Knowledge Spillovers in Health Care: The Geographic Impact of Biomedical Research on Cancer Outcomes*

Katharina Blankart† Johannes S. Kunz‡ Maria Wiśniewska§

June 2026

Preliminary draft

Abstract

Innovation plays a crucial role in health care as it drives continuous improvements in patient outcomes. However, the influence of new ideas largely depends on how quickly they are adopted within professional networks. While the influence of prominent scientists is well-documented, the broader effects of general research activity remain less understood. This study uses the total volume of local biomedical research in cancer as a novel measure of innovation and examines its effect on health outcomes. To assess this relationship we link over 32,000 relevant publications, by their geographic area and studied cancer type, to cancer-related death records from Australian administrative data. Our fixed-effects analysis shows that residing in a research-active area is associated with higher mean death age and lower cancer mortality. To test causality, we employ a shift-share instrumental variable approach that uses exogenous variation from the establishment of new medical schools. The causal estimates confirm that increased local research activity significantly improves health outcomes.
JEL-codes: I11, I12, J24, O31
Keywords: Knowledge spillovers; Biomedical research; Cancer outcomes; Innovation diffusion; Geographic spillovers; Shift-share IV.

Acknowledgments: We thank Rebecca McKibbin, Bianca Devitt and seminar participants at Monash University for very useful comments. Funding by the DAAD is gratefully acknowledged. The results of these studies are based, in part, on data supplied to the ABS under the Taxation Administration Act 1953, A New Tax System (Australian Business Number) Act 1999, Australian Border Force Act 2015, Social Security (Administration) Act 1999, A New Tax System (Family Assistance) (Administration) Act 1999, Paid Parental Leave Act 2010 and/or the Student Assistance Act 1973. Such data may only be used for the purpose of administering the Census and Statistics Act 1905 or the performance of functions of the ABS as set out in section 6 of the Australian Bureau of Statistics Act 1975. No individual information collected under the Census and Statistics Act 1905 is provided back to custodians for administrative or regulatory purposes. Any discussion of data limitations or weaknesses is in the context of using the data for statistical purposes and is not related to the ability of the data to support the Australian Taxation Office, Australian Business Register, Department of Social Services and/or Department of Home Affairs' core operational requirements.

Legislative requirements to ensure privacy and secrecy of these data have been followed. For access to MADIP and/or BLADE data under Section 16A of the ABS Act 1975 or enabled by section 15 of the Census and Statistics (Information Release and Access) Determination 2018, source data are de-identified and so data about specific individuals has not been viewed in conducting this analysis. In accordance with the Census and Statistics Act 1905, results have been treated where necessary to ensure that they are not likely to enable identification of a particular person or organisation.
Ethics: Monash, MUHREC-37998.
CoI: None of the authors has any conflict of interest to declare.

†School of Health Professions. Bern University of Applied Sciences, katharina.blankart@bfh.ch.
‡Centre for Health Economics, Monash Business School, Monash University. johannes.kunz@monash.edu
§Department of Economics, Deakin University. maria.wisniewska@deakin.edu.au.

1 Introduction

Local networks play a crucial role in the diffusion of information in research, often driven by scientific 'superstars' (Azoulay, Graff Zivin and Sampat, 2011). In the medical field, such local opinion leaders are especially influential, conducting research on transformative innovations and accelerating their adoption within local communities (Agha and Molitor, 2018; Vadia and Blankart, 2021). This increased adoption can, in turn, contribute to improved patient health outcomes (Skinner and Staiger, 2015). However, with the increasingly important role of specialisation in modern treatments and the exponentially growing body of medical knowledge,[1] the adoption of 'mainstream' innovations, such as new pharmaceuticals, may not fully account for the variation in health outcomes. This is especially true in disease groups such as cancer, where treatment may differ significantly depending on the individual characteristics of the patient, and is closely related to the physician's exposure to similar cases in their practice (Selby, Gillis and Haward, 1996), and to other practicing peers (Barrenho et al., 2025).

In this study, we evaluate whether the broader community of local medical researchers - including those who are less prominent but actively publish in medical journals and contribute to the understanding of specific diseases - plays a role in improving health outcomes. We take a novel approach to measuring their local scientific activity by using PubMed, a comprehensive repository of the biomedical literature. We build a database of over 32,000 cancer-related articles published between 2000 and 2016 in Australia[2], with over 6% of the sample classified as clinical trials, each linked to the geographic location of its authors and detailed metadata, including the studied cancer type. We hypothesise that areas with a higher volume of publications focused on a specific cancer type are likely to have local experts who specialise and innovate in treatments for that cancer. In turn, this specialised knowledge might be disseminated through local networks, ultimately leading to adoption of the innovation and improved health outcomes for the local patient population.

We also use Australian, population-wide, administrative records from Person Level Integrated Data Asset (PLIDA), specifically each individual's location, demographic characteristics (age and gender), chemotherapy usage, and death records, including the information on the main cause of death and, in turn, enabling the identification of all cancer-related fatalities. We construct several mortality-related variables to study the relationship between scientific activity and health outcomes in cancer on area level: mean age at death, number of cancer-related deaths per 10,000 population and number of premature cancer-related deaths per 10,000 population.

Our identification is based on the variation of both scientific research and health outcomes not only across areas, but also across cancer types. For instance, an area may be considered scientific in the

[1] Medical knowledge is estimated to double every 73 days (Densen, 2011)

[2] Specifically, we concentrate on cancer-related research, as cancer is the leading cause of mortality in Australia and presents significant challenges to the healthcare sector: `https://vizhub.healthdata.org/gbd-compare/`, data for 2019, accessed on 14.03.2025

treatment of breast cancer but nonscientific in the treatment of lung cancer, depending on the nature of research publications produced there. As a result, physicians in different regions are exposed to experts in distinct fields, allowing us to assess the impact of localized scientific activity on medical practices. We use fixed-effects models to account for condition- and area-specific effects on the outcomes of interest, controlling for various factors that may influence the adoption of new medical technologies beyond exposure to influential scientists - such as local demand and need.

Despite accounting for fixed effects in our models, endogeneity remains a concern as the distribution of publications across areas is not random. To mitigate this, we employ a shift-share instrumental variable (SSIV), exploiting the exogenous shock created by the sudden establishment of new medical schools in the early 2000s, aimed at boosting the supply of health care practitioners. This, in turn, expanded the population of publishing medical scientists in an unprecedented way over the subsequent years. We apply the instrument to our models analysing probability of dying from cancer.

Our results show that patients living in scientific areas, on average, die at an older age than people living in non-scientific areas across different model specifications. First, we show a positive relationship between the absolute number of cancer-specific articles in an area and the mean age at death. Second, we show the inverse hyperbolic sine of the number of articles (to control for highly skewed article distribution) is positively correlated with older mean age at death in scientific areas.

Our findings also indicate that residing in scientific areas positively impacts overall cancer mortality. Specifically, each additional publication is associated with a reduction of 0.023 deaths per 10,000 people and a decrease of 0.012 premature deaths per 10,000 people. These results are reinforced by the SSIV analysis.

Next, we conduct a heterogeneity analysis, deriving the effects of scientific activity in cancer on males and females, and across more and less disadvantages areas. We build on the literature regarding inequities in health care, which shows that some groups, such as females and people in disadvantaged areas, may face discrimination in clinical practice - an issue particularly important regarding the publicly-funded nature of most research. We find that both males and females benefit from scientific activity, though males experience a greater advantage. Furthermore, when comparing regions above and below the median Index of Relative Socio-Economic Disadvantage (IRSD), we find that individuals in more disadvantaged areas gain greater benefits from scientific activity than those in less disadvantaged areas, though both groups experience significant positive effects.

Lastly, we exploit the variation in local expertise to study the impact of scientific activity pre- and post-months of listing new cancer medications on the Pharmaceutical Benefits Scheme (PBS). We assume new listings to be exogenous, since all specialists can prescribe the same medications to patients located in any area. We use these identified differences in expert composition on area level to study the effects of new PBS listings: both on the uptake of these medications and on related health outcomes. Subsequently, we

show that prescribing of newly-listed chemotherapy medications follows similar trends for patients living in both scientific and non-scientific areas, with high uptake from the first quarter after the introduction.

Our study contributes to the literature focused on the diffusion of information in healthcare. We build on the well-understood mechanism of how biomedical research might impact health outcomes and on the role of knowledge spillovers in the adoption of new ideas. We assume that biomedical research, conducted by local experts, leads to important developments in healthcare, such as new medications, medical procedures, or devices (i.e., through spillovers from academia to industry (Azoulay, Graff Zivin and Sampat, 2011)) and that these developments contribute to the improvement of health outcomes (Lichtenberg, 2013). Subsequently, we hypothesize that expert knowledge spreads across professionals, potentially leading to improved health outcomes due to the adoption of new solutions. The literature points to various mechanisms of diffusion of innovation. Jaffe, Trajtenberg and Henderson (1993) show evidence that knowledge spreads geographically, with significant primary local invention uptake, and that geographical uptake variation fades over time. Coleman, Katz and Menzel (1957) show that innovation spillovers take place, to some extent, due to professionals' ties and networks.

Second, our study extends the recent literature, by assessing the direct impact of scientific activity on health outcomes. There are two studies that explore similar mechanisms of knowledge spillovers and their impact on health outcomes. The first is Agha and Molitor (2018), in which authors explore the impact of local experts, specifically physician investigators leading clinical trials for new cancer drugs, on the adoption of technology in healthcare. The study finds that patients in the lead investigator's region are initially 36% more likely to receive the new drug, indicating the significant influence of these local leaders. The second study is McKibbin and Weinberg (2021), which addresses the challenge of causally estimating the impact of biomedical research on mortality in the US by linking a research database to administrative death records. Utilising hospital market-level shocks to research activity by disease, the study finds that an additional research publication, on average, reduces local mortality from disease by 0.35%. Our study contributes to this literature by analysing the heterogeneous effects of research activity and measuring the impact on outcomes beyond cancer-related mortality.

Finally, we provide valuable insights for the decision-makers regarding the justification of new investments in cancer-related R&D projects and supporting local research activity. We also present results that might inform policy about existing inequalities in gaining health benefits due to the uneven distribution of expert knowledge and its spillovers in the country.

2 Background

2.1 Measurements of innovation and scientific activity in economics

Selecting an appropriate measure of innovation presents a challenge in our study. In the economics literature, innovation is commonly assessed through measures such as patents (Hall, Jaffe and Trajtenberg, 2005) or R&D expenditures (Jaffe, 1986). These, however, may not be optimal when aiming to study the impact of local innovations in the medical field (usually created by practicing physicians).

R&D resources have some potential limitations regarding the interpretation of the data, since they do not capture the entirety of innovation, only the extent of funds that went towards the development (i.e., the resources are the input in creating innovation, not the output). They also often provide the overall expenditure, instead of product-specific investments, prohibiting from the detailed analysis in particular fields. Patent statistics have limitations regarding the coverage, as many innovations do not have a patent and there's a large variation in the actual value of existing patents (OECD and Eurostat, 2005).

The measure of innovation employed in this study - cancer research publications - addresses several limitations associated with alternative measures such as R&D resources and patent data. Scientific publications can effectively capture innovation, as they represent the initial formal record of specific knowledge within scholarly literature. Unlike patents or R&D expenditure, publications encompass a broader range of scientific activities, including research conducted both with and without dedicated funding. Focusing on publications also facilitates the identification of local experts within narrowly defined fields (specific cancer types), as expertise can be quantified by publication volume; areas with a higher number of publications are considered to have greater expertise in that cancer type. Finally, in a highly specialised and experimental field such as medicine, research that contributes to broader scientific understanding, beyond merely developing new treatments, is equally valuable in advancing disease knowledge.

2.2 Burden of cancer and access to innovation in Australia

Oncology is one of the most crucial areas in need of innovation, given the burden of disease on society and predicted increase in incidence.[3] While innovation in cancer treatments may take different forms - from medication development to analysing rare cases - it is crucial to understand how that expertise spreads across the country and whether it benefits the society in an equal way, especially as government funding plays a pivotal role in advancing research initiatives.[4]

The increasing incidence of cancers in Australia, coupled with the high cost of treatments, has led to

[3]It is estimated that 3 of every 10 deaths in 2023 is due to cancer, `https://www.aihw.gov.au/reports/cancer/cancer-data-in-australia/contents/about`, Accessed on 14.11.2023

[4]The 2020-2021 budget of the Australian Government for medical research is 6.7 bln dollars, part of which is dedicated to rare cancer research, `https://www.health.gov.au/sites/default/files/documents/2021/05/life-saving-research-funding-for-medical-research.pdf`, Accessed on 09.11.2023

a significant financial burden on the healthcare system. In 2015-2016, cancer accounted for a substantial portion of healthcare expenditures, comprising approximately 8.6% of the total disease expenditure[5]. Furthermore, it is a major contributor to mortality, with cancer-related deaths representing roughly 31% of all fatalities in 2019[6].

The projected number of deaths and number of new cases will increase in the next 2 decades in Australia, with a total number of cases increasing by 47.4% and the number of cancer deaths increasing by 36.4% in 2040-44 (Luo et al., 2022), showing the importance and timeliness of our context.

Australian settlement locations are dictated, to some extent, by the natural conditions that occur in the country (e.g., 80% of the country is uninhabitable). These unique circumstances pose challenges in bringing equitable healthcare services to Australian residents given both the variation in population density (ranging from 20,700 people per sq km in inner Melbourne to below 0.1 people per sq km in remote areas in 2018) [7] and significant distances between the regions. Studies also point to the variation in risky behaviours (such as alcohol consumption, excess sun exposure, and physical activity), variation in breast, cervical, and bowel screening programs attendance, and variation in incidence and survival between areas pointing to some potential inequalities and vulnerability of some communities ((Sharplin et al., 2014), (Cramb et al., 2017)).

These geographic disparities extend to cancer mortality. Even after restricting to areas with sufficient cancer-related deaths to ensure reliable estimation, there is a 21.5-year gap in mean age at death from cancer between the highest and lowest ranked areas over the period 2011-2019 - from 80.5 years in the area with the oldest average age at death to just 59 years in the area with the youngest.

Chemotherapy, alongside surgery and radiotherapy, holds a pivotal place in cancer care in modern healthcare systems. While pharmaceutical companies consistently develop new and innovative medications, the accessibility of new chemotherapy medications in Australia follows a rigorous, evidence-based acceptance process (Mellish et al., 2015). First, each new drug must secure approval from the Therapeutic Goods Administration (TGA) and be listed on the Australian Register of Therapeutic Goods (ARTG) to be sold and used in the country. Subsequently, medications are evaluated by the Pharmaceutical Benefits Advisory Committee (PBAC), including cost-effectiveness analysis, in order to be listed on the Pharmaceutical Benefits Schedule (PBS) and subsidised by the Australian government, what reduces the cost to patient to a maximum of AUD 31.60[8].

[5] Australian Institute of Health and Welfare (AIHW). Available from https://www.aihw.gov.au/reports/cancer/health-system-expenditure-cancer-other-neoplasms/summary. (Accessed 26.10.2023)

[6] Institute for Health Metrics and Evaluation (IHME). GBD Compare. Seattle, WA: IHME, University of Washington, 2015. Available from http://vizhub.healthdata.org/gbd-compare. (Accessed 26.10.2023)

[7] `https://www.abs.gov.au`, Accessed on 29.11.2023

[8] `https://www.pbs.gov.au/info/healthpro/explanatory-notes/front/fee`, accessed on 24.03.2025

3 Data

As the basis for our study, we select a subsample of all cancer types and focus on those in top 100 leading causes of death in Australia as of 2019, excluding cancer types with relatively low burden of disease and low scientific activity.[9] Notably, the number of cancer-related deaths within the selected 20 groups of interest comprises around 80% of all cancer deaths (see Figure 7), with tracheal, bronchus, and lung cancer, colon and rectum cancer, prostate cancer, and breast cancer being the main causes of death among all cancer-related deaths (see Figure 8).

Then, to allow for the appropriate linkages between PubMed and PLIDA, we create a crosswalk for all 20 cancer types of interest (See Table 8). As the two sources use different classifications to indicate diseases - PubMed uses MeSH terms and PLIDA uses ICD-10 codes - we exploit the hierarchical structure in both and aggregate the appropriate codes into groups corresponding to either the site of cancer (e.g. breast cancer) or to the histologic type (e.g. leukemia). For example, prostate cancer is coded with C61 in ICD-10 and corresponds to two separate codes in MeSH classification - D011471 (Prostatic Neoplasms) and D064129 (Prostatic Neoplasms, Castration-Resistant).

Lastly, we aggregate our outcomes on Population Health Area (PHA) level.[10] PHA classification is based on Statistical Areas Level 2 (SA2),[11] and includes either entire SA2s or clusters of SA2s - in more rural parts of the country. This allows to better control for the large variation in both number of articles and population between the regions in Australia.

3.1 Cancer-related deaths

We utilise population-wide administrative datasets from the Person Level Integrated Data Asset (PLIDA) to define the cancer-death population. We start by identifying individuals who died between 2011 and 2019, with cancer as the primary cause of death, and link these records to Census data. This linkage enables us to examine the demographic and geographic characteristics of both the general population and the cancer-death population across scientific and non-scientific areas, including their usual place of residence.

Next, we calculate the age at death for each individual who died from cancer, using the year of birth assigned to each person. We then aggregate the data to create two panel datasets: one by PHA-year and another by PHA-year-cancer-group, which show the mean age at death for each category. We then link the aggregated data with publication data, using either all cancer-related research or cancer-type-specific

[9] Institute for Health Metrics and Evaluation (IHME). GBD Compare. Seattle, WA: IHME, University of Washington, 2015. Available from http://vizhub.healthdata.org/gbd-compare. (Accessed 5/12/2022)

[10] Public Health Information Development Unit (PHIDU), Torrens University Australia. Population Health Areas: Correspondence Based on Statistical Areas Level 2 (online) 2016. Accessed 21 March 2023, `https://phidu.torrens.edu.au/help-and-information/help-guides-and-faq/geographical-structures/pha-list`.

[11] Areas with an average population of 10,000: `https://www.abs.gov.au/statistics/standards/australian-statistical-geography-standard-asgs-edition-3/jul2021-jun2026/main-structure-and-greater-capital-city-statistical-areas/statistical-area-level-2`, accessed on 07.04.2025

research. These datasets consist of 10,354 and 104,160 observations, respectively, and are unbalanced due to some areas not recording any cancer-related deaths during our period of interest.

Finally, we incorporate indicators for the post-treatment period, which include only deaths that occurred after all selected papers were published (i.e., after 2016). Additionally, we identify areas that are scientifically active in at least one of the 20 selected cancer groups to serve as an alternative control group.

The initial dataset includes 329,106 cancer-related death records. After linking the data with Census information and selecting cancer groups of interest, the final dataset comprises 231,368 cancer-related death records.

3.2 General population outcomes

We also use PLIDA to derive general population-level outcomes. First, we construct a cross-sectional dataset of all individuals residing in Australia in 2012, excluding those not covered by Medicare (e.g., temporary migrants without public health insurance). To avoid bias from relocation to treatment centers, we use each person's 2012 location as a baseline. We then link this data to their location at the PHA level, gender, year of birth and death records from 2013 to 2019. Individuals who were alive in 2012 and did not die within this period remain in the dataset with missing values in the death columns.

Next, we classify individuals who died from cancer as their primary cause of death into one of 20 cancer groups (using Table 8) and link their PHA of residence in 2012, prior to any deaths being recorded in our data, to local scientific activity in cancer research (measured by publication volume). This allows us to determine whether a person who died from a specific cancer lived in an area with research activity related to that cancer, another cancer, or no relevant scientific activity.

We construct a stacked dataset at the PHA-cancer group level, aggregating population data to capture both the total local population (constant across cancer groups) and the total number of deaths by cancer type. Using these variables, we calculate the number of cancer-related deaths per 10,000 population. Individuals who died from a specific type of cancer remain in the population counts for other cancer types, as they did not pass away from those cancers. This structure allows us to control for both area-specific and cancer-group-specific effects. Our final dataset includes 23,280 records, covering 1,164 areas across 20 cancer groups.

We apply the same approach to determine the number of premature cancer deaths, considering only individuals who passed away before the age of 75. We then generate separate datasets for males and females, maintaining the same structure. Our key outcomes of interest include the absolute number of cancer-related deaths and the number of cancer-related deaths per 10,000 population.

Lastly, we create a binary disadvantage indicator using the Index of Relative Socio-Economic Disadvantage (IRSD). We classify areas as more disadvantaged (below the median score) or less disadvantaged

(above the median score) and link this information to our final dataset.

3.3 Research activity

To measure local research activity, we use PubMed database. Based on the affiliations of researchers' contributions to published research articles in the respective cancer research, we assign them to areas of their affiliation in Australia between 2000 and 2016.

To select relevant records from MEDLINE/PubMed, we used the National Library of Medicine's (NLM) thesaurus classification of Medical Subject Headings (MeSH)[12]. The MeSH terms are NLM's in-house developed keywords for cataloging each publication depending on its subject matter. MeSH terms are assigned to each publication by a team of indexers who systematically scan the title, abstract, and publication. Accordingly, authors cannot select or manipulate the assignment of MeSH. The MeSH classification is provided by a hierarchical tree structure containing 16 main branches, including diseases, drugs, therapeutic equipment, and processes, and each branch contains hierarchical subbranches. In addition to its definition, each MeSH is described by a "DescriptorName" (i.e., "Prostatic Neoplasms") for one word/phrase description and is accompanied by "DescriptorUI" (i.e., D011471), a unique ID number for that particular MeSH.

Per publication, multiple MeSH terms may be assigned with an average of 12 MeSH per article. We relied on the 2020 version of MeSH[13] that includes over 29,000 terms (Greenblatt, N.d.). We included all MeSH terms indexed in the group *C04* (neoplasms), in total 998 descriptors identified by 'DescriptorUI' and, subsequently, selected a subsample of all cancer code that correspond to the ones in top 100 causes of death in Australia.

To obtain the geographical distribution of affiliation of authors by cancer-related MeSH term, we used the publicly available tool MapAffil to assign geocodes (latitudes and longitudes) and the country of origin to each publication based on the authors' affiliations (Torvik (N.d.), Xu et al. (N.d.)). MapAffil is capable of correctly identifying 97.7% of the geolocation of a city reported in the author's affiliation variable provided by PubMed. We relied on the 2018 version of MapAffil.[14]

3.4 Chemotherapy dispensing

To test the potential channel of knowledge spillovers, we analyse all dispensed cancer medications listed on the PBS between 2013 and 2017 (see Table 9) and construct a measure of quarterly dispensed volumes for each PBS item code. To account for variations in introduction timing, we first identify the month in which a medication was either added to the PBS or had an update to its prescribing restrictions and calculate the dispensing volume for the quarter beginning in that month. We then extend this analysis to

[12] https://www.nlm.nih.gov/databases/download/mesh.html
[13] https://nlmpubs.nlm.nih.gov/projects/mesh/2020/
[14] https://databank.illinois.edu/datasets/IDB-2556310

include dispensing volumes for the five quarters preceding and following the introduction quarter. This approach standardizes the timing of medication introductions, allowing for a consistent baseline period and facilitating the analysis of changes in dispensing patterns over time.

Using PBS data for analysing cancer provides relatively high coverage of the cancer population, however, it can differ between types of cancer (from 57% of all chemotherapy captured in PBS data for head and neck cancers to 92% for neurological cancers) (Tervonen et al., 2020). Importantly, the introduction of the Efficient Funding of Chemotherapy in April 2012 improved the coverage as compared to the previous years.

4 Empirical strategy

Our empirical strategy uses highly detailed data that enables the geographic and cancer-group-level linkage of both scientific publications and mortality records. We include all articles published between 2000 and 2016 that are assigned relevant MeSH codes indicating cancer within our 20 groups of interest, alongside the death records.

We adopt a broad inclusion criteria for publications (i.e., all publications between 2000 and 2016), based on two key assumptions. First, scientific research may have a long-lasting impact beyond its publication date, influencing clinical practice and treatment outcomes over time. Second, research findings may begin to have an impact before formal journal publication, as knowledge is often disseminated within scientific and medical communities through conferences and professional networks.

We assume that, due to using PubMed, our selected sample of publications arguably covers the majority of all medically-relevant cancer-related research that was produced in Australia. Characteristics of these publications show that, overall, it contains research of relatively high quality - over 70% of publications are research articles, over 40% were cited in other clinical papers and over 11% perform higher than 75% of their peers (classification derived from (Klopfenstein and Dampier, 2020)). Furthermore, the number of retracted articles accounts for less than 0.03%.

However, as the scientists are not randomly assigned to the geographic areas or cancer groups, comparing the outcomes in scientific and non-scientific areas, without differentiating between cancer types, would not show causal results of scientific activity on health outcomes. Instead, our identification strategy is based on the spatial variation of scientific activity in cancer research. The idea is that individuals in the same geographic area but diagnosed with different types of cancer may benefit differently depending on local research focus. For instance, a person with breast cancer might benefit more in an area with a strong breast cancer research presence compared to someone with another cancer type.

Figure 1a illustrates the geographic distribution of overall scientific activity and cancer-related mortality. We observe that areas "performing" the best - characterised by a relatively high number of research

articles and a relatively low number of cancer-related deaths (indicated by light blue on the map) - are dispersed across the country in a non-systematic manner. Similarly, areas "performing" the worst - those with relatively low research output and high cancer-related mortality - are found in various parts of the country, including major metropolitan regions.

Additionally, Figure 1b shows an example of distribution of research activity within two selected cancer groups in New South Wales: melanoma and breast cancer. As indicated, there is a variation in expert composition between the areas - some contain local expertise only in breast cancer or in melanoma, or both in the overlapping PHAs. A large portion of areas do not have any scientific activity.

We also account for capturing the effects of potential moving of patients to treatment by using their location prior to the first included year of death.

Lastly, we incorporate two potential control groups in our models to account for variations in research activity. The first control group consists of non-scientific areas, where no cancer research activity is observed. The second control group includes areas that are non-scientific in cancer of interest, but scientific in at least one other cancer group (indicated as "only scientific" in results figures). We assume that these areas, with established research capabilities in other cancer domains, possess the technological and professional infrastructure necessary to conduct research, posing as more appropriate counterfactual. This approach allows us to better isolate the effects of cancer-specific research activity on health outcomes while controlling for broader research capabilities.

4.1 Scientific activity models

To examine whether cancer patients die from cancer at an older age in areas with scientific activity, we estimate the following model:

$$y_{c,r,t} = \tau f(\text{publication}_{c,r,t^0}) + \gamma_c + \gamma_r + \gamma_t + \varepsilon_{c,r,t} \tag{1}$$

where $y_{c,r,t}$ represents the health outcome of interest (e.g., mean age at death) for cancer group c in area r at time t. The term $f(\text{publication}_{c,r,t^0})$ captures different transformations of the number of publications, including absolute counts and the inverse hyperbolic sine (arcsinh) transformation to account for skewness.

Fixed effects γ_r (area) and γ_t (year) control for regional and temporal heterogeneity, while γ_c (cancer group) is included when analyzing outcomes at the cancer-group level.

This model framework enables us to assess associations between research activity and health outcomes at both the aggregate and cancer-group levels. However, as the distribution of publications across geographic areas is not random (Hicks, 2012), we do not infer causality from these estimates. Instead, we extend this approach in the next section to explore causal relationships.

4.2 Stacked models

We analyse cross-sectional data from the general population, examining cancer-specific mortality risk relative to individuals who did not die from that particular type of cancer, including those who passed away from other cancer types. In our stacked analysis, we employ the following specification:

$$y_{r,c} = \tau f(\text{publication}_{r,c,t^0}) + \gamma_r + \gamma_c + \varepsilon_{r,c} \tag{2}$$

where γ_c denotes cancer group and γ_r area fixed effects and y corresponds to health outcomes of interest (number of cancer-related deaths and number of cancer-related deaths per 10,000 population) in cancer group c and area r.

To account for factors unrelated to local scientific activity, our models incorporate both area and cancer fixed effects. For example, improved health outcomes for melanoma in regions with less sun exposure may not be solely attributed to research efforts but could also result from lower natural incidence rates compared to the northern parts of the country.

Despite controlling for area and condition fixed effects, potential endogeneity remains, as the distribution of publishing scientists across PHAs is not random. To address this, we employ a shift-share instrumental variable (SSIV), following Borusyak, Hull and Jaravel (2025). SSIV decomposes variation into two components: shift and share. The shift represents a national trend affecting all regions, while the share captures the extent to which each region is exposed to that trend based on pre-determined data.

Our SSIV approach builds on how Australia's medical education system has evolved in response to workforce supply fluctuations. After two decades of stagnant policy leading to medical staff shortages, five new medical schools were established between 2000 and 2007 (Joyce et al., 2007). As a result, the number of medical graduates more than doubled between 2000 and 2013, rising from around 1,500 to almost 4,000.[15]

This surge in medical graduates not only addressed workforce shortages but also influenced medical research, which is primarily conducted by medical professionals. Research plays a crucial role in the career progression of early-stage doctors seeking specialisation McGrail et al. (2019). Consequently, fluctuations in medical graduate numbers directly impact research activity.

To construct the SSIV, we selected both the number of publications and the number of scientists, using the following formula:

$$SSIV_r = \sum_j (S_{c,r,t^0}/S_{c,t^0}) \times \Delta S_{c,t} \tag{3}$$

[15] `https://www.health.gov.au/sites/default/files/documents/2021/03/doctors-australia-s-future-health-workforce-report.pdf`, accessed on 07.03.2025

where S_{c,r,t^0} denotes the number of publications in condition c in region r between 2000 and 2005, S_{c,t^0} indicates the total number of publications in condition c between 2000 and 2005 in the country and $\Delta S_{c,t}$ indicates the number of new publications in condition c at the national level between 2006 and 2016.

We assume that the sudden increase in medical students in the early 2000s, driven by the establishment of new medical schools, constituted a supply shock at the national level, affecting both the availability of medical professionals and the volume of medical research. Since the location of new medical schools was determined by policy decisions rather than local demand for cancer research, this influx can be considered exogenous.

Consequently, the share-weighted average of this exogenous shift—reflecting the extent to which each region was exposed based on pre-determined factors - can also be considered exogenous, making it a valid instrumental variable.

4.2.1 IV validity

The validity of the shift-share instrument rests on two identifying assumptions (Jaeger, Ruist and Stuhler, 2018; Goldsmith-Pinkham, Sorkin and Swift, 2020): that the pre-period exposure shares are exogenous, and that the national shifts are exogenous (Adao, Kolesár and Morales, 2019; Borusyak, Hull and Jaravel, 2022). We assess each in turn.

Exogeneity of exposure shares. The exposure shares in our instrument are constructed from the historical distribution of research output across cancer groups. For these shares to be exogenous, the allocation of research activity across disease categories must reflect scientific and institutional factors rather than local economic or demographic conditions. We provide two pieces of evidence in support of this assumption.

First, we examine whether pre-treatment area-level demographic characteristics predict the distribution of research output across cancer groups. For each cancer group, we regress the local share of female, male, and over-65 population on the number of publications in that group. Table **??** reports the results. Across top 5 cancer groups and all three demographic characteristics, none of the coefficients are statistically significant, and R^2 values are negligible throughout. This confirms that the cross-sectional distribution of research output is unrelated to local population composition, consistent with shares being driven by scientific priorities rather than local conditions.

Second, we assess whether the first stage is sensitive to the influence of any single cancer group. We re-estimate the first stage sequentially, omitting one cancer group at a time. Table **??** reports the resulting coefficients for both the baseline SSIV specification and an alternative that excludes zero-publication observations. Across the leave-one-out samples for top 5 cancer groups, the first stage coefficient remains

stable and precisely estimated, ranging from 0.636 to 0.791 in the baseline and from 0.609 to 0.760 in the alternative, confirming that the instrument captures a broad and pervasive pattern rather than the idiosyncratic influence of a single disease category.

Exogeneity of national shifts. The national shifts in our instrument are constructed from the staggered opening of medical schools across Australia. Figure 2a plots the cumulative number of medical schools from 1862 to 2008, drawing on commencement dates documented by Geffen (2014). Three distinct waves of expansion are visible. The first, spanning the 1860s to the 1960s, saw the gradual establishment of eight schools — one per state — largely shaped by British medical education standards. The second wave, in the 1970s and 1990s, added only two schools, at Flinders and Newcastle, though both introduced significant curricular innovations. The third and most rapid wave began in 2000, when eight new schools opened within a decade, driven primarily by national workforce shortages and rural maldistribution concerns (Geffen, 2014).

The staggered, policy-driven timing of these openings supports the exogeneity of the national shifts. The post-2000 expansion in particular was mandated by centralised workforce planning priorities and responded to national-level shortages and rural maldistribution rather than to local labour market conditions. The key identifying assumption is that the timing and location of new school openings reflected national workforce considerations rather than local economic outcomes, a claim supported by the historical record and the centralised nature of medical education policy in Australia.

Figure 2b provides a visual validation of the instrument's relevance, plotting the shift-share instrument (SSIV) against the actual number of cancer-related articles for each area–cancer group combination. Areas with higher predicted research intensity do indeed produce more articles.

As a further check on the exogeneity of the exposure shares, Figures 11 and 12 in the Appendix examine the relationship between cancer incidence and research output across areas and cancer groups. Figure 11 plots pre-period incidence (1998–2006) against the number of articles produced in the same period, while Figure 12 plots pre-period incidence against post-period articles (2007–2019). In both figures, the relationship between local disease burden and research output is weak and diffuses: areas with higher incidence do not systematically produce more articles, either contemporaneously or in the subsequent period. This is consistent with the view that research activity is allocated according to national scientific and institutional priorities rather than local epidemiological conditions, and supports the identifying assumption that the exposure shares in our shift-share instrument are not driven by local cancer prevalence.

4.3 Chemotherapy adoption model

Following Agha and Molitor (2018), our causal analysis is conducted in a difference-in-differences framework with listings of new PBS chemotherapy medications as the source of exogenous variation. We compare health outcomes around the time of listing in scientific and non-scientific areas by cancer groups of interest, including both area and time fixed effects.

Assessing pathways, we will aim to see how the adoption of newly approved/PBS-covered drugs disperses in space and again how this affects heterogeneous populations as the poor and the disabled tend to live in different areas that may be less connected to advanced research centers. We will thus assess spillovers to populations in such areas. Our models may take the generic form:

$$y_{c,r,t} = \tau \times 1[\text{drug approved}_{c,t}] \times 1[\text{publication}_{c,r,t^0}] + \gamma_c + \gamma_r + \gamma_t + \varepsilon_{c,r,t} \tag{4}$$

where γ_c indicates cancer group, γ_r area and γ_t year fixed effects and y corresponds to health outcomes of interest in cancer group c, area r at time t in the time pre- and post-listing of chemotherapy medications - drug approved$_{c,t}$ - and comparing areas with and without scientific activity - publication$_{c,r,t^0}$. Our coefficient of interest is τ.

5 Results

Descriptive statistics

The descriptive statistics of the general population and of people who died due to cancer in scientific and non-scientific areas are presented in table 5. We find that residents of scientific areas are older than residents of non-scientific areas in both populations. In general population, residents of non-scientific areas are more likely to die from cancer.

We also find consistent differences between scientific and non-scientific areas in the two populations regarding education, income and employment. Residents of scientific areas tend to be more educated, have higher income and are more likely to be employed.

Model results

Relationship between research activity and mean age at death

The results of analysis measuring the relationship between scientific activity and mean age at death are presented in tables 6 and 7. Table 6 presents the results for 3 pooled models: unweighted, weighted by population, and weighted by number of cancer deaths. The first panel shows the relationship across these specifications between the absolute number of cancer-related articles (without differentiating between

cancer groups) and mean age at death for all cancer groups of interest. The second panel (Panel B) shows the same outcome variable across different specifications with the inverse hyperbolic sine of the number of articles as the independent variable to account for the skewed distribution of number of articles across areas. We find statistically significant estimates for all models, which points to the potential of an important relationship between scientific activity and health outcomes. Panel A results can be interpreted as a nominal change in mean age at death - in unweighted model one additional cancer-related article in an area is associated with 0.008 year increase in mean age at death. In same specification in panel B the result indicates that 1% increase in the number of cancer-related articles is associated with a 0.004 year increase in mean age at death.

The results in table 7 are based on the same two panels (Panel A - absolute number of articles and B - inverse hyperbolic sine of the number of articles) and more detailed specifications and include cancer, year, and health area fixed effects. The first model includes cancer-specific mean age at death between 2012 and 2019 on health area level. The second model includes the same data between 2016 and 2019 and the third model adds a new condition that an area needs to have at least 1 article of any kind to be included in the model. Due to the additional conditions there is a decreasing number of observations in the models.

Similarly to the 6, the results for Panel A and model (1) show that the increase of number of articles by 1 in a given cancer group is associated with 0.001 year increase in mean age at death and for Panel B and model (1) show that 1% increase in the number of articles in a given cancer group is associated with 0.001 year increase in the mean age at death.

Stacked analysis and SSIV

The results of the stacked analysis are presented in Table 10. The first panel reports the effects of scientific activity, as estimated using the model specified in 2, on the number of cancer-related deaths per 10,000 population and the number of premature cancer deaths per 10,000 population. The analysis is conducted using two different control group specifications: models (1) and (3) compare scientific areas to all non-scientific areas, while models (2) and (4) use only areas with scientific activity in at least one cancer group as the control group. The second panel presents results for models (2) and (4) as a robustness check, using the inverse hyperbolic sine of the number of publications as the key variable of interest. This transformation accounts for skewness in the distribution of publication counts while preserving interpretability in a logarithmic-like scale.

We find that one additional publication results in a decrease of number of cancer-related deaths by 0.023 per 10,000 population, or 2.3 per 1 million population and a decrease of 0.012 premature cancer-related deaths per 10,000 population, or 1.2 per 1 million population. Results using the inverse hyperbolic sine of number of articles reinforce these findings, showing slightly more moderate effects. Specifically, the

estimates suggest a reduction of approximately one cancer-related death per 10,000 population, further supporting the relationship between scientific activity and improved health outcomes.

Results for a panel data analysis, where we stack area and cancer-group-level data by years are presented in Appendix Tables 12 and 13. They further reinforce our findings that research activity has a positive impact on health outcomes.

Table 2 presents the results for our two main outcomes of interest when incorporating the SSIV based on publications, as defined in 3. The first-stage results and identification tests - including the weak identification test and the Anderson-Rubin (AR) test, reported in the middle section of the table - confirm the validity of the selected instrument for use in our empirical strategy. Overall, the SSIV estimates closely align with those obtained from the OLS models, reinforcing the robustness of our findings.

The SSIV results using the number of researchers instead of the number of publications are presented in Appendix Table 11. While these results confirm the positive effects of scientific activity on health outcomes, they differ from the main analysis. Diagnostic tests indicate that the number of researchers is a weaker instrument compared to the number of publications in our empirical strategy.

Heterogeneous effects of scientific activity

Our heterogeneity analysis provides a clearer understanding of how the effects of local research activity vary across different population groups. Specifically, we explore differences in impact between males and females, as well as between residents of more and less disadvantaged areas. While women generally utilise healthcare services more frequently than men and tend to have higher overall healthcare expenditures (Owens, 2008), they also face significant disparities in medical research and treatment. One well-documented issue is their underrepresentation in clinical trials, which creates gaps in understanding how diseases manifest differently across genders and how treatments affect women specifically (Holdcroft, 2007). Additionally, research funding has historically been skewed toward conditions that predominantly affect men, leading to a relative underinvestment in diseases that disproportionately impact women—an issue that is particularly concerning in the context of cancer treatment (Donington and Colson, 2011).

Beyond gender differences, socioeconomic disparities also play a crucial role in shaping health outcomes. Research consistently shows that individuals living in more disadvantaged areas of Australia experience worse cancer outcomes compared to those in more affluent regions (Afshar et al., 2020). This suggests that access to healthcare, early diagnosis, and treatment quality may be unevenly distributed, further reinforcing the importance of local research activity in addressing these disparities. By examining these differences, our study aims to shed light on whether and how scientific activity contributes to reducing - or potentially exacerbating - existing inequalities in healthcare.

Tables 3 and 4 present the results of the heterogeneity analysis by gender and socio-economic disadvantage. The findings indicate that both males and females experience significant reductions in cancer-related

mortality associated with scientific activity, with estimated decreases of 4.6 deaths per 1 million for males and 2.5 per 1 million for females. The larger effect observed for males suggests they may derive greater benefits from research activity, highlighting potential disparities in how scientific advancements translate into health outcomes across genders.

These results are further supported by the SSIV analysis, which shows similar reductions - 2.7 deaths per 1 million for females and 3.8 per 1 million for males. Additionally, when restricting the analysis to areas with some level of research activity, we observe comparable effects, with reductions of 2.2 per 1 million for females and 3.7 per 1 million for males.

When focusing on premature mortality, we find more moderate effects overall; however, the gender pattern remains consistent, with males exhibiting larger reductions in premature deaths than females.

Our findings on socio-economic disparities indicate that residents in more disadvantaged areas benefit relatively more from scientific activity than those in less disadvantaged areas. While cancer mortality rates are higher in disadvantaged areas - averaging 5.3 deaths per 10,000 population compared to 3.8 deaths per 10,000 in less disadvantaged areas - the reduction associated with scientific activity is estimated at 0.029 and 0.013 deaths per 10,000 population, respectively. These effects correspond to a 0.5% reduction in cancer-related deaths in disadvantaged areas and a 0.3% reduction in less disadvantaged areas, suggesting that research activity may play a role in narrowing health inequities.

Chemotherapy adoption model

The results for chemotherapy adoption model are presented graphically in Figures 3, 4, 5 and 6. All models are conducted on an area level with outcomes aggregated on a quarterly level and with time fixed effects to control for overall improvement in cancer treatment over time. Period indicated as 0 on x axis is the first quarter when a chemotherapy medication was listed on the PBS (with either a new item number or change to restriction). We included 5 quarters before and 5 quarters after the listing in each case.

The first model (in Figure 3) shows the uptake in all PHAs with very clear adoption from the first period after listing. A similar pattern is presented for the number of patients treated with new medications (Figure 4). Subsequently, we show the comparison for the same outcomes between scientific and non-scientific areas in Figures 5 and 6. For both outcomes there are comparable results in scientific and non-scientific areas with an increase in the outcome of interest right after chemotherapy listing.

6 Discussion

Scientific research is the primary method for understanding various aspects of the world in an objective way. It plays a fundamental role in data-driven decision-making and making technological advancements, with its growing relevance reflected in the exponential rise in publication volume (Bornmann and Mutz,

2015). In medicine, a highly experimental field, research is essential for addressing complex health challenges, yet its broader impact on health outcomes remains unclear. While major breakthroughs, such as life-saving medications and medical devices, are widely acknowledged, it is crucial to assess how the overall substantial investments - such as the USD 24.5 billion spent globally on cancer research between 2016 and 2020 (McIntosh et al., 2023) - translate into societal benefits.

Our results show positive overall effects of local scientific activity on health outcomes in cancer. While the effects might seem moderate - showing that one additional publication prolongs life of people that die from cancer by 0.009 of a year and decreases number of cancer-deaths by around 2.3 per 1 million population - these are additional benefits to an already well-established field of treatments, with an over 80-year history in chemotherapy Scott (1970) and over 100-year history in radiotherapy Holsti (1995).

We interpret these results as evidence that local researchers with expertise in a given cancer type make valuable contributions to medical knowledge and that this knowledge efficiently spreads to the local physicians.

While the results regarding adoption of newly-listed chemotherapy treatments do not show any differences in scientific and non-scientific areas in the volume of administered medications, there might be some differences regarding the appropriateness of use of these medications that could be studied using more detailed patient data.

Our heterogeneity analysis reveals that patients living in more disadvantaged areas experience relatively greater benefits from research advancements compared to those in less disadvantaged areas. This finding presents a strong case for increasing investment in medical research, where the impact of innovation can be even more pronounced.

However, our analysis also highlights a concerning disparity: female patients tend to benefit less from research activity than male patients. This underscores the urgent need for more equitable clinical studies that not only ensure adequate female representation but also recognize gender-specific differences in disease progression and treatment efficacy. For example, the lack of gender-inclusive research can result in severe consequences, such as adverse drug reactions that disproportionately affect women. In some instances, medications have been withdrawn from use in female patients after post-market analysis revealed significantly higher risks for women compared to men Simon (2005). Addressing these gender disparities in healthcare research is essential for developing equitable and effective healthcare solutions.

While our study uses very detailed individual-level data, we do not have access to the diagnoses made by physicians during treatments. We think that a study using diagnosis data might disentangle the effects even more - as it can show how the volume of research might impact survival rates or curing from cancer.

References

Adao, Rodrigo, Michal Kolesár and Eduardo Morales. 2019. "Shift-Share Designs: Theory and Inference." *Quarterly Journal of Economics* 134(4):1949–2010.

Afshar, Nina, Dallas R. English, Tony Blakely, Vicky Thursfield, Helen Farrugia, Graham G. Giles and Roger L. Milne. 2020. "Differences in cancer survival by area-level socio-economic disadvantage: A population-based study using cancer registry data." *PLoS One* 15(1):e0228551.

Agha, Leila and David Molitor. 2018. "The local influence of pioneer investigators on technology adoption: evidence from new cancer drugs." *Review of Economics and Statistics* 100(1):29–44.

Azoulay, Pierre, Joshua S Graff Zivin and Bhaven N Sampat. 2011. The Diffusion of Scientific Knowledge Across Time and Space: Evidence from Professional Transitions for the Superstars of Medicine. Working Paper 16683 National Bureau of Economic Research.
http://www.nber.org/papers/w16683

Barrenho, Eliana, Eric Gautier, Marisa Miraldo, Carol Propper and Christiern Rose. 2025. "Innovation Diffusion Among Coworkers: Evidence from Senior Doctors." *Management Science* .
https://doi.org/10.1287/mnsc.2023.00496

Bornmann, Lutz and Rüdiger Mutz. 2015. "Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references." *Journal of the Association for Information Science and Technology* 66(11):2215–2222.
https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.23329

Borusyak, Kirill, Peter Hull and Xavier Jaravel. 2022. "Quasi-Experimental Shift-Share Research Designs." *Review of Economic Studies* 89(1):181–213.

Borusyak, Kirill, Peter Hull and Xavier Jaravel. 2025. "A Practical Guide to Shift-Share Instruments." *Journal of Economic Perspectives* 39(1):181–204.
https://www.aeaweb.org/articles?id=10.1257/jep.20231370

Coleman, James, Elihu Katz and Herbert Menzel. 1957. "The Diffusion of an Innovation Among Physicians." *Sociometry* 20(4):253–270.
http://www.jstor.org/stable/2785979

Cramb, Susanna M., Paula Moraga, Kerrie L. Mengersen and Peter D. Baade. 2017. "Spatial variation in cancer incidence and survival over time across Queensland, Australia." *Spatial and Spatio-temporal Epidemiology* 23:59–67.
https://www.sciencedirect.com/science/article/pii/S1877584516301174

Densen, P. 2011. "Challenges and opportunities facing medical education." *Transactions of the American Clinical Climatological Association* 122:48–58.
https://pmc.ncbi.nlm.nih.gov/articles/PMC3116346/

Donington, Jessica S. and Yolonda L. Colson. 2011. "Sex and Gender Differences in Non-Small Cell Lung Cancer." *Seminars in Thoracic and Cardiovascular Surgery* 23(2):137–145.
https://www.sciencedirect.com/science/article/pii/S1043067911000864

Geffen, Laurence. 2014. "A brief history of medical education and training in Australia." *Medical Journal of Australia* 201(1):S19–S22.

Goldsmith-Pinkham, Paul, Isaac Sorkin and Henry Swift. 2020. "Bartik Instruments: What, When, Why, and How." *American Economic Review* 110(8):2586–2624.

Greenblatt, Wesley. N.d. "Building on Solid Ground: Foundational Knowledge and the Dynamics of Innovation.".
https://papers.ssrn.com/abstract=3919866

Hall, Bronwyn H., Adam Jaffe and Manuel Trajtenberg. 2005. "Market Value and Patent Citations." *The RAND Journal of Economics* 36(1):16–38.
http://www.jstor.org/stable/1593752

Hicks, Diana. 2012. "Performance-based university research funding systems." *Research Policy* 41(2):251–261.
https://www.sciencedirect.com/science/article/pii/S0048733311001752

Holdcroft, Anita. 2007. "Gender bias in research: how does it affect evidence based medicine?" *Journal of the Royal Society of Medicine* 100(1):2–3. PMID: 17197669.
https://doi.org/10.1177/014107680710000102

Holsti, L. R. 1995. "Development of Clinical Radiotherapy Since 1896." *Acta Oncologica* 34(8):995–1003.
https://pubmed.ncbi.nlm.nih.gov/8608037/

Jaeger, David A., Joakim Ruist and Jan Stuhler. 2018. "Shift-share instruments and the impact of immigration." *NBER Working Paper No. 24285* .

Jaffe, Adam B. 1986. "Technological Opportunity and Spillovers of R D: Evidence from Firms' Patents, Profits, and Market Value." *The American Economic Review* 76(5):984–1001.
http://www.jstor.org/stable/1816464

Jaffe, Adam B., Manuel Trajtenberg and Rebecca Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *The Quarterly Journal of Economics* 108(3):577–598.
http://www.jstor.org/stable/2118401

Joyce, Catherine M, Johannes U Stoelwinder, John J McNeil and Leon Piterman. 2007. "Riding the wave: current and emerging trends in graduates from Australian university medical schools." *Medical Journal of Australia* 186(6):309–312.
https://onlinelibrary.wiley.com/doi/abs/10.5694/j.1326-5377.2007.tb00907.x

Klopfenstein, D. and Will Dampier. 2020. "Commentary to Gusenbauer and Haddaway 2020: Evaluating retrieval qualities of Google Scholar and PubMed." *Research synthesis methods* 12.

Lichtenberg, Frank R. 2013. "The impact of therapeutic procedure innovation on hospital patient longevity: Evidence from Western Australia, 2000–2007." *Social Science & Medicine* 77:50–59.
https://www.sciencedirect.com/science/article/pii/S0277953612007538

Luo, Qingwei, Dianne O'Connell, Xue Qin Yu, Clare Kahn, Michael Caruana, Francesca Pesola, Peter Sasieni, Paul Grogan, Sanchia Aranda, Citadel Cabasag, Isabelle Soerjomataram, Julia Steinberg and Karen Canfell. 2022. "Cancer incidence and mortality in Australia from 2020 to 2044 and an exploratory analysis of the potential effect of treatment delays during the COVID-19 pandemic: a statistical modelling study." *The Lancet Public Health* 7:e537–e548.

McGrail, Matthew Richard, Belinda G O'sullivan, Hollie R Bendotti and Srinivas Kondalsamy-Chennakesavan. 2019. "Importance of publishing research varies by doctors' career stage, specialty and location of work." *Postgraduate Medical Journal* 95(1122):198–204.
https://doi.org/10.1136/postgradmedj-2019-136473

McIntosh, Stuart, Fareeha Alam, Laura Adams, Ian Boon, Jonathan Callaghan, Isabella Conti, Ellen Copson, Victoria Carson, Mitchell Davidson, Helen Fitzgerald, Ashram Gautam, Christopher Jones, Saffron Kargbo, Amatta Mirandari, Natasha Moore, Ross Moore, Aidan Murray, Lydia Newman, Stephen Robinson and Michael Head. 2023. "Global funding for cancer research between 2016 and 2020: a content analysis of public and philanthropic investments." *The Lancet Oncology* 24:636–645.

McKibbin, Rebecca and Bruce A Weinberg. 2021. Does Research Save Lives? The Local Spillovers of Biomedical Research on Mortality. Working Paper 29420 National Bureau of Economic Research.
http://www.nber.org/papers/w29420

Mellish, Leigh, Emily Karanges, Melisa Litchfield, Andrea Schaffer, Bianca Blanch, Benjamin Daniels, Alicia Segrave and Sallie-Anne Pearson. 2015. "The Australian Pharmaceutical Benefits Scheme data collection: A practical guide for researchers." *BMC Research Notes* 8.

OECD and Eurostat. 2005. *Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data, 3rd Edition*.
https://www.oecd-ilibrary.org/science-and-technology/oslo-manual_9789264013100-en

Owens, Gary. 2008. "Gender Differences in Health Care Expenditures, Resource Utilization, and Quality of Care." *Journal of Managed Care Pharmacy* 14(3 Supp A):2–6. PMID: 18439060.
https://doi.org/10.18553/jmcp.2008.14.S3-A.2

Scott, R. B. 1970. "Cancer chemotherapy–the first twenty-five years." *BMJ* 4(5730):259–265.
https://www.bmj.com/content/4/5730/259

Selby, P, C Gillis and R Haward. 1996. "Benefits from specialised cancer care." *The Lancet* 348(9023):313–318.
https://doi.org/10.1016/S0140-6736(96)02482-8

Sharplin, Greg, Samantha Bannister, Marion Eckert, David Roder and Brenda Wilson. 2014. *Cancer Forum* 38(2):143–149.
https://search.informit.org/doi/10.3316/informit.490522784977980

Simon, Viviana. 2005. "Wanted: Women in Clinical Trials." *Science* 308(5728):1517–1517.
https://www.science.org/doi/abs/10.1126/science.1115616

Skinner, Jonathan and Douglas Staiger. 2015. "Technology Diffusion and Productivity Growth in Health Care." *The Review of Economics and Statistics* 97(5):951–964.

Tervonen, Hanna, Nicola Creighton, George Zhao, Megumi Ng and David Currow. 2020. "Capture of systemic anticancer therapy use by routinely collected health datasets." *Public Health Research & Practice* 30.
http://www.jstor.org/stable/2785979

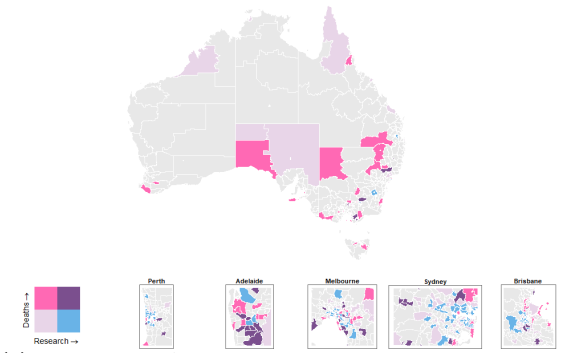
Torvik, Vetle I. N.d. "MapAffil: A Bibliographic Tool for Mapping Author Affiliation Strings to Cities and Their Geocodes Worldwide." . Forthcoming.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4861081/

Vadia, Rucha and Katharina Blankart. 2021. "Regional Innovation Systems of Medical Technology - A knowledge production function of cardiovascular research and funding in Europe." 8:57–81.

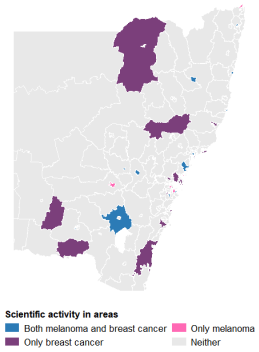
Xu, Jian, Sunkyu Kim, Min Song, Minbyul Jeong, Donghyeon Kim, Jaewoo Kang, Justin F. Rousseau, Xin Li, Weijia Xu, Vetle I. Torvik, Yi Bu, Chongyan Chen, Islam Akef Ebeid, Daifeng Li and Ying Ding. N.d. "Building a PubMed Knowledge Graph." . Forthcoming.
https://www.nature.com/articles/s41597-020-0543-2

7 Tables and Figures

Figure 1: Geographic distribution of cancer research and mortality in Australia

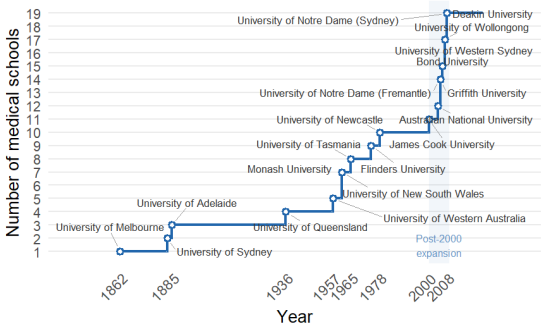


(a) Distribution of cancer-related research and cancer deaths across Australia

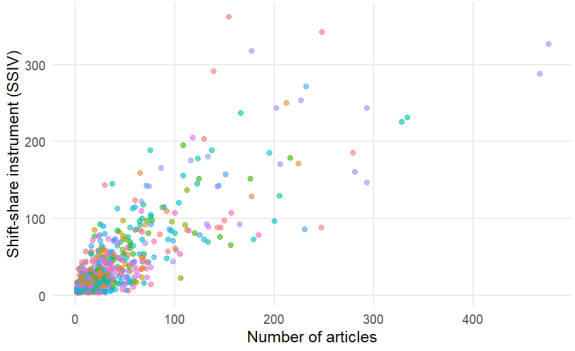


(b) NSW scientific and non-scientific areas in breast cancer and melanoma

Figure 2: Instrument construction and relevance



(a) New medical school openings in Australia



(b) Number of articles and SSIV

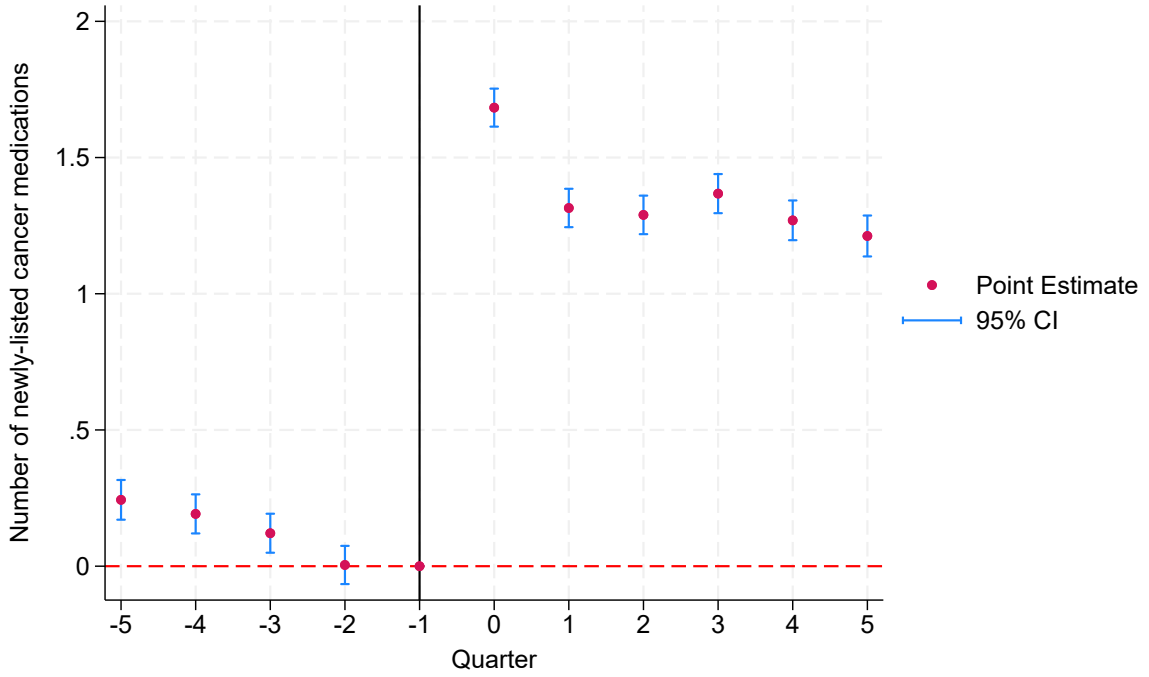


Figure 3: Event study of number of newly-listed chemotherapy medications pre- and post-listing in all PHAs in Australia.

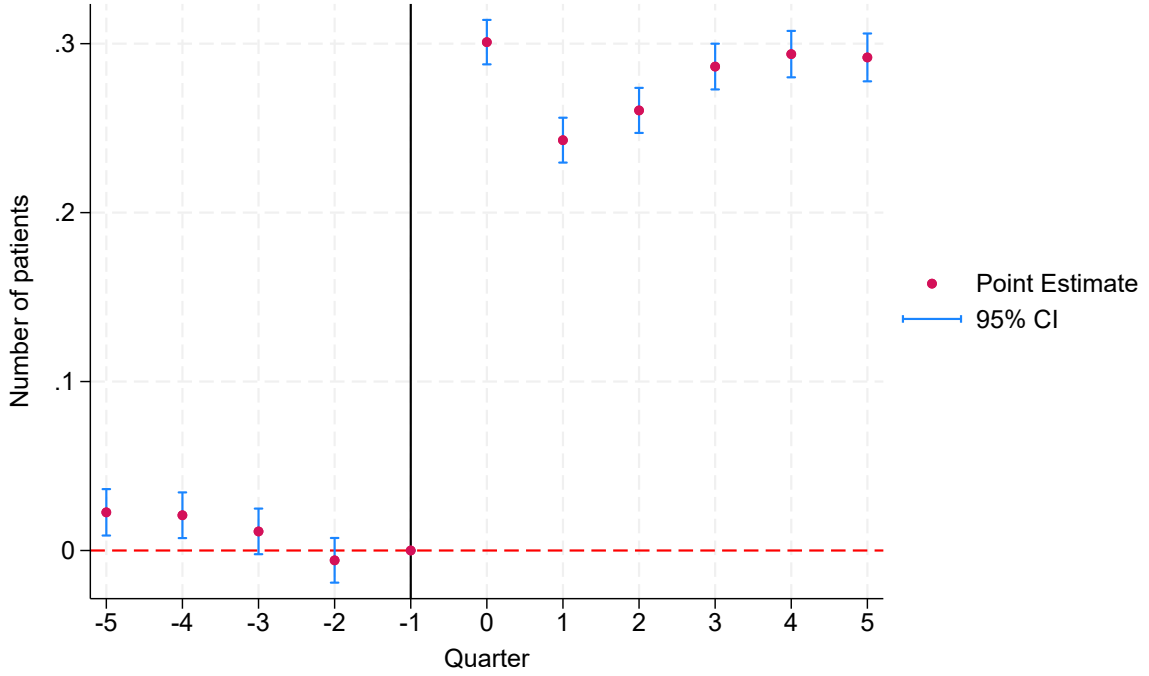


Figure 4: Event study of number of patients receiving newly-listed chemotherapy pre- and post-listing in all PHAs in Australia.

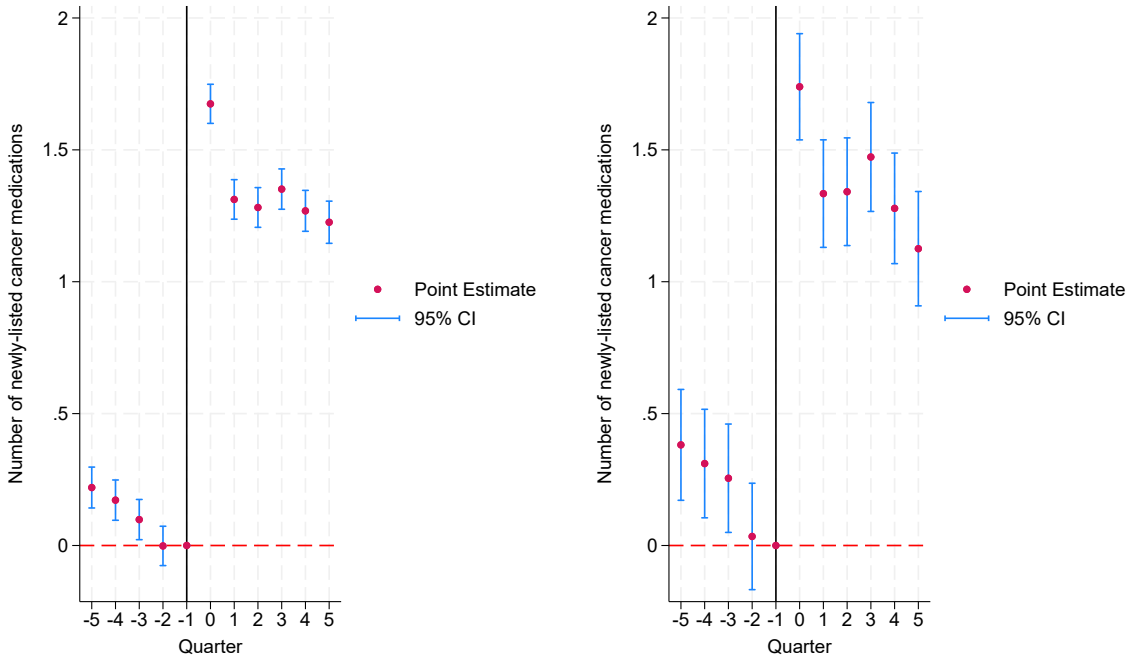


Figure 5: Event study of number of newly-listed chemotherapy medications pre- and post-listing in non-scientific (left) and scientific areas (right).

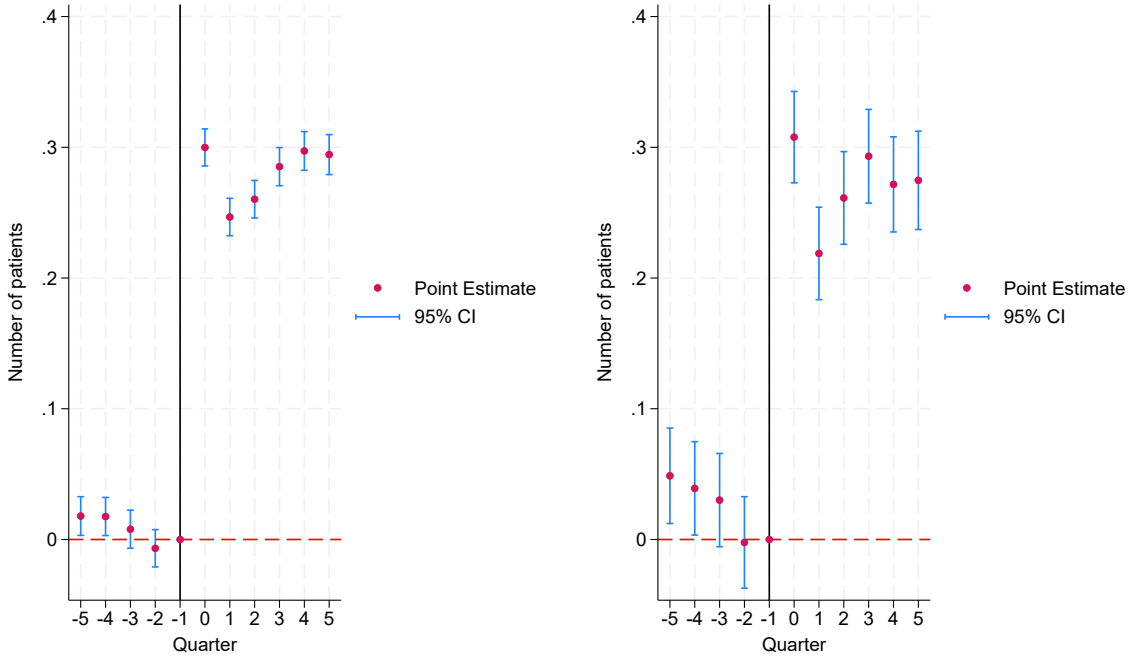


Figure 6: Event study of number of patients receiving newly-listed chemotherapy pre- and post-listing in non-scientific (left) and scientific areas (right).

Table 1: IV validity checks: leave-one-out first stage and exogeneity of exposure shares

	Leave-one-out first stage		Demographic predictors		
Cancer group	SSIV with 0	SSIV without 0	Female share	Male share	Over-65 share
Tracheal, bronchus and lung	0.7909***	0.7605***	2.854	−2.854	0.725
	(0.0712)	(0.0714)	(1.967)	(1.967)	(1.635)
Colon and rectum cancer	0.7384***	0.7080***	2.642	−2.642	−0.690
	(0.0720)	(0.0711)	(3.179)	(3.179)	(2.345)
Prostate cancer	0.7704***	0.7401***	−1.090	1.090	−0.770
	(0.0799)	(0.0796)	(2.076)	(2.076)	(1.534)
Breast cancer	0.6362***	0.6087***	3.516	−3.516	−2.452
	(0.0577)	(0.0560)	(4.931)	(4.931)	(4.004)
Pancreatic cancer	0.7515***	0.7206***	−0.133	0.133	−0.211
	(0.0750)	(0.0746)	(0.582)	(0.582)	(0.211)
N	22,116	3,306	1,164	1,164	1,164

Notes: The left panel reports leave-one-out first stage coefficients omitting one cancer group at a time, for the baseline SSIV specification (with zeros) and an alternative excluding zero-publication observations. The right panel reports coefficients from regressions of area-level demographic shares on the number of publications in each cancer group; none are statistically significant at conventional levels. Only the five largest cancer groups by number of deaths are shown. Standard errors in parentheses. * p<0.05, ** p<0.01, *** p<0.001

Table 2: SSIV results: publications as IV

	(1)	(2)		(3)	
	First stage	Cancer deaths per pop.		Premature deaths per pop.	
		OLS	IV	OLS	IV
Number of articles	0.772***	-0.023***	-0.023***	-0.012***	-0.012***
	(0.077)	(0.002)	(0.003)	(0.001)	(0.001)
N	23,280	23,280	23,280	23,280	23,280
Mean	1.213	4.572	4.572	2.280	2.280
F-stat	100.3	-	-	-	-
F-stat, weak ID			95.670	-	95.670
F-stat, AR	-	-	40.736	-	47.911
P-value, AR	-	-	0.0000	-	0.0000
Condition FEs	✓	✓	✓	✓	✓
PHA FEs	✓	✓	✓	✓	✓

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes:

Table 3: Relationship between number of publications and cancer deaths among males and females

	(1) Cancer deaths per pop.	(2) Cancer deaths per pop.	(3) Premature deaths per pop.	(4) Premature deaths per pop.
Panel a. Number of articles:				
Females	-0.025***	-0.022***	-0.014***	-0.011***
	(0.003)	(0.003)	(0.002)	(0.002)
Males	-0.046***	-0.037***	-0.019***	-0.014***
	(0.004)	(0.004)	(0.003)	(0.002)
Panel b. SSIV: publications				
Females	-0.027***	-	-	-
	(0.004)	-	-	-
Males	-0.038***	-	-	-
	(0.006)	-	-	-
N	23,280	8,206	23,280	8,206
Cancer group FEs	✓	✓	✓	✓
Health area FEs	✓	✓	✓	✓
Only scientific		✓		✓

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes:

Table 4: Heterogeneity: SEIFA

	(1) Cancer deaths per pop.; IRSD above median	(2) Cancer deaths per pop.; IRSD below median	(3) Cancer deaths per pop.; IRSD above median	(4) Cancer deaths per pop.; IRSD below median
Panel a. OLS				
Number of articles	-0.013***	-0.029***	-0.011***	-0.026***
	(0.003)	(0.003)	(0.003)	(0.002)
Panel b. SSIV: publications				
Number of articles	-0.012***	-0.025***	-	-
	(0.003)	(0.004)	-	-
N	11,640	11,640	4,686	3,520
m	3.840	5.316	3.706	4.978
Cancer group FEs	✓	✓	✓	✓
Health area FEs	✓	✓	✓	✓
Only scientific			✓	✓

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes:

A Appendix

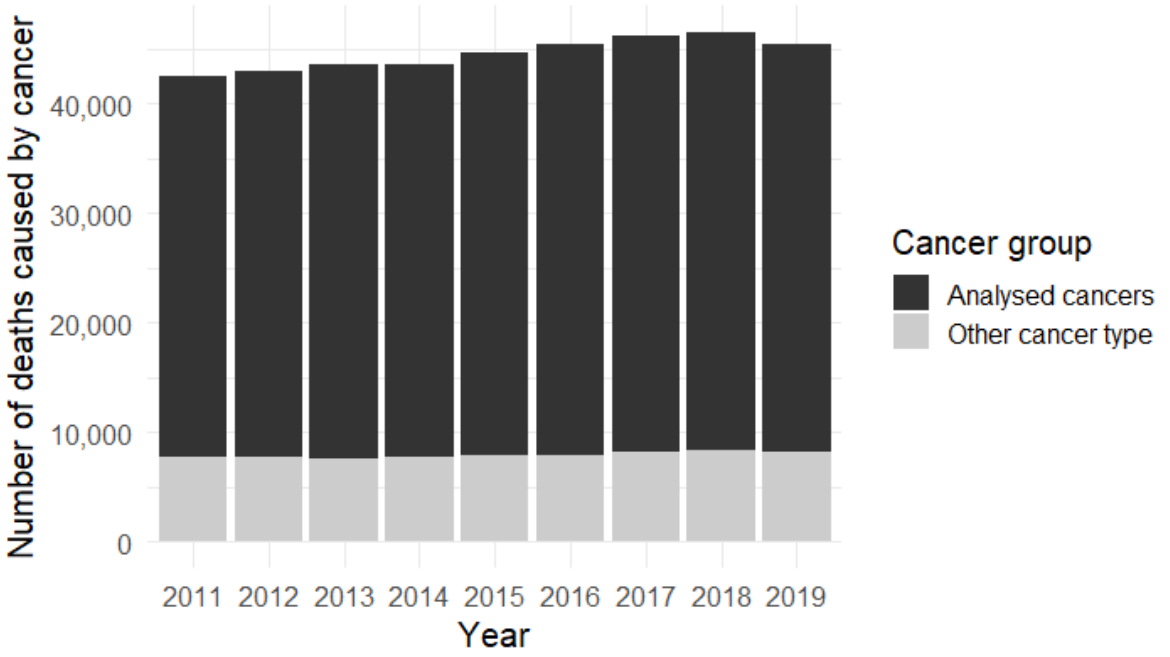


Figure 7: Number of all cancer deaths reported between 2011 and 2019, including the number of cancer deaths with reported ICD-10 of interest as the main cause of death ("Analysed cancers") and other ("Other cancer type").

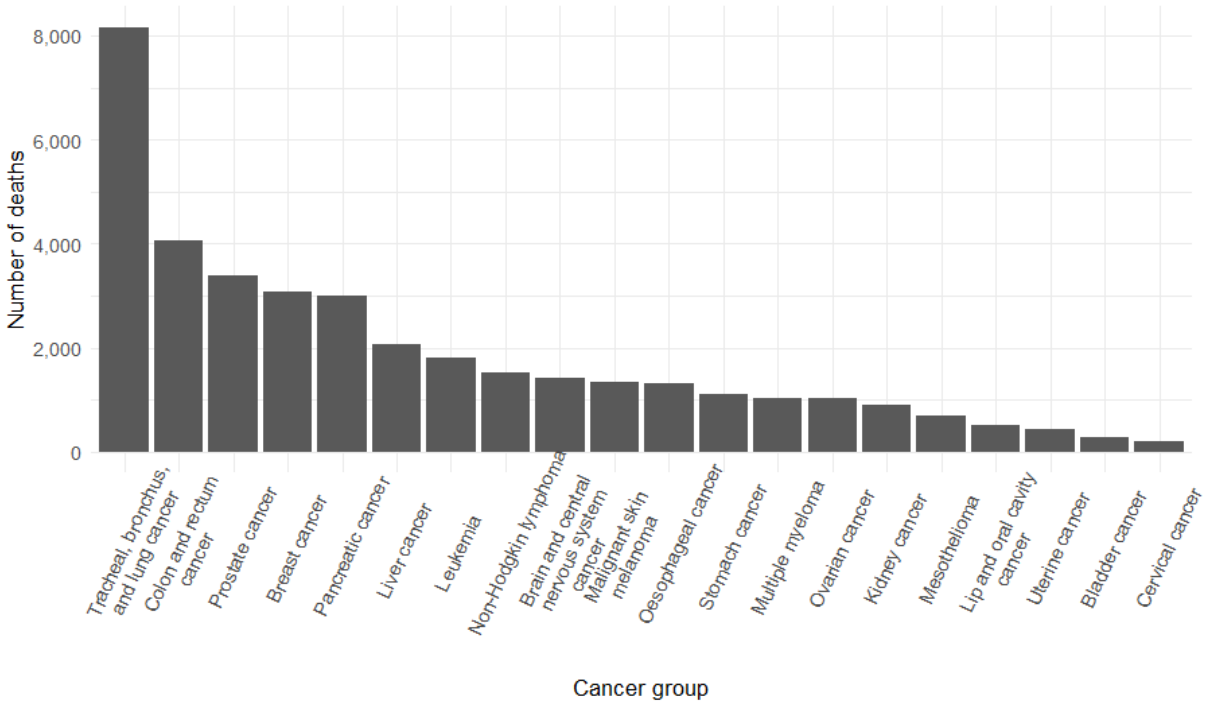


Figure 8: Number of deaths caused by cancer by cancer group (2019). Only cancer groups of interest are presented.

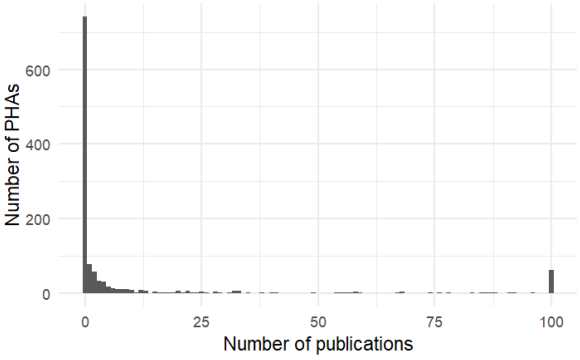


Figure 9: Distribution of number of articles by PHA. All PHAs with over 100 articles are censored at 100.

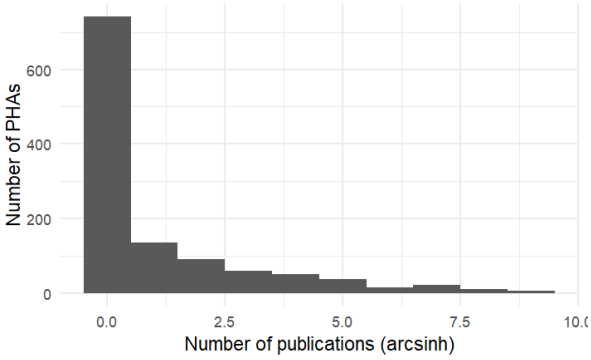


Figure 10: Distribution of inverse hyperbolic sine of number of articles by PHA.

Table 5: Population Characteristics by Scientific and Non-Scientific Areas

	Cancer-Death Population				General Population			
	All	Scientific	Non-scientific	Diff.	All	Scientific	Non-scientific	Diff.
Age in 2011	69.85	70.31	69.52	-0.78***	37.97	38.17	37.82	-0.34***
Probability of dying from cancer	1.00	1.00	1.00	0.00	0.01	0.01	0.01	0.00***
% Degree or higher	0.04	0.05	0.03	-0.02***	0.14	0.17	0.12	-0.05***
% Advanced degree	0.06	0.07	0.06	-0.01***	0.08	0.08	0.08	-0.00***
% Certificate	0.07	0.06	0.07	0.01***	0.11	0.09	0.12	0.03***
% Year 12 or below	0.22	0.21	0.23	0.02***	0.26	0.24	0.28	0.04***
% Low income	0.38	0.37	0.40	0.03***	0.25	0.25	0.26	0.01***
% Middle income	0.41	0.40	0.41	0.01**	0.28	0.27	0.29	0.02***
% High income	0.12	0.14	0.11	-0.03***	0.28	0.30	0.26	-0.04***
% Employed	0.15	0.15	0.14	-0.00	0.51	0.52	0.50	-0.02***
N	257,571	107,036	150,535	257,571	21,851,354	9,525,734	12,325,620	21,851,354

Notes: Statistical significance levels are indicated by *** (p¡0.01), ** (p¡0.05), and * (p¡0.1).

Table 8: MeSH to ICD-10 codes crosswalk

GBD Cancer Group	ICD10	MeSH Codes
Tracheal, bronchus, and lung cancer	C33-C34	D001984, D002283, D002289, D008175, D010178, D010997, D014134, D016066, D018202, D047868, D054363, D055613, D055752
Colon and rectum cancer	C18-C21	D001005, D001063, D002430, D003110, D004379, D005736, D007078, D007580, D012004, D012811, D015179
Prostate cancer	C61	D011471, D064129
Breast cancer	C50	D001943, D018567, D018270, D018275, D058922, D064726
Pancreatic cancer	C25	D003969, D007340, D007516, D010190, D013005, D015408, D018273, D021441

GBD Cancer Group	ICD10	MeSH Codes
Non-Hodgkin lymphoma	C82-C86	D002051, D008224, D008228, D008230, D009182, D012751, D016393, D016399, D016400, D016403, D016410, D016411, D016483, D017728, D017731, D018442, D020522, D054391, D054446, D054685, D058527
Oesophageal cancer	C15	D004938
Malignant skin melanoma	C43	D008545, D018327, D018328
Stomach cancer	C16	D013274
Brain and central nervous system cancer	C70-C72	D001254, D004806, D005909, D005910, D009837, D018316, D020339, D001254, D004806, D005909, D005910, D009837, D018316, D020339, D001254, D004806, D005909, D005910, D009837, D018316, D020339, D001932, D002528, D002551, D007029, D008577, D008579, D010871, D010911, D013120, D015173, D015174, D015192, D016080, D016543, D016545, D018306, D020288, D020295, D020863, D056364, D003390, D019574, D020339, D003390, D019574, D020339
Kidney cancer	C64-C65	D002292, D007680, D009396, D018201, D030321
Bladder cancer	C23-C24	D001749
Multiple myeloma	C88-C90	D007952, D009101
Ovarian cancer	C56	D006106, D008539, D010051, D013798, D018310, D018311
Leukemia	C91-C95	D007938, D007943, D007945, D015448, D015451, D015452, D015456, D015458, D015459, D015461, D015463, D054066, D054198, D054218, D007946, D001752, D004915, D007946, D007947, D007948, D007951, D015464, D015465, D015466, D015470, D015471, D015473, D015477, D015479, D023981, D054429, D054438, D007952, D007953
Lip and oral cavity cancer	C00-C08	D005887, D007972, D008048, D009062, D010157, D010307, D012468, D013362, D013365, D014062
Mesothelioma	C45	D008654, D018261

GBD Cancer Group	ICD10	MeSH Codes
Liver cancer	C22	D018197, D018281, D018285, D006528, D008113, D018248
Uterine cancer	C54	D014594, D016889, D018269
Cervical cancer	C53	D002583

Table 9: PBS listings between 2013 and 2017.

Molecule	Listing date	PBS item code	Cancer group
arsenic	1/04/2016	10699D, 7241D, 10691Q, 4371C	leukaemia
bendamustine	1/05/2016	10763L, 10760H	non-Hodgkin's lymphoma
bevacizumab	1/08/2014	10120P, 10115J, 10114H, 10121Q	ovarian cancer
bevacizumab	1/09/2016	10885X, 10881Q	cervical cancer
blinatumomab	1/05/2017	11115B, 11116C, 11119F, 11117D, 11118E, 11120G	leukaemia
brentuximab	1/12/2014	10166C, 10171H	lymphoma
brentuximab	1/04/2017	11067L, 11080E, 11086L, 11089P, 11073T, 11079D, 11087M, 11096B	Hodgkin's lymphoma
ceritinib	1/02/2017	11056X	lung cancer
cetuximab	1/01/2015	4436L, 4731B, 7242E, 7273T	colorectal cancer
cetuximab	1/06/2015	10262D, 4436L, 4731B, 10265G, 7242E, 7273T	colorectal cancer
cobimetinib	1/04/2017	11074W, 11075X	melanoma
crizotinib	1/07/2015	10323H, 10322G	lung cancer
dabrafenib	1/12/2013	10003L, 2846T, 2954L, 2963Y	melanoma
dabrafenib	1/04/2016	10003L, 2846T, 2954L, 2963Y	melanoma
denosumab	1/04/2014	10061M	bone tumour
enzalutamide	1/12/2014	10174L	prostate cancer
eribulin	1/10/2014	10140Q, 10144X	breast cancer
erlotinib	1/01/2014	10022L, 10028T, 10020J, 10019H, 10014C, 10025P	lung cancer

Molecule	Listing date	PBS item code	Cancer group
everolimus	1/06/2014	2819J, 2985D	breast cancer
everolimus	1/09/2014	10133H, 10131F, 10132G, 10135K	kidney cancer
everolimus	1/04/2015	10131F,10132G,10133H,10135K	pancreas cancer
gefitinib	1/01/2014	8769M	lung cancer
lapatinib	1/05/2016	9148L	breast cancer
lenalidomide	1/02/2017	11036W, 11063G, 11042E, 11055W, 11029L, 11064H, 11062F, 11041D	multiple myeloma
lenvatinib	1/12/2016	10952K	thyroid cancer
leuprorelin	1/03/2016	10656W	prostate cancer
leuprorelin	1/12/2016	10962Y, 10963B, 10969H	prostate cancer
mercaptop-urine	1/03/2015	10214N	leukaemia
nivolumab	1/05/2016	10748Q, 10775D, 10745M, 10764M	melanoma
nivolumab	1/08/2017	11143L, 11152Y, 11157F, 11159H, 10775D, 11150W, 11153B, 11160J	lung cancer, kidney cancer
obinu-tuzumab	1/08/2015	10418H, 10407R	leukaemia
ofatumumab	1/04/2015	10249K, 10236R, 10252N, 10240Y, 10237T, 10239X	leukaemia
olaparib	1/02/2017	11034R, 11050N	ovarian cancer
paclitaxel	1/11/2014	10150F, 10165B	pancreas cancer
panitu-mumab	1/04/2014	10069Y, 10082P	colorectal cancer
panitu-mumab	1/01/2015	10082P, 10069Y	colorectal cancer
pem-brolizumab	1/09/2015	10424P, 10475H, 10436G, 10493G	melanoma
pem-brolizumab	1/04/2017	10424P, 10475H, 10436G, 10493G	melanoma

Molecule	Listing date	PBS item code	Cancer group
pertuzumab	1/07/2015	10268K, 10308M, 10334X	breast cancer
pomalido-mide	1/08/2015	10417G, 10386P, 10406Q, 10387Q	multiple myeloma
ponatinib	1/11/2015	10520Q, 10523W, 10524X, 10530F	leukaemia
rituximab	1/12/2014	9544H, 9611W	lymphoma
rituximab	1/04/2015	10193L, 7257Y, 7258B, 7259C, 10179R, 4613T, 4614W, 4615X	non-Hodgkin's lymphoma
sunitinib	1/12/2013	10004M, 10009T, 2959R, 2842N, 2837H, 10010W	pancreas cancer
tamoxifen	1/10/2016	10911G, 1880Y	breast cancer
temozolo-mide	1/04/2014	10062N	brain tumour
trametinib	1/08/2015	10385N, 10403M, 10382K, 10405P	melanoma
trametinib	1/04/2016	10382K, 10385N, 10403M, 10405P	melanoma
trastuzumab	1/07/2015	10270M, 10296X, 10324J, 10281D, 10269L, 10346M, 10362J, 10282E	breast cancer
trastuzumab	1/01/2016	10575N, 10589H, 10597R, 10581X, 10588G, 10595P	stomach cancer
trastuzumab	1/04/2016	10682F, 10721G	breast cancer
trastuzumab	1/05/2016	10281D, 10282E	breast cancer
vemurafenib	1/04/2017	11076Y, 11081F	melanoma
vorinostat	1/07/2017	11138F, 11141J	lymphoma

Table 6: Scientific activity and mean age at death

Dependent variables: Mean age at death due to cancer in the Population Health Area			
	Pooled	Weighted	
	Raw	Pop	Deaths
	(1)	(2)	(3)
Panel A. Number of cancer articles	0.008***	0.005***	0.007***
	(0.002)	(0.000)	(0.001)
Panel B. Arcsinh of number of cancer research articles weighted by population	0.399***	0.337***	0.303***
	(0.048)	(0.000)	(0.007)
N	10,354	10,354	10,354
Cancer group FEs			
Year FEs	✓	✓	✓
Health area FEs	✓	✓	✓

Notes: The Table presents coefficients estimates from equation (1)

Table 7: Scientific activity and mean age at death by cancer group

Dependent variables: Mean age at death due to specific cancer in cancer groups in the Population Health Area	By cancer	Post 2016	(4), (5), Any research
	(4)	(5)	(6)
Panel A. Number of cancer articles	0.001***	0.009***	0.007***
	(0.000)	(0.000)	(0.000)
Panel B. Arcsinh of number of cancer research articles weighted by population	0.101***	0.120***	0.070***
	(0.001)	(0.001)	(0.001)
N	104,160	49,561	19,488
Cancer group FEs	✓	✓	✓
Year FEs	✓	✓	✓
Health area FEs	✓	✓	✓

Notes: The Table presents coefficients estimates from equation (1)

Table 10: Relationship between number of articles and cancer deaths

	(1) Cancer deaths per pop.	(2) Cancer deaths per pop.	(3) Premature deaths per pop.	(4) Premature deaths per pop.
Panel a.				
Number of articles	-0.023***	-0.020***	-0.012***	-0.009***
	(0.002)	(0.002)	(0.001)	(0.001)
Panel b.				
asinh(Number of articles)	-0.443***	-	-0.252***	-
	(0.054)	-	(0.027)	-
N	23,280	8,206	23,280	8,206
m	4.572	4.251	2.280	2.031
Cancer group FEs	✓	✓	✓	✓
Health area FEs	✓	✓	✓	✓
Only scientific			✓	✓

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Data from the Person Level Integrated Data Asset (PLIDA). Presented results are based on model defined in (2).

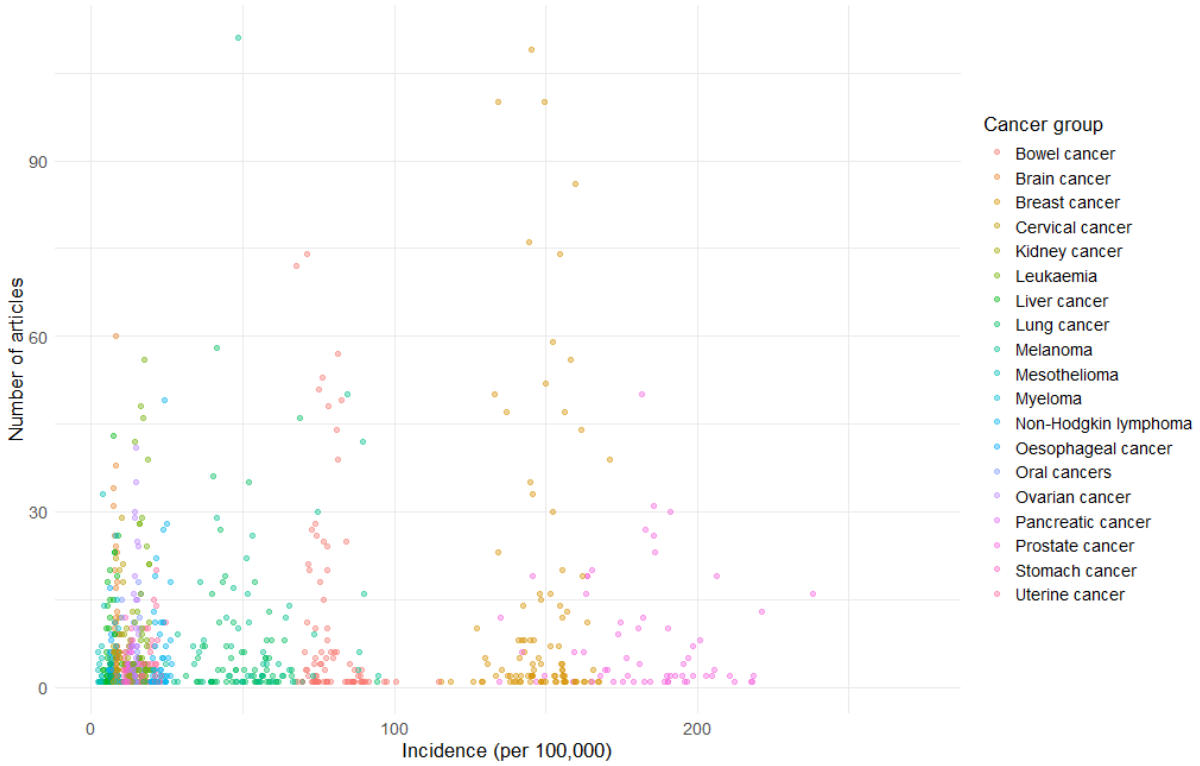


Figure 11: Incidence (per 100,000 population) and number of articles by PHA and cancer type in the pre-period (1998 - 2006)



Figure 12: Incidence (per 100,000 population) in the pre-period (1998 - 2006) and number of articles in the post-period (2007-2019) by PHA and cancer type

Table 11: SSIV results: scientists as IV

	(1) First stage	(2) Cancer deaths per pop.		(3) Premature deaths per pop.	
		OLS	IV	OLS	IV
Number of articles	0.006***	-0.023***	-0.098***	-0.012***	-0.059***
	(0.001)	(0.002)	(0.025)	(0.001)	(0.016)
N	23,280	23,280	23,280	23,280	23,280
Mean	1.213	4.572	4.572	2.280	2.280
F-stat	19.905	-	-	-	-
F-stat, weak ID	-	-	19.189	-	19.189
F-stat, AR	-	-	25.306	-	19.700
P-value, AR	-	-	0.0000	-	0.0000
Condition FEs	✓	✓	✓	✓	✓
PHA FEs	✓	✓	✓	✓	✓

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes:

Table 12: Relationship between number of articles and cancer deaths based on panel data

	(1) Cancer deaths per pop.	(2) Cancer deaths per pop.	(3) Cancer deaths per pop.
Panel a.			
Number of articles	-0.002***	-0.001***	-0.001**
	(0.000)	(0.000)	(0.000)
Panel b.			
asinh(Number of articles)	-0.443***	-	-
	(0.0083)	-	-
N	162,960	93,120	52,220
Cancer group FEs	✓	✓	✓
Health area FEs	✓	✓	✓
Year FEs	✓	✓	✓
Only scientific			✓
Post 2016		✓	

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes:

Table 13: Relationship between number of articles and premature deaths based on panel data

	(4) Premature deaths per pop.	(5) Premature deaths per pop.	(6) Premature deaths per pop.
Panel a.			
Number of articles	-0.004***	-0.002**	-0.003**
	(0.001)	(0.001)	(0.001)
Panel b.			
asinh(Number of articles)	-0.053***	-	-
	(0.016)		
N	162,960	93,120	52,220
Cancer group FEs	✓	✓	✓
Health area FEs	✓	✓	✓
Year FEs	✓	✓	✓
Only scientific			✓
Post 2016		✓	

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes:

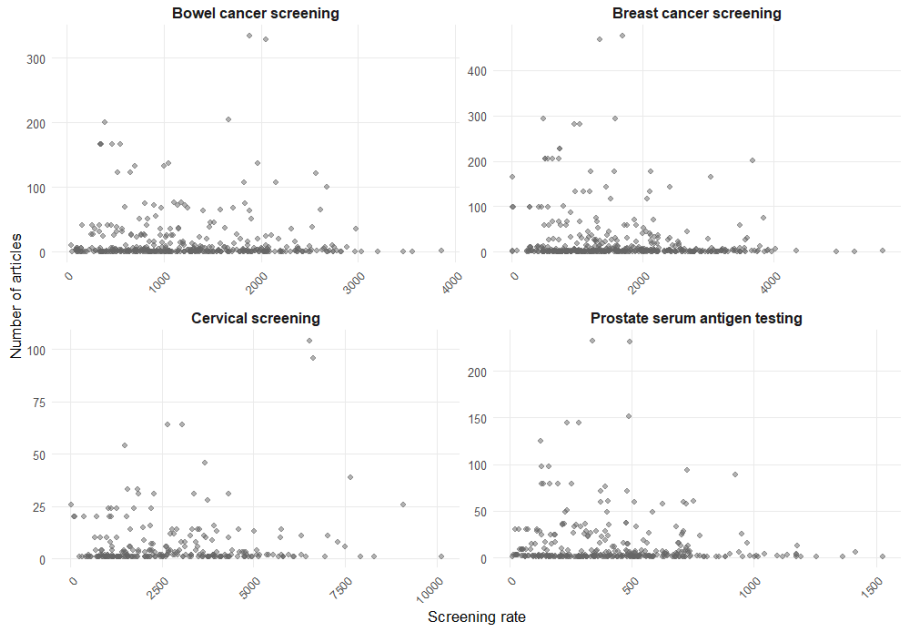


Figure 13: Screening estimates (2017-18 or 2020) and number of articles (2000-2016) in corresponding cancer groups