1537.

Horta, H., & Santos, J. M. (2016). The impact of publishing during PhD studies on career research publication, visibility, and collaborations. Research in Higher Education, 57(1), 28-50.

Huggins, R., & Johnston, A. (2009). The economic and innovation contribution of universities: a regional perspective. Environment and Planning C: Government and Policy, 27(6), 1088-1106.

Hughes, T., Webber, D., & O'Regan, N. (2019). Achieving wider impact in business and management: Analysing the case studies from REF 2014. Studies in Higher Education, 44(4), 628-642.

Hunt, J., Garant, J. P., Herman, H., & Munroe, D. J. (2012). Why don't women patent? (No. w17888). National Bureau of Economic Research.

Khan, A., Sina Önder, A., & Ozcan, S. (2026). Performance-based research funding and gender diversity in research: evidence from UK universities. Oxford Economic Papers, 78(1), 1-18.

Kim, J., & Lee, S. (2015). Patent databases for innovation studies: A comparative analysis of USPTO, EPO, JPO and KIPO. Technological Forecasting and Social Change, 92, 332-345.

Kwon, D. (2025). How many PhDs does the world need? Doctoral graduates vastly outnumber jobs in academia. Nature, 643, 3.

Larivière, V. (2012). On the shoulders of students? The contribution of PhD students to the advancement of knowledge. Scientometrics, 90(2), 463-481.

Lax-Martínez, G., de Juano-i-Ribes, H.S., Yin, D., Feuvre, B.L., Hamdan-Livramento, I., Saito, K. & Raffo, J. (2021). Expanding the World Gender-Name Dictionary: WGND 2.0.

Lax Martínez, G., Raffo, J., & Saito, K. (2016). Identifying the gender of PCT inventors (WIPO Economic Research Working Paper No. 33). World Intellectual Property Organization. https://www.wipo.int/publications/en/details.jsp?id=4125&plang=EN

Marcella, R., Lockerbie, H., & Bloice, L. (2016). Beyond REF 2014: The impact of impact assessment on the future of information research. Journal of Information Science, 42(3), 369-385.

Marti, S. and Peneoasu, A.M. (2025) Doctoral education in Europe today: enhanced structures and practices for the European knowledge society. 2025 Survey report I. European University Association, Council for Doctoral Education.

Marx, M., & Fuegi, A. (2020a). Reliance on science: Worldwide front-page patent citations to scientific articles. Strategic Management Journal, 41(9), 1572-1594.

Marx, M., & Fuegi, A. (2022b). Reliance on science by inventors: Hybrid extraction of in-text patent-to-article citations. Journal of Economics & Management Strategy, 31(2), 369-392.

Muric, G., Lerman, K., & Ferrara, E. (2021). Gender disparity in the authorship of biomedical research publications during the COVID-19 pandemic: retrospective observational study. Journal of medical Internet research, 23(4), e25379.

Official Journal of the European Union – OJEU (2023) Council Recommendation of 18 December 2023 on a European framework to attract and retain research, innovation and entrepreneurial talents in Europe (C/2023/1640).

Park, C. (2005). New variant PhD: The changing nature of the doctorate in the UK. Journal of higher education policy and management, 27(2), 189-207.

UK Research and Innovation - UKRI (2024, November 13). Major investment to support the next generation of researchers. UKRI. https://www.ukri.org/news/major-investment-to-support-the-next-generation-of-researchers/

Annex A: UK DGCI Methodology

Introduction

This Annex provides a summary of the data sources employed and key methodological steps of this project. The methodology follows the DOC-TRACK project (https://doc-track.eu/; Corsini et al, 2025), a European-wide study of the inventive activities of STEMM doctoral graduates in Europe funded by the European Patent Office Academic Research Programme (EPO-ARP).

EThOS repository and metadata

EThOS, managed by the British Library, is a bibliographic database of electronic theses with metadata that can be searched and downloaded as an open dataset (last available version November 2023; https://doi.org/10.23636/rcm4-zk44). The dataset covers approximately 98% of all PhDs awarded by UK Higher Education institutions since 1787 and is available under a Creative Commons license.

Table A1 summarises the available bibliographic metadata, which covers name of the graduate, awarding institution, year of publication, discipline, title, subject and abstract (for 72% of entries) and thus provides a comprehensive and structured dataset suitable for large-scale quantitative analysis of doctoral education in the UK.

Supervisor information is available but shows gaps, particularly in older records, with less than 10% of pre-2010 records and less than 50% of post-2009 records containing supervisor details. We systematically retrieved missing supervisor data from UK universities' library portals, employing a three-step methodology focussing on the period since 2010, as pre-2010 information was largely unavailable or inaccessible through existing URLs. Our approach combined automated web extraction, PDF acknowledgement extraction via name entity recognition, and targeted manual checks. The procedures identified supervisors for 72,454 dissertations and achieved a validated accuracy of 95.44% in confirming at least one supervisor for theses from 2010 onwards. Across all three steps, i.e., web-crawling, acknowledgement extraction, and manual checking, we recovered 64.3% of missing data for the period of 2010-2020, reducing the supervisor missing rate in that period to 18.7%.

Subject information was missing for a significant share of thesis in the most recent cohorts. We employed a two-stage approach to assign these dissertations to one of the 19 EThOS subject disciplines: 1) via linked publication data, and 2) machine learning-based classification model.

Among the cases with missing subject data, 58% could be linked to OpenAlex publications. We assigned the subject from a publication published closest to the thesis publication year. This method was based on the assumption that a researcher's publication topics would be most similar to their doctoral research during that period. For the remaining theses, we employed a machine learning approach, achieving ~72% accuracy.

Table A1: Information coverage in EThOS Metadata (in %)

		2000-2020	2020	2010	2000
Thesis	Title	100	100	100	100
	Abstract	72.8	96.8	60.5	36.4
	Discipline	98.4	87.0	99.9	99.9
	Institution	100	100	100	100
	Year	100	100	100	100
	Link to local repository	70.2	99.1	61.0	37.8
Graduate	Full Name	93.2	97.3	90.7	94.0
	Surname + Initial	100	100	100	100
Supervisor	Full Name	31.0	56.9	23.8	2.9
	Surname + Initial	33.0	67.4	24.9	4.9
Thesis*	Discipline (after processing)	100	100	100	100
Supervisor^	Name (after processing)	53.9	93.2	53.0	4.9

* predicted using publication matching and machine learning-based classification.

^ extracted via webscraping, PDF acknowledgement extraction, and manual search

Gender attribution

We applied a multi-stage process to assign gender attributions to UK PhD graduates, aiming to maximise accuracy and coverage across culturally diverse naming patterns. We primary resource was the World Gender-Name Dictionary (WGND 2.0) (Lax-Martinez et al., 2016, 2021), selected for its extensive international scope. The WGND classifies names into three categories: male ("M"), female ("F"), and uncertain ("?"). Pre-processing steps included checking for prediction consistency within the dictionary, handling conflicting predictions and prediction uncertainty.

After cleaning and consolidating the WGND dictionary, we matched this to graduates' first names, obtaining an initial gender attribution for 94.14% of PhD students; however, 13.04% of these matches came from names with conflicting predictions. We conducted additional cleaning to refine the matches, removing cases where gender was inferred only from initials and resolving uncommon or ambiguous names. To improve reliability, we applied a minimum prediction-weight threshold of 0.8. After this full procedure, we achieved a final gender attribution rate of 86.1% (299,649 out of 347,838 graduates), or 91.4% for theses with full first names.

Linking UK PhD theses to scientific publications

Our methodology for linking UK PhD theses to their corresponding scientific publications is based on DOC-TRACK's supervised machine learning classification approach (Corsini et al. 2025). The method can be divided into several key phases: parsing, merging, model training, and classification (Figure A1).

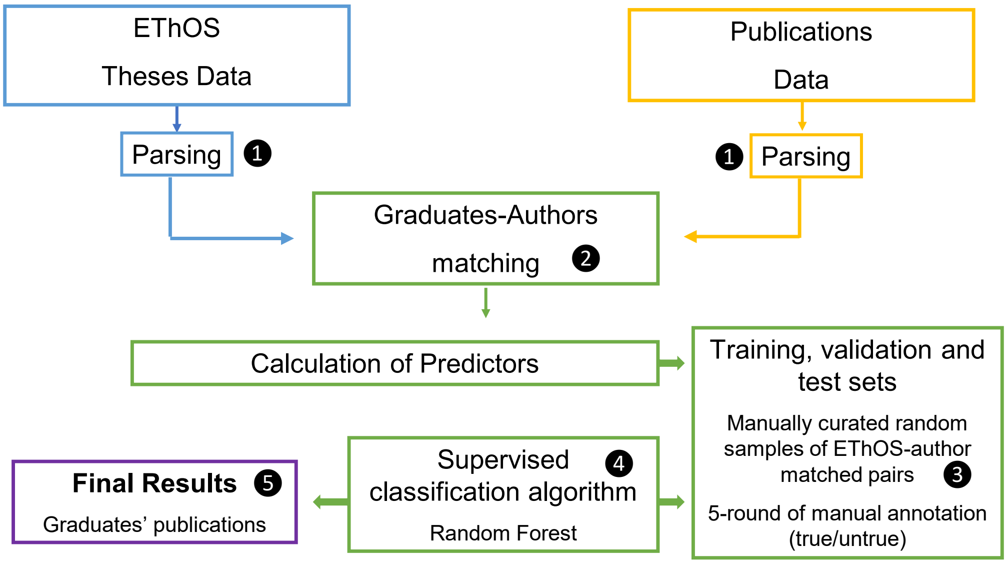
EThOS
Theses Data
Parsing ❶
Publications
Data
❶ Parsing
Graduates-Authors
matching ❷
Calculation of Predictors
Training, validation and test sets
Manually curated random samples of EThOS-author matched pairs ❸
5-round of manual annotation (true/untrue)
Supervised classification algorithm ❹
Random Forest
Final Results ❺
Graduates' publications

Figure A1: Theses-publication matching methodology (adapted from Corsini et al., 2025)

Step 1: data parsing and standardisation

To ensure uniformity across datasets, particularly for author names, which are central to matching, we standardised all thesis and publication author names by converting them to lowercase, removing non-alphabetic characters, and stripping titles and honorifics. To account for variations in name entry, we generated several variants for each individual, including forms with initials and alternative given-family name permutations. Publication authors underwent the same cleaning and variant-generation process, and entries lacking a first or last name were removed. Each record was then normalised into a unified relational schema.

Step 2: initial merging and filtering

To identify potential matches, we conduct a many-to-many exact match using the standardised and expanded name variants from both datasets. This approach achieves high recall, linking 96% of UK PhD graduates to at least one publication author with a similar name. To manage the large number of potential matches and reduce the number of false positives, we apply a rule-based filtering system, which includes a temporal filter to only retain publications within a plausible time window relative to the PhD graduation year (at least one publication in the period from 5 years before and 3 years after the PhD), an outlier filter to remove cases where a single graduate is linked to an unusually high number of authors, and a name consistency filter to resolve conflicts between full names and initials. Finally, we apply a geographic filter to only retain matches where the publication author has institutional ties to the UK, arriving at a set of 1.5M possible Scopus author matches.

Step 3: construction of the gold standard dataset

To train and validate our machine learning model, we created a gold standard dataset of manually annotated thesis–author pairs. A stratified random sample of 19,482 matched pairs (based on 3000 theses authors) was drawn and evaluated by two independent annotators, who classified each pair as a true match, false match, or uncertain. Disagreements were resolved by a third annotator or reviewed manually to ensure consistency. This rigorous annotation process provided high-quality labelled data for supervised learning with an error of less than 1%.

Step 4: model training and evaluation

To distinguish between correct and incorrect linkages, we employ a Random Forest classification algorithm. This model is well-suited for our task because it is robust to class imbalance, handles high-dimensional feature spaces effectively, and is less prone to overfitting. We train the model on our annotated gold standard dataset using 10 features, incl. the presence of the supervisor amongst co-authors, geographic alignment between author affiliation and thesis institution, title similarity and discipline overlap, author name similarity, and publication productivity during the PhD period. We use 90% of the annotated gold standard sample for training and validation, implementing a stratified 5-fold cross-validation procedure combined with an exhaustive grid search to fine-tune the model's hyperparameters. This is tested against the remaining 10%. The best-performing model, determined by the highest F1 score, is then selected for the final classification task. Our model achieved a strong performance, demonstrating high precision and recall on the test set (96.7% precision, 92.3% recall).

Step 5: Final classification and sample construction

The selected model was applied to the full set of possible thesis-author pairs. Finally, our sample consists of 213,825 thesis authors with highly probable author matches in Scopus, corresponding to 62% of graduates. All identified publications' DOIs are linked to OpenAlex, an open database of scientific publications for further processing, dataset building and analysis (99% match rate).

Linking scientific publications to patents

Our methodology for linking UK PhD theses to their corresponding scientific publications that are subsequently cited by patents relies primarily on the Reliance on Science database (Marx and Fuegi, 2020a, 2020b).[9] This database provides extensive publication identifiers, which we used to establish matches with OpenAlex records. Pre-processing steps included standardising publication identifiers for both Reliance on Science and OpenAlex. To track patent identifiers, we considered both USPTO and EPO patent documents, retaining only patent identifiers with the us- and ep- suffixes to ensure accurate jurisdictional mapping. After these steps, we found that 5.17% of the identified unique publications were connected to patents, which corresponds to 17% of our PhD graduates having at least one publication cited by a patent. When focusing specifically on STEMM theses, the share of graduates with at least one publication cited by a patent was 34.5%. Among all citing patents, we further link UK applicants to ORBIS-IP to identify location of patent applicant.

[9] Accessed at https://relianceonscience.org/ and downloaded on July 31st, 2025

Annex B: Complementary analyses by UK-DGCI

The AI Thesis Network: A glimpse into gender dynamics of UK STEMM PhD theses.

MIOIR Blog as part of the UK Doctoral Graduates' Contribution to Innovation project, co-authored by Sanya Panda and Emily Johnson, research interns working with the UK DGCI team. Published on 19 Sep 2025 https://blogs.manchester.ac.uk/mioir/2025/09/19/the-ai-thesis-network-a-glimpse-into-gender-dynamics-of-uk-stemm-phd-theses/

Findings: The findings in this blog suggest that the underrepresentation of women in AI is not merely a reflection of broader STEMM participation rates, but points to structural barriers within the AI research ecosystem itself. One such barrier relates to the significant gender imbalance among PhD supervisors in AI, where male supervisors vastly outnumber female counterparts. Our data reveals that 73.4% of AI-relevant dissertations were supervised by men while only 26.6% had female supervisors. This could potentially influence mentorship opportunities, research topic selection (e.g. steering women away from AI), and access to networks. Addressing these disparities requires interdisciplinary strategies such as promoting diverse inclusive mentorship models, opening up opportunities for AI research in STEMM fields with stronger female representation, and ensuring institutional support for equitable opportunities.

Funding for PhDs: Shaping the Future of Science and Innovation.

MIOIR Blog as part of the UK Doctoral Graduates' Contribution to Innovation project, authored by Emily Johnson, research intern working with the UK DGCI team.
Published on 15 Dec 2025 https://blogs.manchester.ac.uk/mioir/2025/12/15/funding-for-phds-shaping-the-future-of-science-and-innovation/

Findings: The analysis shows that doctoral funding in the UK is unevenly distributed across subjects, genders, and regions, and that these domains are deeply interconnected. STEM disciplines, particularly Physics and Technology, consistently attract the highest levels of support, while Humanities and Social Sciences remain underfunded. This disciplinary weighting intersects with gender, since STEM is historically male-dominated, with men accounting for around two-thirds of funded PhDs. At the same time, STEM funding is concentrated in the 'Golden Triangle' of London, Oxford, and Cambridge, reinforcing regional disparities. These disparities matter because they extend beyond doctoral study into the wider research workforce.

www.ircaucus.ac.uk

Email info@ircaucus.ac.uk Twitter @IRCaucus

INNOVATION &
RESEARCH
CAUCUS

Delivered with
ESRC and
Innovate UK