# Knowledge Growth and Specialization: Evidence from Oncologists*

Maya Lozinski
Massachusetts General Hospital
Harvard Medical School

October 2026

**Abstract**

This paper investigates how expert workers respond to knowledge growth. Medical oncology has seen rapid growth in knowledge: cancer treatment guidelines grew more than fivefold between 2002 and 2020. I estimate how oncologists' specialization responds to knowledge shocks using unexpected growth in the guidelines for the cancer types they treat. Less specialized oncologists did not become meaningfully more specialized, either over time or in response to knowledge shocks. In contrast, the most specialized oncologists became more specialized over time, but the estimates neither confirm nor rule out that knowledge shocks contributed to this rise. The highest levels of specialization are concentrated in large markets and large academic organizations. Median specialization among oncologists, however, is low and changes little with market or organization size outside academic medicine, consistent with limited returns to specialization for the typical oncologist. More specialized oncologists increasingly prescribe newer drugs and have more experience with each drug they prescribe than less specialized oncologists, suggesting a growing gap in technology adoption and quality of care. Specialization therefore appears under-rewarded, yielding limited private returns despite potential gains for patients.

*Keywords*: healthcare, innovation, oncology, scale economies, specialization, technology adoption
*JEL codes*: I1, J2, L2, O3

*The author gratefully acknowledges Zarek Brot-Goldberg, Joshua Gottlieb, David Meltzer, Dan Black, Tamara Konetzka, Pauline Mourot, Keith Ericson, and Jonathan Skinner for their valuable feedback and insightful comments. Special thanks are extended to Stephen Pardy for his technical assistance in digitizing the guideline citations. This study was supported by the NIH Medical Scientist Training Program (T32GM007281) and the NIA Program in Medicine, the Social Sciences, and Aging (T32AG051146). Author's email address: mlozinski@mgh.harvard.edu.

# 1 Introduction

Many disciplines, including computer science, molecular biology, and medicine, have experienced rapid expansion in their knowledge bases, accompanied by the proliferation of specialized subfields. Expert workers at the frontier of fields with significant innovation must contend with the growing challenge of remaining current. As knowledge expands, the costs associated with staying updated across numerous subfields increase substantially. One possible response is specialization—that is, narrowing one's scope of practice—which can mitigate the cognitive and informational burden. However, not all workers have the flexibility to specialize due to organizational or market constraints. In such cases, experts face a dilemma: either bear the growing costs of maintaining expertise across multiple domains or risk falling behind the frontier.

This paper investigates how expert workers respond to knowledge growth in the context of medical oncology. It develops novel empirical measures of knowledge growth and assesses how oncologists—specialist physicians who administer systemic cancer therapies such as chemotherapy and immunotherapy—adapt to this expansion. These clinicians are highly trained knowledge workers, requiring a minimum of 9 years of post-baccalaureate education in the United States. Some oncologists informally subspecialize and treat only select types of cancers while others remain in a "general" oncologic practice. For both types of oncologists, their practice centers on complex cognitive and communicative tasks, and they are primarily consumers, not producers, of scientific knowledge (Taylor, 2004; Jain et al., 2019; Schlafly and Sebro, 2022).

Despite these significant initial investments in human capital, oncologists must learn continually to remain at the forefront of their rapidly evolving field. Cancer affects a substantial portion of the population: the lifetime risk of developing cancer is approximately 40% (American Cancer Society, 2024). This disease burden fuels innovation and rapid knowledge growth. One article in a popular science magazine observed that "incorporating this torrential evolution of knowledge is an impossible, Sisyphean task" (Wallis, 2022).

An oncologist interviewed in the course of this research remarked, "Some fields of cancer treatment are as complex now as the whole of oncology was 20 years ago." The oncology setting, therefore, provides a rich context in which to measure rapid knowledge accumulation and examine the consequences, with implications for clinical care and patient health.

To evaluate the relationship between knowledge growth and specialization empirically, I construct measures based on changes in the National Comprehensive Cancer Network Clinical Practice Guidelines in Oncology (NCCN Guidelines®), which constitute the most authoritative source of clinical guidance for cancer management in the United States. Since 2001, these guidelines have offered a comprehensive framework for screening, diagnosis, and treatment by cancer type (NCCN, 2023*b*). In this paper, changes in guideline length and citation counts proxy for the expansion of clinically relevant knowledge. A principal advantage of these measures is that they reflect the operational body of knowledge that oncologists must engage with in their daily practice.

Section 4 provides evidence of substantial growth in oncology-related knowledge. Between 2002 and 2020, the total number of pages in NCCN screening and treatment guidelines increased from 1,105 to 5,581 pages. This expansion corresponds closely with an increase in guideline citations to the biomedical literature, the vast majority of which do not overlap across cancer types. It also coincides with substantial growth in the number of FDA-approved cancer therapies, with the cumulative count rising from 90 in 2002 to 243 in 2020.

To examine how oncologists respond to this knowledge growth, I measure annual specialization at the physician level using a 21-year panel of Medicare claims. Specialization is measured using the Herfindahl–Hirschman Index (HHI) over the time an oncologist spends managing different cancer types, so a higher HHI reflects a narrower scope of practice. The median and the highest levels of specialization have moved in different directions over time. Highly specialized oncologists have become increasingly specialized. In contrast, the median oncologist has become, if anything, more general over time and has treated a mix of cancers about as diverse as the cancers in the overall patient population.

To explore the causal relationship between knowledge growth and specialization, I construct a physician-level measure of exposure to knowledge growth. The measure is the cumulative increase in guideline length across the cancer types treated by the oncologist in a baseline year. For example, if a physician treats kidney and lung cancer, and each guideline expands by 20 pages, the physician is exposed to 40 pages of knowledge growth. This metric combines two sources of variation: variation in the sets of cancer types that physicians treat and in the rates of knowledge growth across cancer types.

However, this measure of exposure to knowledge growth may suffer from omitted variable bias, as physicians' choice of cancer types to treat may not be random. To address this, I apply a recentering procedure adapted from Borusyak and Hull (2023). The procedure removes predictable variation in exposure to knowledge growth based on the number and characteristics of the cancer types a physician treats. The residual variation is interpreted as exposure to "unpredictable" knowledge shocks (e.g., success in clinical trials or discovery of novel molecular targets). The procedure allows physicians' choice of cancer types to be non-random without threatening identification, as long as the cancer-level knowledge shocks are unrelated to other factors that drive specialization. Reassuringly, this procedure also removes the joint correlation between exposure to knowledge shocks and observable physician confounders such as market size and organization size.

I examine how oncologists respond when exposed to unexpected knowledge shocks. This is done separately by initial level of specialization because the median and the highest levels of specialization moved in different directions over time. I find that oncologists who start out more general specialize little or no more when exposed to larger knowledge shocks. The estimated effect of a one standard deviation rise in exposure to knowledge growth is an increase in HHI of 0.03 standard deviations, and the data rule out increases larger than 0.09 standard deviations. More general oncologists also do not meaningfully narrow the set of cancer types they treat: the data rule out dropping even one cancer type for a two standard deviation rise in exposure to knowledge growth. In contrast, for oncologists who start out

more specialized, the data neither confirm nor rule out a meaningful specialization response to knowledge growth. Although the estimated effect of knowledge shocks is small (0.03 standard deviations) and not statistically distinguishable from zero, the data cannot rule out increases in HHI as large as 0.22 standard deviations for a one standard deviation rise in exposure to knowledge growth.

More general oncologists, who typically work outside academic organizations, appear to face persistent barriers to specialization that are not reduced by larger organizations or markets. Overall, median specialization remains nearly flat with respect to market size. In non-academic settings, specialization also does not increase with organization size. Price incentives for producing high-quality care are weak: during the study period, Medicare paid only for quantity of care, not quality. Specializing may do little to attract additional patients: most patients do not see an oncologist who specializes in their cancer type, even when one practices nearby. These findings suggest that, for the majority of oncologists, the returns to specialization are low or the costs prohibitively high.

In contrast, specialization is concentrated among oncologists in large academic organizations (academic medical centers and their affiliates) and in large markets. Within academic organizations, specialization rises steeply with organization size. Moreover, the highest levels of specialization within a market scale with market size, and this relationship has steepened over time. Consequently, geographic disparities in access to highly specialized care have grown, implying increasing market-level economies of scale in oncology. Outside the largest markets, most cancer types have too few patients to support a fully specialized oncologist.

Given the growing dispersion in specialization, an important question arises: does specialization affect patient care? Analysis of prescribing behavior indicates that it does. Over time, the most specialized oncologists have more experience with each drug they prescribe than less specialized oncologists. Furthermore, less specialized oncologists have lagged in adopting newer cancer therapies, including breakthrough drugs. The differences in drug experience and drug age have widened over time, suggesting that specialization is increasingly associated

with higher-quality care.

This paper contributes to the economics literature on innovation, healthcare, and specialization. The paper offers a novel measure of knowledge growth based on guideline evolution, a more comprehensive and clinically relevant alternative to counts of drug approvals or patents. Recent medical literature has documented that NCCN Guidelines for four common cancers have grown longer and include more citations over time (Kann et al., 2020). I extend this line of inquiry by quantifying the growth in content across all cancer types and documenting the turnover in citations over time. In the economics literature, drug approvals and patents have proxied for innovation (Glied and Lleras-Muney, 2008; Alcacer and Gittelman, 2006; Roach and Cohen, 2013). In contrast to these measures, clinical guidelines reflect the breadth of innovation and knowledge growth in oncology, encompassing not only new drug approvals but also off-label uses, novel drug combinations, biomarkers, delivery strategies, and more. In addition, guidelines are curated to include only clinically relevant knowledge, excluding outdated information, such as drugs no longer in use.

This paper also studies the accumulation of many technologies, which has received less attention in the empirical literature on technological change in health care. Prior research has typically focused on the adoption of specific technologies (e.g., Skinner, 2011; Chandra et al., 2014; Skinner and Staiger, 2015; Arrow et al., 2020). This paper uses guideline growth to measure accumulation in the knowledge base and set of treatments that oncologists draw on.

Notably, this paper examines a question similar to that posed by the innovation literature, namely how the growth of knowledge influences specialization, but arrives at different conclusions. Jones (2009) and Ganguli et al. (2026) find that inventors have become more specialized and differentiated as the depth of knowledge has increased. In contrast, most oncologists remain general and do not specialize over time or in response to knowledge growth. One explanation may be that inventors are primarily creators of new knowledge, whereas most oncologists are primarily users of existing knowledge. Users of knowledge may experience the burden of knowledge more lightly than creators, as applying existing knowledge is generally

less demanding than producing it. Furthermore, this paper finds that knowledge users may opt to lag behind the technological frontier when the cost of remaining up to date becomes prohibitively high. Consequently, the implications of knowledge growth for the many workers who use knowledge may differ from previous findings on workers who create knowledge.

This paper also speaks to a growing economics literature on what makes decisions complex and how complexity changes choices (Gabaix and Graeber, 2024; Agranov, Schotter, and Trevino, 2025). The evidence in this literature comes from theory and laboratory experiments. This paper brings similar questions about decision complexity to the field. Agranov, Schotter, and Trevino (2025) show that the same problem can be easy for some people and too hard for others. Gabaix and Graeber (2024) show that people facing more complex decisions adjust their choices less when the best choice changes. Specialization can be viewed as a way for oncologists to make their decisions easier and less complex, by reducing the scope of problems they encounter regularly. Consistent with this prediction, less specialized oncologists adopt newer therapies more slowly than more specialized oncologists.

This paper also contributes to the literature on labor specialization. Consistent with prior work, this paper finds that, even within a tightly defined occupation, the highest levels of specialization rise with market size (Baumgardner, 1988; Garicano and Hubbard, 2009; Atalay et al., 2024). The paper extends this cross-sectional evidence by showing that specialization in large markets has risen over time. Subsequent work in the medical literature has confirmed the rise in oncologist specialization over time and in large markets documented here (Karadakic et al., 2025). These findings on market size align with the substantial market-level economies of scale in health care identified by Dingel et al. (2023) and suggest that these economies of scale are growing over time.

In addition, this paper finds that physician specialization is associated with markers of higher-quality care, in contrast to the findings of Baicker and Chandra (2004). The difference may partly reflect the periods studied, as Baicker and Chandra (2004) examine evidence from 1995 and 1996, while this paper uses data from 1999 to 2019 and finds that the extent and

the realized benefits of specialization appear to have increased over time. Recent work by Karadakic et al. (2026) builds on this paper's evidence that specialization affects patient care, estimating that access to an oncologist who specializes in the patient's cancer type reduces three-year mortality by about 10%. Together with the growing differences in specialization by market size, these findings suggest that cancer outcomes may depend increasingly on where patients live and on which oncologists they see.

# 2 Conceptual Framework

This section presents institutional context and a theoretical framework. It first characterizes medical oncologists, who treat cancer primarily with pharmaceutical interventions. It then discusses how they can specialize horizontally (by cancer type) and vertically (within knowledge hierarchies). As clinical guidelines become more complex, the cost of acquiring knowledge may increase and push oncologists toward a narrower scope. It next explains how knowledge growth may widen differences in specialization across markets of different sizes. Finally, it explores how frictions in referrals and matching may limit specialization.

This research focuses on medical oncologists, who differ from other cancer care providers on two main dimensions: disease focus and treatment modality. In terms of disease, medical oncologists specialize in the treatment of cancer. In terms of modality, medical oncologists primarily use drug-based therapies, unlike surgical oncologists or radiation oncologists. Their training pathways also differ substantially.[1] Thus, medical oncology is a distinct specialty. For simplicity, the term "oncologist" in this study refers to medical oncologists.

Medical oncologists may specialize by cancer type, for example, focusing exclusively on breast cancer. Such specialization reduces redundant learning, as each oncologist needs to stay updated only on advancements within their narrow domain. This focus also facilitates

[1] In the United States, surgical and medical oncologists both complete 4 years of medical school. Surgical oncologists first specialize in surgery in a specific anatomical region and then in cancer surgery. In contrast, medical oncologists complete 3 years of internal medicine residency followed by at least 2 years of fellowship in medical oncology.

"learning by doing," enabling, for instance, a breast cancer specialist to become particularly adept at managing complex and potentially toxic drug regimens.

In classic economic theory, this specialization is considered productive and is generally constrained by the size of the market (Smith, 1776; Kim, 1989). Consequently, specialization is expected to increase with market size.

However, Becker and Murphy (1992) argue that the division of labor distributes responsibilities across a greater number of workers, thereby incurring coordination costs. For example, handing off a patient with a complex cancer history costs the receiving oncologist the time to learn that history. The optimal level of specialization is determined by weighing its productivity benefits against these coordination costs. In some cases, these costs may limit specialization, so that specialization rises little with market size. Nevertheless, knowledge growth increases the benefits of specialization, thereby raising specialization in equilibrium.

The organization of labor within medical oncology can also be examined through vertical knowledge hierarchies, as described by Garicano (2000) and Garicano and Santos (2004). All oncologists are formally licensed to manage a broad range of cancers. However, informal knowledge hierarchies exist. In theory, general oncologists manage routine cases, while more complex cases are referred to subspecialists with deeper expertise.

The depth of these knowledge hierarchies is shaped by the cost of acquiring knowledge. Clinical guidelines affect these costs. Initially, the introduction of guidelines lowers these costs by making relevant knowledge easier to find. However, as guidelines become more complex, the cost of extracting and applying relevant knowledge rises. These guidelines function as summaries of external expertise, and their length proxies for the volume of underlying knowledge. As the cost of acquiring knowledge for a given clinical area rises, oncologists may specialize further, thereby deepening vertical knowledge hierarchies.

Knowledge growth may also widen differences in specialization across market size. For instance, a breast cancer specialist needs enough breast cancer patients to support a practice. As the complexity of breast cancer care increases, this specialist may further narrow her

focus to subtypes of breast cancer. However, smaller markets may lack the patient volume necessary to support such a narrow practice, so knowledge growth may widen the gap in specialization between small and large markets.

Even in markets large enough to support specialists, frictions in referrals and matching may limit the patient volume specialists attract, and with it the returns to specialization. Anti-kickback regulations prohibit compensation for referrals, meaning physicians are not reimbursed for any lost earnings when referring patients to other providers. Anecdotal evidence suggests that organizations are often reluctant to refer patients externally, for financial reasons. These frictions can result in under-referral, consistent with Garicano and Santos (2004), who argue that in a spot market with zero-priced referrals, equilibrium referral falls short of the social optimum. Organizations can potentially address this inefficiency by formalizing referral and compensation arrangements in advance (Garicano and Santos, 2004). Patients face a different friction: search costs. It can be difficult for a patient to determine which oncologists are experts in their cancer type.

# 3 Data and Measurement

The empirical strategy integrates cancer treatment guidelines, Medicare claims, and Medicare provider data to examine trends in knowledge growth and specialization within medical oncology. Section 3.1 explains how knowledge growth is proxied through changes in page length and citation counts in cancer treatment guidelines. Section 3.2 describes the Medicare provider and claims data. Section 3.3 explains how cancer guidelines and Medicare data are combined to generate physician-level estimates of exposure to knowledge growth across a multi-year panel. Section 3.4 explains how physician specialization is assessed using Medicare data. Appendices B.1–B.4 provide more detail on data construction, and Appendix B.5 describes how the code was developed and checked.

## 3.1 Cancer Guidelines Data

To measure knowledge growth, I constructed a dataset based on historical cancer treatment guidelines issued by the National Comprehensive Cancer Network (NCCN). The NCCN is an alliance of 33 leading cancer centers (NCCN, 2023*a*). Since 1996, the NCCN has published cancer treatment guidelines (NCCN, 2023*b*). The guidelines span the full spectrum of cancer care, including "diagnosis, imaging, pharmacotherapy, radiation, and surgical interventions" (McGivney, 2008). These guidelines are widely used: in one survey, 96% of oncologists reported that the NCCN Guidelines are important for treatment decisions (McGivney, 2008).

Each set of NCCN Guidelines reflects a consensus of expert opinion and is overseen by a panel of about 20 specialists (NCCN, 2011). These guidelines are updated at least annually to incorporate the latest evidence and therapies. Substantive updates require approval by a quorum of panel members. In practice, the majority of recommendations are approved by more than 85% of panel members (NCCN, 2024).

I use changes in NCCN Guideline page length and citation counts as proxies for the expanding volume of clinically relevant information. Page lengths and citation counts were measured by cancer type from each year's first version of the NCCN Guidelines for cancer screening, treatment, and risk reduction. As a robustness check, I use changes in two citation counts, built as described in Appendix B.4: citations of any type, and citations to scholarly works identifiable in OpenAlex, a database of research literature (Priem, Piwowar, and Orr, 2022). Both track the volume of scientific literature that underpins the guidelines.

The volume of relevant knowledge measured in this way is distinct from the advancement of treatment technologies. The guidelines drop outdated content, so they measure the current stock of relevant clinical knowledge rather than its historical accumulation. In theory, if future treatments cure all cases with minimal side effects, guidelines could be shorter. In practice, this scenario is rare: some patients still present with cancers that are incurable or progress after first-line treatments. Consequently, later lines of therapy build up over time, contributing to longer guidelines.

Guideline expansion has also been driven by increasing capacity to subtype cancers, tailor targeted treatments accordingly, and expand the therapeutic options within each subtype. For instance, lymphomas, cancers of the lymphatic system, were once classified as either "Hodgkin lymphoma" or "non-Hodgkin lymphoma," a catch-all category for every other lymphoma (Armitage, 2010). Over time, this catch-all category has come to encompass a growing number of biologically distinct subtypes, each requiring unique treatment approaches (Jaffe, 2009). These forces make the treatment decisions the guidelines describe more complex in the sense of Gabaix and Graeber (2024): more considerations bear on each decision.

Figures A.1 and A.2 provide examples of how the NCCN Guidelines changed between 2009 and 2014. The table of contents for non-Hodgkin lymphoma became longer and incorporated new subtypes. The diagnostic workup flowchart for lobular carcinoma in situ, a breast cancer subtype, added a new diagnostic step and introduced more decision splits.

Several assumptions are necessary for changes in page length to serve as a valid proxy for changes in the volume of clinically relevant knowledge. First, this measure assumes that each page corresponds to a comparable amount of clinical content. Reassuringly, the formatting of the NCCN Guidelines remained consistent across cancer types throughout the core study period (2008–2014). This consistency supports the interpretation that increases in page length reflect added content rather than formatting changes. If each page holds a comparable amount of clinical content, a 75-page increase confronts an oncologist with the same amount of new content whether the baseline guideline had 35 pages or 200.

Second, the measure assumes that relevant information not captured in the NCCN Guidelines is either proportional to the information included or remains constant across cancer types. For instance, while the guidelines may not capture every detail about a chemotherapy drug, they are assumed to capture a similar share of the detail for every treatment. Likewise, the NCCN Guidelines rely on foundational knowledge in internal medicine and cancer care. Even if the volume of general medical knowledge changes over time, this assumption holds as long as those changes affect all cancer types equally.

Finally, growth in page count is a net measure of knowledge growth. It assumes that knowledge is primarily added rather than replaced. For example, if a 20-page guideline grows to 30 pages, this is interpreted as 10 pages of new knowledge. If, however, 10 pages are removed and replaced with 20 pages of new content, the net measure would understate total knowledge growth. This assumption is evaluated empirically in Section 4.

## 3.2 Medicare Data

Oncologists' clinical activity is observed in a 20% sample of Medicare Part B claims from 1999 to 2019. Additional physician characteristics are obtained from the Medicare Data on Provider Practice and Specialty (MD-PPAS) dataset for 2008–2014. Medicare provides health insurance to disabled adults and most U.S. adults aged 65 and older. This age group has the highest incidence of cancer, making it an ideal population for studying cancer care (Centers for Disease Control and Prevention, 2019).

From these data, I construct an annual panel of oncologists. The sample is restricted to physicians whose most frequently listed specialty in a given year is Hematology, Medical Oncology, or Hematology–Oncology. I limit the sample to physicians observed to treat at least 25 unique cancer patients per year. This threshold protects patient privacy and limits measurement error.

For the primary regression analysis, I focus on a balanced panel from 2008 to 2014. The panel begins in 2008, the first year with many physician characteristics available in the data. It ends in 2014 because the claims data switch from ICD-9 to ICD-10 diagnosis codes in mid-2015. The switch introduces a discontinuity in measured specialization, as described in Section 4.3.

For 2008–2014, the MD-PPAS provides each physician's sex, birth year, and the top tax identification (ID) number under which they bill. Organization size is proxied by the number of oncologists billing under a tax ID. One limitation of this measure is that the number of oncologists linked to a tax ID may understate the size of the organization's medical oncology

practice, as detailed in Appendix B.3. Tax ID names are also used to identify academic organizations, meaning academic medical centers and their affiliates. The resulting measure of academic status is a broadly inclusive proxy rather than a definitive measure, as detailed in Appendix B.3.

Next, I measure the time each oncologist devotes to each cancer type. The analysis focuses on outpatient evaluation and management claim lines billed by oncologists. These visits involve high-level medical decision-making. Each evaluation and management code is mapped to work Relative Value Units (wRVUs), which for these services reflect the physician time Medicare expects the visit to require.

Each visit is assigned to a cancer type based on the patient's primary cancer diagnosis. The primary cancer keeps its name when it metastasizes, or spreads from where it started to a distant part of the body (NCI, 2026). NCCN Guidelines are organized by cancer type on the same basis. The guideline for lung cancer, for example, covers cancers originating in the lung wherever they have spread. Appendix B.2 provides details about how each patient is assigned a primary cancer.

An advantage of this setting is that cancer diagnoses are standardized and determined independently of an oncologist's specialization. A pathologist studies the patient's cells and tissues and "provides the definitive cancer diagnosis" (NCI, 2022). This limits concerns that the oncologist's focus might influence the type of cancer diagnosed.

## 3.3 Measuring Relevant Knowledge Growth for Each Oncologist

I construct the main treatment variable: the growth in relevant knowledge experienced by each oncologist. For each oncologist $i$, let $C_i$ denote the set of cancer types they manage in outpatient settings in 2008, the baseline year. For each cancer type $c$, define $\Delta k_c$ as the change in the number of guideline pages between 2009 and 2014. I assume that the full guidelines and their updates for cancer $c$ are relevant to all oncologists who treated cancer $c$ in 2008. Under this assumption, oncologist $i$'s exposure to knowledge growth is $\Delta k_i = \sum_{C_i} \Delta k_c$.

Suppose oncologist $i$ treated only lung and breast cancer in 2008. Suppose also that between 2009 and 2014 the lung cancer guidelines expanded by 50 pages and the breast cancer guidelines by 100 pages. Oncologist $i$ is then exposed to 150 pages of knowledge growth.

To link cancer types in the claims data to guideline categories, I construct custom groupings of diagnosis codes. These custom groupings follow the organization of the NCCN Guidelines. Additional details are provided in Appendix B.1.

## 3.4 Measuring Physician Specialization

The main outcome is oncologist specialization, measured using the Herfindahl–Hirschman Index (HHI) based on the time an oncologist spends managing different cancer types. Let $\tau_{ic}$ be the share of oncologist $i$'s wRVU-weighted evaluation and management time spent managing cancer type $c$. Then $\text{HHI}_i = 10{,}000 \sum_c \tau_{ic}^2$.

The HHI ranges from 0 to 10,000. A value of 10,000 indicates the oncologist treats only one cancer type, the most specialization the data can detect. If time is split equally across ten cancer types, the HHI is 1,000. As the count of cancer types increases and time is more evenly distributed, the HHI approaches 0.

This measure has several advantages. First, it is continuous and grounded in actual clinical activity. Second, it captures changes on the intensive margin of specialization. For example, it detects if an oncologist begins focusing more heavily on a specific cancer type without completely discontinuing care for other patients. Unlike measures based on formal titles or specialty designations, the HHI therefore tracks changes within a physician over time and differences across physicians who share the same title.

The HHI captures horizontal specialization across cancer types. The claims data record a cancer's anatomical site of origin but not how far it has spread, how complex the case is, or other clinical features. The complexity of the cases an oncologist sees is therefore unobserved, so vertical specialization, depth of expertise within a cancer type, cannot be measured directly. However, a narrower horizontal focus typically allows deeper vertical expertise.

The HHI is estimated with sampling error. The claims cover a 20% sample of Medicare patients, so each oncologist's observed patients are a sample of those she treats. The time shares observed in the claims, $\hat{t}_{ic}$, are therefore sample estimates of the true shares $\tau_{ic}$, and the index computed from them, $\widehat{\text{HHI}}_i = 10{,}000 \sum_c \hat{t}_{ic}^2$, is a sample estimate of $\text{HHI}_i$. The rest of this subsection draws on Appendix C, which shows how this sampling error biases the HHI, corrects the bias with a jackknife, and tests the correction.

This sampling error does not average to zero but inflates measured specialization systematically, creating an upward finite-sample bias. The size of this finite-sample bias depends on the patient count as well as the true time shares. At the extreme, an oncologist observed to have a single patient mechanically receives an estimated HHI of 10,000, regardless of her true time shares. The HHI is estimated without sampling error for an oncologist who treats only one cancer type but is biased upward for one whose time is split across multiple cancer types.

This sampling error is not classical measurement error. Its mean depends on the patient count and on the true HHI, both of which may be correlated with an oncologist's exposure to knowledge growth, so it could bias the estimated effect rather than only add noise.

To reduce this finite-sample bias, the HHIs reported in this paper, written $\widehat{\text{HHI}}_i^{\text{jack}}$, are jackknife corrected (Cameron and Trivedi, 2005). The jackknife removes the part of the bias that shrinks with the patient count, leaving a smaller remainder that shrinks with the square of the patient count.

# 4 Descriptive Patterns

This section examines the expansion of clinical knowledge and specialization among oncologists over time. Section 4.1 shows that NCCN guidelines grew more than fivefold between 2002 and 2020. Section 4.2 presents summary statistics about oncologists. Section 4.3 shows that, between 1999 and 2019, the most specialized oncologists specialized further, while the median oncologist did not. Section 4.4 shows that, in 2008, the most specialized oncologists were

also more often academic, in large markets, and in large organizations.

## 4.1 Knowledge Growth Over Time

Cancer treatment guidelines have grown substantially longer since 2002. Figure 1a illustrates the lengths of these guidelines for selected cancer types, and Table A.1 reports page counts across all cancer types. Total guideline length grew more than fivefold, from 1,105 pages in 2002 to 5,581 pages in 2020.

One contributor to the increase in guideline length appears to be the rapid expansion in the number of anti-cancer drugs approved by the U.S. Food and Drug Administration (FDA). Figure A.3 shows the cumulative count, which grew from 90 in 2002 to 243 in 2020.

A growing body of scientific literature is another contributor to the expansion of guidelines. Figure 1b shows that the total number of citations for colon, kidney, lung, and skin cancer rose roughly tenfold between 2002 and 2020. Guideline length rose with the number of citations, across cancer types and over time (Table A.2). Between 2009 and 2014, each additional citation corresponded to an increase of about one-fifth of a page ($R^2 = 94\%$). Page growth thus closely tracked citation growth.

The majority of references are to published scholarly works. Of the 17,320 citations in the 2009 and 2014 guidelines, 87% can be linked to scholarly works. The remaining 13% of citations are to drug package labels, conference posters, professional society reports, government documents, and papers that cannot be reliably linked.

Between 2009 and 2014, the guidelines mostly added scholarly works rather than replacing them. The 2009 guidelines cited 3,574 unique scholarly works, with 2,506 (70%) appearing in the 2014 guidelines (Figure A.4). The other 1,068 (30%) were dropped. Additionally, 7,155 scholarly works were added, several times the number dropped. These changes suggest that clinical knowledge expands primarily by addition, so that net knowledge growth is close to gross knowledge growth.

Citation growth between 2009 and 2014 came primarily from adding references to new

and recent scholarly works. Figure A.5 groups the references added, retained, and removed by the year the cited work was published. Most added references were published shortly before or after 2009. Of the added references, 49% were new (published in or after 2009), 30% were recent (published within 5 years prior to 2009), and 21% were older (published more than 5 years prior). The ages of the added references suggest that guideline growth largely reflects newly published research rather than the late addition of older literature.

Measured by the scholarly works that guidelines cite, knowledge overlaps little across cancer types. In both 2009 and 2014, 90% of scholarly works were referenced by only one cancer type's guidelines. The scientific literature behind each cancer type's guideline thus appears largely distinct.

A limitation of guideline length is that some cancer types still had no guideline in 2009, although the NCCN says it had released a comprehensive library of guidelines in 2001 (NCCN, 2023*b*). For example, mesothelioma, a rare cancer, had no guideline until 2010. Evidence on how to treat the disease was very limited in 2010 (Mulcahy, 2010), likely delaying its first guideline. Mesothelioma was nonetheless being treated before 2010. Zero guideline pages therefore do not mean zero knowledge about how to treat a cancer type.

The empirical strategy does not assume that zero-page guidelines indicate zero knowledge, only that more pages correlate with more knowledge. However, when a first guideline appears, its entire length counts as page growth, which may overstate knowledge growth for that cancer type. In 2009, only 3.95% of patients with a specific cancer diagnosis had a cancer type with no guideline, so the potential overstatement affects a small share of patients. Appendix E reports a robustness check that treats a guideline introduced during the study period as documentation of knowledge that already existed rather than as knowledge growth.

## 4.2 Oncologist Characteristics

Table 1 presents summary statistics for the balanced panel of oncologists used in the regressions of Section 6. The panel includes the 3,425 oncologists who saw at least 25 unique cancer

patients in each of 2008, 2009, and 2014. In 2008, each oncologist treated an average of 57 unique cancer patients in the outpatient setting. The average oncologist treated 14 different cancer types, though this number ranged from as few as one to as many as 25 types. Oncologists worked in organizations of widely varying size, with an average of 18 oncologists. The smallest organizations had a single oncologist, and the largest had 219.

The guideline pages relevant to the average oncologist doubled between 2009 and 2014, matching the growth of the guidelines as a whole. In 2009, the NCCN Guidelines totaled 1,832 pages. This total doubled to 3,589 pages by 2014. The guideline pages for the cancers each oncologist treated in 2008 also doubled, from an average of 1,121 pages per oncologist in 2009 to 2,187 pages by 2014.

## 4.3 Oncologist Specialization

HHI had a mean of 1,474 in 2009. Some oncologists approached the maximum possible value of 10,000 (max = 9,886), so they managed almost exclusively one type of cancer over a year. The HHIs reported in this paper are jackknife-corrected. Across the sample, the jackknife correction lowers measured HHI by 366 points, or 18.9%, relative to the uncorrected value. The jackknife correction leaves the ordering of oncologists largely unchanged, with a rank correlation of 0.931 between corrected and uncorrected values.

In a simulation calibrated to the claims data, the jackknife correction removes most of the finite-sample bias in the uncorrected HHI (Appendix C). The uncorrected HHI of a simulated general oncologist, the case with the most bias, exceeds the true HHI on average by 68% at 25 patients and 35% at 50 patients. The same oncologist's jackknife-corrected HHI exceeds the true HHI on average by only 4.8% at 25 patients and 1.2% at 50 patients. The jackknife correction reduces bias at the cost of some variance, raising the standard deviation of the general oncologist's estimated HHI at 25 patients by 12%.

The small finite-sample bias that remains depends on the patient count and on the true HHI. It could bias the estimated effect of knowledge growth in Section 6 if either predicts

exposure to knowledge growth. This upward bias is present in both the initial and the final HHI. It therefore partially cancels out in the main estimates, which use each oncologist's change in HHI.

Between 1999 and 2019, the most specialized oncologists in the distribution specialized further, while the median oncologist did not. Figure 2 shows the 10th, 25th, 50th, 75th, and 90th percentiles of HHI in each year, with 95% confidence intervals from bootstrapping oncologists. Each year's sample includes every oncologist who saw at least 25 unique cancer patients that year. The figure also shows the population-level HHI, the HHI of all cancer patients in the sample pooled together. An oncologist whose HHI equals the population-level HHI treats a mix of cancers as diverse as the cancers in the overall patient population.

Despite the rapid expansion of clinical knowledge and treatment options, the typical oncologist has not become more specialized over time. The median and lower percentiles of the HHI distribution remained flat or declined between 1999 and 2019. Throughout those years, the 10th, 25th, and 50th percentiles remain within a 600-point HHI range that includes the population-level HHI. The typical oncologist thus appears to have remained general, treating a mix of cancers about as diverse as the cancers in the overall patient population.

In contrast, oncologists at the upper end of the HHI distribution have become far more specialized. The figure omits 2015, when the transition from ICD-9 to ICD-10 diagnosis codes introduced a discontinuity in measured HHI. From 1999 to 2014, the 90th percentile of HHI rose by 1,483 [1,276, 1,881] points (95% confidence interval in brackets). Between 2016 and 2019, it rose by a further 309 [−72, 630] points. The same trends appear when specialization is measured by shares of patients rather than shares of time (Figure A.6).

These divergent trends suggest that more and less specialized oncologists may respond differently to the expansion of medical knowledge. Section 6 tests this hypothesis by splitting the panel into generalist and specialist oncologists, defined in Section 4.4, and estimating their responses to knowledge growth separately. Although the split distinguishes more from less specialized oncologists, all oncologists in the sample are medical subspecialists. The

analysis therefore uses variation in specialization within an already specialized group.

## 4.4 Differences Between More and Less Specialized Oncologists

Table 1 summarizes the baseline characteristics of generalist and specialist oncologists in 2008, the first year of the balanced panel. Generalists are the oncologists below the 75th percentile of the 2008 HHI distribution, and specialists are those at or above it.

The two groups differed substantially in specialization and academic affiliation. The average HHI was 993 for generalists and 2,865 for specialists, nearly three times as high. The generalist average was equivalent to splitting time equally across about ten cancer types, and the specialist average to splitting it across three or four cancer types. HHI was also far more dispersed among specialists than among generalists, a difference that influences the precision of the estimates for specialists in Section 6. Only 5% of generalists were academic, compared with 22% of specialists.

Furthermore, specialists practiced in larger markets and organizations. In 2008, 59% of specialists practiced in the largest 20% of markets, ranked by hospital referral region population, compared with 46% of generalists. Specialists' organizations were also larger, with 23 oncologists on average versus 16 for generalists. Large markets and organizations thus appear to facilitate specialization, a question Section 7 takes up in more depth.

Notably, generalists and specialists did not differ meaningfully in age, averaging 49 and 50 years old. Oncologists thus do not appear to evolve systematically toward a more general or specialized practice as they age.

Specialists and generalists differ slightly in the cancers they treat (Figure A.7). Breast cancer is the largest difference. In 2008 breast cancer accounted for 31% of specialists' time and 18% of generalists' time. For every other cancer type, the two groups' shares of time differ by 3 percentage points or less.

Despite similarities in age, the observable and likely unobservable differences between specialists and generalists indicate that they are distinct groups. These baseline differences

motivate estimating responses to knowledge growth separately for the two groups.

# 5 Identification

This section presents the strategy for estimating the effect of knowledge growth on specialization. Section 5.1 explains how changes in specialization are regressed on exposure to knowledge growth at the physician level. Section 5.2 states the identifying assumptions and the threats to them. Section 5.3 explains how the exposure measure is recentered to remove predicted, or non-random, exposure to knowledge growth. This adjustment follows the general approach outlined in Borusyak and Hull (2023). Section 5.4 explains why randomization inference is needed for inference and how it is implemented. Section 5.5 shows that recentering and randomization inference largely resolve concerns that observable physician characteristics predict exposure to knowledge growth.

## 5.1 Estimating Equation

Let $\Delta\widehat{\text{HHI}}_i^{\text{jack}}$ denote physician $i$'s change in specialization between 2009 and 2014, measured by the jackknife-corrected HHI. Let $C_i$ be the set of cancers treated by physician $i$ in 2008, and let $\Delta k_c$ be the change in page length of cancer type $c$'s guideline between 2009 and 2014. Physician $i$'s exposure to knowledge growth sums the page growth over the cancers they treat, $\Delta k_i = \sum_{C_i} \Delta k_c$. Let $X_i$ denote physician characteristics, entered in some versions of the estimating equation as fixed effects for age group, sex, hospital referral region, decile of patient count, organization size, and academic affiliation, all measured in 2008.

The estimating equation is

$$\Delta\widehat{\text{HHI}}_i^{\text{jack}} = \alpha + \beta \Delta k_i + \theta X_i + \epsilon_i.$$

The equation is estimated using ordinary least squares. Confidence intervals come from both heteroskedasticity-robust standard errors and randomization inference (Section 5.4).

The coefficient of interest is $\beta$, which represents the effect of knowledge growth on specialization. $\beta$ is estimated using variation in exposure to knowledge growth ($\Delta k_i$), which arises from two sources. The first source is that physicians work in different sets of cancer subfields. For instance, some physicians treat breast cancer, others kidney cancer, and some both. The second source is that different cancer subfields experience varying levels of knowledge growth. For example, the breast cancer guideline grew more than the kidney cancer guideline, so physicians who treated breast cancer exclusively were more exposed to knowledge growth than those who treated kidney cancer exclusively. $\beta$ also uses within-physician variation in specialization over time. If specialization responds with a lag, the estimate of $\beta$ is attenuated.

The intercept $\alpha$ absorbs any global shocks to knowledge. It includes the impact of knowledge growth that affects all cancer fields or the general practice of medicine. Consequently, $\beta$ captures the effect of knowledge growth within different cancer subfields. The intercept $\alpha$ also accounts for changes in the composition of oncology patients. The population-level HHI fell modestly from 2009 to 2014, a change of $-109$ HHI points $[-122, -99]$. This level change is absorbed by $\alpha$.

## 5.2 Identifying Assumptions

This approach makes two assumptions. First, individual oncologists' specialization choices do not change how fast individual guidelines grow. Second, a physician's exposure to knowledge growth is as good as random after the predictable part is removed.

The first assumption could be violated if causation ran in reverse, with increased specialization driving knowledge growth. Several features of the setting attenuate this worry. About 91% of the balanced panel are non-academic oncologists, for whom research is not typically an expected job function. The roughly 9% who are academic oncologists vary in how much research they do and how much time they spend on it. Physicians who report research as their primary activity were only 1.5% of the U.S. physician workforce in 2012

(National Institutes of Health, Advisory Committee to the Director, 2014; Jain et al., 2019).

The guideline process is also large relative to any one oncologist. As of 2011, roughly 44 NCCN panels wrote the guidelines, with nearly 900 clinicians (NCCN, 2011). An individual member could influence a guideline but is unlikely to move a process of that size. The guidelines also draw on thousands of scholarly works, far beyond any one oncologist's contribution.

New cancer treatments come out of an industry pipeline. One approved drug costs roughly $2.6 billion to develop and takes more than a decade from discovery to approval (DiMasi, Grabowski, and Hansen, 2016). Drug approvals during the sample period therefore largely reflect research decisions made before it began. Drug development also draws on a wide range of professionals outside clinical oncology, including laboratory scientists, regulatory experts, and the investors who fund development.

The second assumption, that exposure to knowledge growth is as good as random, could be violated in two ways. Even when cancer-level shocks are unpredictable, a physician's exposure to them is not random. Physicians who treat more cancer types are, by construction, exposed to more knowledge growth. For example, a breast-only oncologist may be more likely to practice in a large market. Oncologists in smaller markets may have a broader practice and different specialization trends unrelated to knowledge growth. Market size would then be correlated with both exposure to knowledge growth and changes in specialization, biasing the estimate of $\beta$.

Cancer-level knowledge growth may also be predictable and thus correlated with physician specialization. For example, an oncologist who expects rapid advances in breast cancer may choose to specialize in it. The next subsection removes the predictable part of a physician's exposure to knowledge growth. That part comes from how many cancer types the physician treats and from growth that cancer characteristics predict.

## 5.3 Recentering Exposure to Knowledge Growth

A physician's recentered exposure to knowledge growth is their exposure to knowledge growth with the predictable part removed (Borusyak and Hull, 2023). Most cancer types' guidelines grew between 2009 and 2014. Part of that growth may have been predicted, and the rest is the knowledge shock. I sum the knowledge shocks over the cancers the physician treats to get recentered exposure to knowledge growth.

For example, consider an oncologist who treats only breast cancer and another who treats both breast and kidney cancer. Suppose the breast cancer guideline grew by 30 pages and the kidney cancer guideline by 10 pages. The breast-only oncologist has 30 pages of exposure to knowledge growth, which I call unadjusted exposure. The oncologist who treats both types of cancer has 40 pages of exposure to knowledge growth. This second oncologist has more exposure only because they treat one more cancer type. The same pattern appears in the sample, where generalists, who treat more cancer types, have higher unadjusted exposure than specialists (Table 1).

Suppose oncologists expect every cancer type's guideline to grow by the same amount. Predicted growth for each guideline can then be estimated by the mean growth across all cancer types. Suppose that mean is 25 pages. Subtracting 25 pages from each leaves a breast knowledge shock of 5 pages and a kidney knowledge shock of $-15$ pages. The breast-only oncologist's recentered exposure is the breast knowledge shock, 5 pages more than expected. The oncologist who treats both types of cancer is exposed to both knowledge shocks, which sum to $-10$ pages, or 10 pages less than expected. The oncologist who treats both types of cancer now has less exposure to knowledge shocks than the breast-only oncologist. Treating more cancer types no longer mechanically raises recentered exposure. Rather, an oncologist's total knowledge shock reveals whether the guidelines relevant to them grew more or less than expected.

Formally, let $\mathbb{E}[\Delta k_c]$ be the predicted growth of cancer type $c$'s guideline, and $\widehat{\mathbb{E}}[\Delta k_c]$ the estimate of that predicted growth. The knowledge shock for cancer type $c$ is the growth

left after subtracting the predicted growth from the realized growth, $\Delta\tilde{k}_c = \Delta k_c - \widehat{\mathbb{E}}[\Delta k_c]$. Physician $i$'s recentered exposure to knowledge growth sums the knowledge shocks over the cancers they treated in the baseline year (2008), $\Delta\tilde{k}_i = \sum_{C_i} \Delta\tilde{k}_c$. The part removed is the predictable part of the physician's exposure to knowledge growth, where omitted variables bias is most likely to enter.

I replace unadjusted exposure ($\Delta k_i$) with recentered exposure ($\Delta\tilde{k}_i$) in the estimating equation, which gives a consistent estimate of $\beta$ (Borusyak and Hull, 2023):

$$\Delta\widehat{\text{HHI}}_i^{\text{jack}} = \alpha + \beta\Delta\tilde{k}_i + \theta X_i + \epsilon_i. \tag{1}$$

The identification assumption is that recentered exposure to knowledge growth is uncorrelated with the error term, including omitted variables ($\Delta\tilde{k}_i \perp \epsilon_i$). What this assumption means depends on how I estimate predicted growth by cancer type, $\mathbb{E}[\Delta k_c]$. The rest of this subsection presents three progressively richer estimates of $\mathbb{E}[\Delta k_c]$ and their implications for the identification assumption. The main analysis in Section 6 uses the richest estimate.

The simplest assumption is that no cancer type has predicted growth ($\mathbb{E}[\Delta k_c] = 0$). Under this assumption, recentered exposure ($\Delta\tilde{k}_i$) equals unadjusted exposure ($\Delta k_i$). The trend in knowledge growth before 2009 (Figure 1a) makes no predicted growth after 2009 implausible.

A more plausible assumption is that every cancer type's guideline will grow, but no one can predict which will grow more or less than average. Under this assumption, predicted knowledge growth ($\mathbb{E}[\Delta k_c]$) is the mean growth across all cancer types. This prediction removes the part of unadjusted exposure that comes from the count of cancer types a physician treats, but not growth that cancer characteristics predict.

A richer assumption lets predicted knowledge growth ($\mathbb{E}[\Delta k_c]$) vary with cancer characteristics. Acemoglu and Linn (2004) find that health conditions with more potential patients tend to experience more innovation, implying that more common cancers may experience greater knowledge growth. Budish, Roin, and Williams (2015) find that cancers

with high-quality surrogate endpoints also tend to experience more innovation. Surrogate endpoints let researchers judge whether a treatment works sooner and more cheaply than a primary outcome such as mortality. These two traits also influence a cancer's potential for specialization. A common cancer supplies enough patients to support more specialists, and faster research may make specialization more valuable. I incorporate both factors into the estimate of predicted knowledge growth, $\widehat{\mathbb{E}}[\Delta k_c]$.

I estimate predicted knowledge growth with a Poisson regression of each guideline's page count increase on cancer prevalence, high-quality surrogate endpoints, and their interaction. Let $p_c$ be the log of the number of Medicare patients with cancer type $c$ in 2008, which equals log prevalence up to an additive constant. Let $s_c$ indicate high-quality surrogate endpoints. The regression, which I call the predicted knowledge growth regression, is

$$\ln \mathbb{E}[\Delta k_c] = \delta_0 + \delta_1 s_c + \delta_2 p_c + \delta_3 s_c p_c. \tag{2}$$

The intercept $\delta_0$ captures growth common to all cancer types.

Table 2 shows the coefficient estimates. The regressors explain 71% of the variation in knowledge growth across cancer types. I attribute the remaining variation to randomness in research and development. For instance, some clinical trials and drug development lines fail, while others succeed.

The identification assumption now requires that the estimated knowledge shocks are unrelated to physicians' specialization decisions, except through knowledge growth's direct effect on specialization. Clinical research outcomes are hard to predict, which makes this assumption plausible. For example, only an estimated 14% of investigational drugs succeed (Schuhmacher et al., 2025).

## 5.4 Randomization Inference

Conventional robust standard errors overstate the precision of the estimates in this design. The data contain 35 independent cancer-level shocks, not thousands of independent physician-level shocks. For example, breast cancer's knowledge shock is shared by every oncologist who treats breast cancer, so their errors move together. Robust standard errors treat each physician's exposure to knowledge growth as independent and thus overreject the null hypothesis of no relationship (Adão, Kolesár, and Morales, 2019). In the balance tests, the shocks would look unbalanced even when they are balanced. In the causal regressions, estimated effects would look statistically significant even when they are not. Borusyak and Hull (2023) recommend randomization inference to overcome this challenge for this type of design.

Randomization inference randomly reassigns the knowledge shocks across cancer types. For example, suppose breast cancer received kidney cancer's knowledge shock and kidney cancer received breast cancer's. Every physician's recentered exposure is rebuilt from the reassigned shocks, and the relevant regression is run again. Each reassignment is a placebo shock realization and gives one placebo coefficient. Many reassignments show the distribution of how large a placebo coefficient chance alone produces. For measures recentered with a Poisson regression, the object reassigned is the standardized Poisson residual, so that each reassigned knowledge shock is sized to the predicted growth of the cancer type that receives it. Appendix D gives the details.

The balance tests in the next subsection regress exposure to knowledge growth on physician characteristics and report placebo intervals. The placebo interval runs from the 2.5th to the 97.5th percentile of the placebo coefficients. An estimate outside its placebo interval would indicate the characteristic is significantly correlated with knowledge shocks. Each balance regression also reports the $p$-value of a joint test of significance for all the physician characteristics, which compares the observed $F$-statistic with the $F$-statistics from the placebo shock realizations. A small $p$-value means the physician characteristics, taken together, predict exposure to knowledge growth more than they do in the placebo shock realizations.

Confidence intervals for $\beta$ in the main results come from test inversion, a procedure from Borusyak and Hull (2021). Test inversion repeats randomization inference for every candidate value of $\beta$, not only zero, and creates a confidence interval with every $\beta$ that is not rejected. The baseline version of randomization inference holds the coefficients that predict knowledge growth fixed and so ignores the error in estimating those coefficients. Two extensions allow for the estimation error. One averages the randomization inference $p$-values over draws of the coefficients from their estimated normal distribution. The other is more conservative and takes the largest $p$-value over all values of the coefficients that the data cannot rule out (Borusyak and Hull, 2021). Appendix D explains the randomization inference in full.

## 5.5 Balance

Tables 3 and 4 test whether physician characteristics predict exposure to knowledge growth. They regress exposure to knowledge growth on physician characteristics measured in the baseline year, namely market population, organization size, age, patient count, sex, and academic affiliation. Column (1) uses unadjusted exposure, and each later column uses exposure recentered with a progressively richer estimate of predicted knowledge growth. Table 3 reports generalists and Table 4 specialists.

Academic affiliation provides a test of the identification assumption. Academic oncologists may specialize faster for reasons outside knowledge growth, such as career incentives. Those reasons threaten the estimate of $\beta$ only if they are correlated with the knowledge shocks. Before recentering, academic affiliation shows a particularly large imbalance in unadjusted exposure to knowledge growth among specialists. Academic specialists' unadjusted exposure (column 1) is 232 pages lower than non-academic specialists', far outside the placebo interval $[-46.8, 47.9]$. The difference shrinks to 1.6 pages when exposure is recentered with the predicted knowledge growth regression (column 4), well inside the placebo interval $[-55, 53.7]$. This insignificant difference suggests that academic affiliation and its associated incentives are not correlated with exposure to knowledge shocks after recentering.

The joint test rejects for unadjusted exposure in both groups (column 1, $p = 0.001$ for specialists and 0.001 for generalists). Unadjusted exposure is therefore predictable from who the oncologist is. After recentering with the predicted knowledge growth regression, the joint test does not reject (column 4, $p = 0.857$ for specialists and 0.708 for generalists), so the shocks land essentially at random with respect to the characteristics observed.

After recentering (column 4), no specialist characteristic falls outside its placebo interval. Specialists' patient count has a $t$-statistic of $-3.2$, significant by conventional standards. However, the coefficient of $-0.33$ is well inside its placebo interval $[-0.68, 0.68]$, an example of the overrejection described in Section 5.4. One generalist characteristic falls outside its placebo interval: a generalist is exposed to about 0.41 fewer pages for each year of age, barely outside the placebo interval $[-0.398, 0.427]$. To account for this potential correlation, one version of the estimating equation includes age group fixed effects.

Identification requires that no unobserved physician characteristic is related to both the knowledge shocks and specialization. Only observed characteristics can be tested. With randomization inference, the balance tests show that recentering largely removes the relationship between observed physician characteristics and exposure to knowledge growth. This finding suggests that recentering also largely removes the relationship between unobserved physician characteristics and exposure to knowledge growth.

# 6 Effect of Knowledge Growth on Specialization

This section asks whether knowledge growth drives specialization. The estimates compare physicians with higher and lower recentered exposure to knowledge growth. Confidence intervals in this section come from the baseline version of randomization inference (Section 5.4) unless stated otherwise. Generalists more exposed to knowledge growth than their peers specialize little or no more. Nor do generalists meaningfully narrow the set of cancer types they treat. In contrast, for specialists, the data neither confirm nor rule out a meaningful

specialization response to knowledge growth.

For generalists, the estimated effect of exposure to knowledge growth on HHI is small and not statistically distinguishable from zero (Figure 3 and Table 5). Every 100 pages of recentered exposure to knowledge growth is associated with a 29-point increase in HHI (s.e. = 18). For scale, this is a 2.9% increase in HHI for a 9.7% increase in relevant guideline content. Equivalently, this is a 0.03 standard deviation increase in HHI (s.d. = 559 HHI points) per standard deviation increase in exposure to knowledge growth (s.d. = 66 pages). For generalists, the data rule out a meaningful specialization response to knowledge growth. Across specifications, the confidence intervals rule out increases above 70 HHI points per 100 pages, or 0.09 standard deviations per standard deviation of exposure.

In contrast, for specialists, the estimates show minimal change in specialization in response to knowledge growth, but the data cannot rule out a meaningful response. Every 100 pages of recentered exposure to knowledge growth is associated with a 47-point increase in HHI (s.e. = 60). For scale, this is a 1.6% increase in HHI for a 12.1% increase in relevant guideline content. Equivalently, this is a 0.03 standard deviation increase in HHI (s.d. = 1,334 HHI points) per standard deviation increase in exposure to knowledge growth (s.d. = 76 pages). Across specifications, the data cannot rule out increases as large as 386 HHI points per 100 pages, or 0.22 standard deviations per standard deviation of exposure.

The specialist confidence intervals are wider than the generalist ones for at least two mechanical reasons: specialists' changes in HHI vary more across physicians, and the specialist sample is one third the size of the generalist sample. Beyond these mechanical reasons, specialists are more often academic, and academic career incentives may add variation in specialization unrelated to knowledge growth. In each specification, the specialist estimate is larger than the generalist estimate, but the difference is not statistically significant under conventional inference. The confidence intervals from randomization inference cannot rule out a much larger specialist response. As such, the data neither confirm nor rule out that knowledge growth contributed to the rise in specialization among the most specialized

oncologists (Section 4.3).

HHI reflects two margins of specialization: shifting time among the cancer types a physician already treats (the intensive margin) and changing how many cancer types they treat (the extensive margin). Regressions of the count of cancer types isolate the extensive margin.

For generalists, the data rule out a meaningful reduction in the count of cancer types they treat (Table 6). In the baseline year, generalists treat an average of 15 cancer types. Every 100 pages of recentered exposure to knowledge growth is associated with a decrease of 0.08 cancer types (s.e. = 0.10). For scale, this is a 0.6% decrease in cancer types for a 9.7% increase in relevant guideline content. The 95% confidence interval spans −0.60 to 0.33 cancer types per 100 pages, and the results with fixed effects are similar. For even a two standard deviation increase in exposure, these specifications rule out reductions greater than 0.9 cancer types.

In contrast, for specialists, the data cannot rule out a slightly larger reduction in the count of cancer types they treat. In the baseline year, specialists treat an average of 12 cancer types. Every 100 pages of recentered exposure to knowledge growth is associated with a decrease of 0.13 cancer types (s.e. = 0.15). For scale, this is a 1.1% decrease in cancer types for a 12.1% increase in relevant guideline content. The confidence intervals are wider than the generalist ones. Across specifications, the data rule out only reductions greater than 1.3 cancer types per 100 pages. Specialists may therefore still respond to large knowledge shocks on the extensive margin by dropping one or two cancer types.

The main findings rest on several design choices. Each robustness check in Appendix E changes one of these choices and compares its estimates with the main estimates. The choices include how the confidence intervals are built, how knowledge growth is measured and recentered, when a physician counts as exposed to a cancer's guideline growth, and which visits build the HHI. Across the robustness checks, a one standard deviation rise in exposure to knowledge growth changes generalist HHI by 0.00 to 0.04 standard deviations.

The generalist confidence intervals rule out increases above 0.12 standard deviations, except the worst-case confidence intervals, which rule out nothing. For specialists, the estimates range from −0.06 to 0.08 standard deviations. The specialist confidence intervals include zero and reach increases of up to 0.22 standard deviations.

# 7 Constraints to Specialization

Section 6 shows that generalists do not significantly increase their specialization as knowledge grows, which suggests that the private returns to specialization are low or the costs of specializing are high for most oncologists. This section examines several factors that may constrain specialization, including market size, organization size and type, and the returns to specialization. Section 7.1 shows that median specialization is largely invariant to market size, while top specialization rises with market size. Section 7.2 shows that among non-academic organizations, specialization does not rise with organization size. In contrast, among academic organizations specialization rises steeply with organization size. Section 7.3 shows that, outside academic organizations, the returns to specialization in patient care appear limited, whether through prices, patient volume, or referrals.

## 7.1 Market Size

If patient volume constrains specialization, oncologists should be more specialized in larger markets. Figure 4a and Table A.3 show median specialization by market size. Median specialization has remained largely invariant to market size. In 1999, the gradient of median specialization was 3 HHI points (s.e. = 10) per doubling of market size, statistically indistinguishable from zero. By 2014, the gradient had risen only slightly, to 70 HHI points (s.e. = 23) per doubling of market size. Despite the extensive knowledge growth between 1999 and 2014, the median oncologist in a large market remained little more specialized than one in a small market.

Top specialization, by contrast, rises with market size, and the gradient has steepened over time. Figure 4b shows top specialization, the 90th percentile of specialization, by market size. In 1999, top specialization increased by 221 HHI points (s.e. = 44) per doubling of market size. The gradient is consistent with the classical economic argument that specialization is limited by the extent of the market (Section 2 and Figure A.8a). By 2014, the gradient had steepened by an additional 449 HHI points (s.e. = 86) per doubling of market size. As a result, growth in specialization has been concentrated in large markets, improving local access to subspecialist care there, while access in smaller markets has been static or has declined on average. The steepening gradient of top specialization is consistent with the prediction in Section 2 that knowledge growth may widen the gap in specialization between small and large markets. These patterns are consistent with market size and patient volume constraining top specialization but not median specialization.

Outside the largest regions, most cancer types have too few patients to support a fully specialized oncologist. Figure A.10 shows the average share of cancer types that can support one, by region size in terms of cancer care volume. In a given region, a cancer type is considered able to support a fully specialized oncologist when the available patients generate at least the median annual work across oncologists. In the smallest tenth of regions, 3% of cancer types can support a fully specialized oncologist, and in the largest tenth, 53% can. These patterns are consistent with market size limiting how far specialization can grow outside the largest markets.

Accordingly, patients have limited access to oncologists who specialize in their cancer type. Figure A.11 shows how far cancer patients live from the nearest dedicated specialist in their cancer type in 2014. I define a dedicated specialist as an oncologist who devotes more than half of their work to one cancer type. Only 31% of patients live within 25 miles of a dedicated specialist in their cancer type, and 37% have none within 100 miles. In short, local access to a dedicated specialist in the patient's cancer type is the exception rather than the norm.

## 7.2 Organization Size and Type

If organizations reduce coordination costs, oncologists should be more specialized in larger organizations. Figure 5 and Table A.4 show median and top specialization by organization size and type in 2014. Within non-academic organizations, median and top specialization are flat across organization size; neither gradient is statistically distinguishable from zero. Non-academic organizations thus do not appear to facilitate specialization, consistent with high internal coordination costs or with low returns to specialization.

In contrast, among academic organizations median and top specialization rise steeply with organization size. Median specialization rises by 41 HHI points (s.e. = 9) per additional oncologist, and top specialization by 60 HHI points (s.e. = 11). Figure 5 shows that top specialization in the largest academic organizations reaches about 8,500 HHI points. The most specialized oncologists there thus devote at least 85% of their work to one cancer type. One interpretation of the rise in specialization with organization size is that academic organizations lower coordination costs or raise the returns to specialization. Once the returns outweigh the costs, organization size is a binding constraint on specialization, and loosening this constraint raises specialization.

Academic oncologists are more specialized than non-academic oncologists, which may partly reflect that academic organizations sit in larger markets. Figure A.12 and Table A.5 show specialization by market size for academic and non-academic oncologists in 2014. Among non-academic oncologists, specialization rises modestly, by 123 HHI points (s.e. = 33) per doubling of market size. The average specialization of non-academic oncologists ranges from about 1,000 to 1,800 HHI points across market sizes. Even in the largest markets, where patient volume can support specialization, the average non-academic oncologist is not particularly specialized, consistent with something other than patient volume limiting their specialization.

Oncologists in academic organizations are substantially more specialized than those in non-academic organizations at the same market size, and the gap appears to widen as market

size grows. At the mean of log market size, specialization in academic organizations is 1,845 HHI points (s.e. = 190) higher than in non-academic organizations. Specialization in academic organizations rises by 520 HHI points (s.e. = 191) per doubling of market size. Market size thus does not explain why oncologists in academic organizations are more specialized, consistent with academic organizations changing the costs or returns to specialization, or with more specialized oncologists selecting into working there.

Collectively, these results point to scale as a constraint on specialization. The market-size results imply that specialized cancer care has large economies of scale, which appear to be growing over time. In addition, specialization appears to require scale on multiple dimensions, in both market size and organizational scale.

Specialists are disproportionately employed in academic organizations (Section 4). Outside patient care, academic careers may reward a narrow focus through research and promotion, so oncologists in academic organizations may face higher returns to specialization. Peer-reviewed publications appear in the promotion and tenure guidelines of 95% of 92 universities with biomedical faculties (Rice et al., 2020). Among 10,241 academic physicians, salary was positively associated with publications, clinical-trial participation, and Medicare payments, a proxy for clinical volume (Jena, Olenski, and Blumenthal, 2016). The salary evidence suggests that both clinical and research work have returns in academic organizations, where the two compete for the same limited time. Specialized oncologists in academic organizations may both respond to knowledge growth and produce it. As knowledge grows, both kinds of work potentially become harder to do broadly, and for oncologists in academic organizations it is difficult to empirically separate responding to knowledge growth from producing it.

Oncologists who prefer focused expertise, scholarly work, or teaching may also select into academic organizations. In contrast, those who prioritize income or prefer clinical variety may choose non-academic organizations. The data cannot separate career incentives from selection. Oncologists in non-academic organizations likely lack promotion rewards for publications and earn returns to specialization mainly through patient care.

## 7.3 Returns to Specialization

Price incentives for producing high-quality care are blunted. During the study period, traditional Medicare reimbursement for oncology did not incorporate quality metrics, and more recent payment models have only offered limited, opt-in participation in quality-based payment systems (Chong et al., 2022). Although Medicare is not the sole payer, it represents a large share of the oncology patient population: 54% of U.S. cancers diagnosed in 2009 were in people aged 65 and older (CDC, 2013). Even if specialization improves quality, prices do not fully reward that quality, limiting the returns to specialization.

Patients are rarely matched to a dedicated specialist in their cancer type. Figure A.11 shows that 99% of patients in 2014 have a dedicated specialist in their cancer type somewhere in the country, but only 7% are treated by one. Even among patients with a dedicated specialist within 10 miles, only 21% are treated by one. Low matching is consistent with several possibilities, including high search costs, a capacity constraint on specialists, or a low value placed on specialization by patients or referring physicians. In a setting with these frictions, specialization has limited returns in terms of attracting patient volume.

Matching rises with income, suggesting that a match is desirable but costly. Among patients with a dedicated specialist within 10 miles, 18% in the bottom fifth of income (based on ZIP code of residence) are treated by one (Figure A.13). In the top fifth of income, the share rises to 24%. The higher share in the top fifth is consistent with higher-income patients being better able to overcome search frictions.

Disincentives to refer patients may also limit the private returns to specialization for oncologists in non-academic organizations. Oncologists in non-academic organizations often face incentives to maximize their own patient volume. However, a specialist depends on referrals from other oncologists, and those oncologists lose revenue when they refer a patient away. Anti-kickback regulations prohibit compensation for referrals across organizations (Section 2). Community oncologists describe a compensation plan that removes competition for patients within an organization as a requirement for specialization, and one that many

practices struggle to meet (Gesme and Wiseman, 2011). As a result, oncologists in non-academic organizations may keep the patients who come to them rather than divide patients by cancer type with their colleagues. In contrast, academic organizations appear able to sustain a referral equilibrium in which oncologists refer patients to one another, through some combination of incentives and norms.

In summary, oncologists differ widely in how specialized they are. Specialization varies substantially across markets and organizations, with increasing divergence over time. Whether this divergence matters depends on whether specialists provide better care. If they do, disparities in outcomes may be growing. The next section asks whether specialists differ in the care they provide, as reflected in their use of anti-cancer drugs.

# 8 Technology Use and Specialization

The variation in specialization across markets and organizations raises a question: are there returns to specialization in the quality of patient care? The data allow an examination of how the anti-cancer drugs that oncologists prescribe vary with specialization. Anti-cancer drugs are central to the treatment of cancer and influence clinical outcomes. They are also challenging to administer due to their complex dosing protocols and toxicity profiles. Among physicians, medical oncologists are uniquely trained to manage these drugs. The choice of anti-cancer drugs is therefore a core part of oncologists' practice and of their use of technology.

The findings indicate that specialization plausibly enhances the quality of care. Section 8.1 explains how the analysis estimates characteristics of the typical set of anti-cancer drugs that oncologists prescribe over time. Section 8.2 shows that specialists prescribe a narrower range of drugs and have more experience with each drug than the most general oncologists. Specialists also prescribe newer drugs and a higher share of breakthrough-designated drugs. Section 8.3 shows that growing differences in these prescribing outcomes primarily reflect a rise in specialization among the most specialized oncologists. Section 8.4 discusses the limitations,

outside evidence that specialization improves patient outcomes, and why specialization appears under-rewarded.

## 8.1 Data and Measurement

The analysis estimates characteristics of the typical set of anti-cancer drugs that oncologists prescribe over time. Drugs are identified by the World Health Organization's Anatomical Therapeutic Chemical (ATC) codes. A drug counts as anti-cancer if it appears on the list of anti-cancer drugs in Pantziarka et al. (2021). Prescriptions come from a 20% sample of Medicare Part B and Part D claims, which identify prescribing physicians by their National Provider Identifiers (NPIs). The prescription data begin in 2008, the first year in which Part D claims broadly recorded prescribers' NPIs.

The criteria for including oncologists differ from those in Section 3. Each year, an oncologist must have prescribed anti-cancer drugs to at least 11 patients and seen at least 11 cancer patients. The thresholds ensure enough patients to measure each oncologist's prescribing and HHI, and comply with privacy standards. The analysis uses an unbalanced rather than a balanced panel, keeping every eligible oncologist in each year, and extends the period studied to 2008–2019. Because the oncologists in the sample change from year to year, each oncologist's specialization is measured by the HHI in that year.

## 8.2 Prescribing Patterns

Specialization allows oncologists to narrow the range of drugs they prescribe. Figure 6a shows the average number of unique anti-cancer drugs that oncologists prescribe each year, with oncologists grouped each year by their HHI percentile. In 2008, the first year of the panel, specialists (the top quarter by HHI each year) prescribe slightly fewer unique anti-cancer drugs than the most general oncologists (the bottom half by HHI each year). Specialists prescribe an average of 15.9 (0.2) unique anti-cancer drugs per year. In contrast, the most general oncologists prescribe an average of 18.3 (0.1) drugs, approximately 15% more than

specialists.

Over time, specialists and the most general oncologists differ more in the number of unique anti-cancer drugs they prescribe. By 2019, the average number of unique anti-cancer drugs that specialists prescribe falls to 12.3 (0.3) per year. In contrast, the most general oncologists prescribe an average of 19.9 (0.2) unique anti-cancer drugs per year, 61% more than specialists. The growing difference is consistent with specialization narrowing the set of drugs that oncologists prescribe.

A narrower scope of drugs allows specialists to accumulate greater experience with the drugs they prescribe. An oncologist's experience with a drug is measured as the number of patients to whom the oncologist prescribes the drug within a calendar year. The measure captures only recent experience, reflecting evidence that physician-specific human capital deteriorates quickly (Facchini, 2022; Hockenberry and Helmchen, 2014). To compare across drugs given to varied numbers of patients, experience is normalized across oncologists within each drug and year, resulting in a relative experience measure with a mean of 1.

Figure 6b shows average relative experience by level of specialization. In 2008, specialists and the most general oncologists differ by less than 1% in average relative experience. Over time, however, specialists accumulate more experience relative to the most general oncologists. By 2019, specialists have approximately 42% higher average relative experience with the drugs they prescribe than the most general oncologists. Given the complexity and toxicity risks of many anti-cancer drugs, greater experience with specific drugs is likely to enhance clinical management and reduce adverse effects.

Specialization also appears to enhance oncologists' ability to stay closer to the technological frontier, as reflected in the age and quality of the drugs they prescribe. Drug age is defined as the number of years since the first FDA approval for the drug's ATC code, regardless of indication, dosage, or delivery form.[2] Drug quality is proxied by whether the drug's active molecule had ever received a "Breakthrough Therapy" designation from the FDA as of 2025.

[2]This approach likely overestimates drug age compared with methods that account for new indications. However, the data do not identify prescribing indication.

The FDA grants the designation on preliminary clinical evidence of substantial improvement over existing therapies. The designation is positively associated with measures of clinical benefit, which makes it a credible indicator of drug quality (Feng and Maini, 2016; Molto et al., 2020). Congress created the designation in 2012, and the FDA first approved a drug under it in late 2013 (U.S. Food and Drug Administration, 2024). Every designated drug prescribed before 2013 was therefore designated after it reached the market. I count these drugs intentionally, because the designation signals drug quality even if granted years later. For each oncologist and year, I measure the average age of the drugs prescribed and the share of them that are breakthrough-designated, weighting each drug by the number of the oncologist's patients who received it that year.

Figure 7a shows the average age of the anti-cancer drugs that oncologists prescribe by level of specialization over time. In 2008, the drugs that specialists prescribe are on average about 0.9 year newer than those of the most general oncologists. The average age of an oncologist's drugs would rise by one year per year if the oncologist kept prescribing the same drugs to the same shares of patients. Both groups shift toward newer drugs, so the average age of their drugs rises by less than one year per year. The rise in age is faster for the most general oncologists than for specialists. By 2019, the drugs that the most general oncologists prescribe are on average 2.9 years older than those that specialists prescribe. The growing difference suggests that specialization facilitates more rapid adoption of new drugs.

Figure 7b shows that specialists also prescribe a higher share of breakthrough-designated drugs than the most general oncologists. Both the share and the difference in shares between the two groups vary across years, but specialists are persistently more likely to incorporate breakthrough-designated drugs into their practice.

## 8.3 Specialization and Patients Over Time

The growing differences in the three prescribing outcomes (drug age, relative experience, and unique anti-cancer drugs) primarily reflect a rise in specialization among the most specialized

oncologists. Table 7 non-parametrically decomposes the growing difference in each prescribing outcome into three parts: the change in the distribution of HHI, the change in outcomes at a given HHI, and a remainder. A majority of the growing differences come from the change in the distribution of HHI: 84% for drug age, 55% for relative experience, and 67% for unique anti-cancer drugs. A minority of the growing differences come from the change in outcomes at a given HHI. Specialists and the most general oncologists prescribed more similarly in 2008 because they were closer in specialization. The differences in their prescribing grew primarily because their specialization diverged. The decomposition is consistent with specialization helping oncologists stay closer to the technological frontier, since the two groups' prescribing diverged primarily because their specialization diverged.

Specialists and the most general oncologists also treat increasingly different cancers, while their patients' other characteristics differ more modestly. Figure A.14 plots the distance between the cancers the two groups treat, measured as the share of time the average oncologist in either group would have to shift to other cancer types to match the average oncologist in the other group. The distance rises from 19% in 2008 to 24% in 2019. Besides cancer type, the two groups' patients differ in age, sex, chronic conditions, and ZIP-code income (Figure A.15). Patients of specialists are about 2 years younger and 5 percentage points more likely to be female than patients of the most general oncologists, but these differences are stable over time and so are unlikely to account for the growing differences in prescribing. In 2008, patients of specialists also have 6% fewer chronic conditions and 13% higher ZIP-code income, and these differences grow to 11% and 19% by 2019, more slowly than the prescribing differences.

## 8.4 Discussion

Specialists prescribe increasingly differently from the most general oncologists. Specialists prescribe fewer unique anti-cancer drugs, have more experience with each drug they prescribe, and prescribe newer drugs and a higher share of breakthrough-designated drugs. The growing

differences primarily reflect a rise in specialization. The two groups' patients differ modestly in characteristics other than cancer type. The prescribing differences suggest that specialists may deliver higher-quality care, increasingly so. The findings are also consistent with specialization enabling oncologists to stay closer to the technological frontier.

A key limitation of this analysis is that it is observational. Patients sort to specialists, so a causal comparison of treatment differences across oncologists would need an additional identification strategy. The analysis also does not directly test whether these prescribing differences translate into better patient outcomes. The claims do not record cancer stage, which is the strongest predictor of survival, so I judged that a comparison of outcomes would be too exposed to unmeasured severity to be credible or precise.

Other studies examine patient outcomes and link specialization with better survival. Associational evidence suggests that specialists deliver better care: guideline-concordant care improves survival and is more common where specialists concentrate (Voss et al., 2017; Thiels et al., 2020). In a working paper, Karadakic et al. (2026) instrument for access to a subspecialist with distance and estimate that access to an oncologist subspecialized in the patient's cancer type reduces three-year mortality by about 10% without raising Medicare spending. They also find that access to a subspecialist raises the use of recently approved therapies, replicating the faster adoption of new drugs by specialists shown in Figure 7a.

Taken together, the evidence so far suggests that specialization improves patient outcomes. The returns to specialization likely accrue mainly to patients and other parties in a setting with weak price incentives for quality, capacity constraints on specialists, and further market frictions discussed in Section 7. Specialization therefore appears under-rewarded, which may help explain why most oncologists remain generalists as knowledge grows.

# 9 Conclusion

Over two decades, clinical knowledge for cancer care grew rapidly. The median oncologist did not become more specialized over time or in response to knowledge growth. The most specialized oncologists specialized further, but this paper's estimates neither confirm nor rule out that knowledge growth contributed to this rise in specialization. Specialization is concentrated in academic organizations and large markets. Outside academic organizations, the returns to oncologists from specializing in patient care are limited. However, specialization appears to improve patient care. Compared with general oncologists, specialized oncologists have more experience with each drug they prescribe and adopt newer drugs faster. Specialization therefore appears under-rewarded, likely reflecting low private returns to most oncologists rather than low social returns.

The findings have key limits in what they reveal about knowledge growth and patient outcomes. This paper estimates the response to unexpected knowledge shocks, not to predicted knowledge growth. Oncologists may respond differently to predicted knowledge growth than to unexpected knowledge shocks, since they are more able to anticipate it and adjust their practice. Given limitations in the available data, this paper also does not directly test whether specialization improves patient outcomes. Other studies link specialization with improved survival. Causally identified tests of whether specialization improves survival are left for ongoing and future work.

Knowledge growth may affect knowledge users and knowledge creators in different ways. For knowledge creators, such as academic researchers, careers may reward a narrow focus. In contrast, workers who primarily apply knowledge, such as most oncologists, face a trade-off: specializing lowers the cost of staying up to date but narrows the market they can serve. Wherever knowledge grows rapidly, workers who apply it may risk falling behind the technological frontier.

## Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the author used Claude (Anthropic) in order to revise the text, write and check analysis code, and extract citations from guideline pages. The author reviewed and edited the output as needed and takes full responsibility for the content of the published article.

## References


Acemoglu, Daron, and Joshua Linn. 2004. "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry." *The Quarterly Journal of Economics*, 119(3): 1049–1090.

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

Agranov, Marina, Andrew Schotter, and Isabel Trevino. 2025. "Complex for Whom? An Experimental Approach to Subjective Complexity." Working paper.

Alcacer, Juan, and Michelle Gittelman. 2006. "Patent Citations as a Measure of Knowledge Flows: The Influence of Examiner Citations." *The Review of Economics and Statistics*, 88(4): 774–779.

American Cancer Society. 2024. "Lifetime Probability of Developing (2017–2019) or Dying (2018–2020) from Cancer." *Cancer Facts & Figures 2024, Supplemental Data*, https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2024/sd4-lifetime-probability-2024.pdf. Accessed January 18, 2024.

Armitage, James O. 2010. "Early-stage Hodgkin's lymphoma." *New England Journal of Medicine*, 363(7): 653–662.

Arrow, Kenneth J, L Kamran Bilir, and Alan Sorensen. 2020. "The impact of information technology on the diffusion of new pharmaceuticals." *American Economic Journal: Applied Economics*, 12(3): 1–39.

Atalay, Enghin, Sebastian Sotelo, and Daniel Tannenbaum. 2024. "The Geography of Job Tasks." *Journal of Labor Economics*, 42(4): 979–1008.

Baicker, Katherine, and Amitabh Chandra. 2004. "The Productivity of Physician Specialization: Evidence from the Medicare Program." *American Economic Review*, 94(2): 357–361.

Baumgardner, James R. 1988. "Physicians' Services and the Division of Labor Across Local Markets." *Journal of Political Economy*, 96(5): 948–982.

Becker, G. S., and K. M. Murphy. 1992. "The Division of Labor, Coordination Costs, and Knowledge." *The Quarterly Journal of Economics*, 107(4): 1137–1160.

Berger, Roger L., and Dennis D. Boos. 1994. "P values maximized over a confidence set for the nuisance parameter." *Journal of the American Statistical Association*, 89(427): 1012–1016.

Borusyak, Kirill, and Peter Hull. 2021. "Non-Random Exposure to Exogenous Shocks: Theory and Applications." National Bureau of Economic Research Working Paper 27845. Revised December 2021.

Borusyak, Kirill, and Peter Hull. 2023. "Nonrandom Exposure to Exogenous Shocks." *Econometrica*, 91(6): 2155–2185.

Budish, Eric, Benjamin N Roin, and Heidi Williams. 2015. "Do firms underinvest in long-term research? Evidence from cancer clinical trials." *American Economic Review*, 105(7): 2044–2085.

Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications.* New York:Cambridge University Press.

CDC. 2013. "Invasive cancer incidence—United States, 2009." *MMWR Morbidity and Mortality Weekly Report*, 62(7): 113–118. Centers for Disease Control and Prevention.

Centers for Disease Control and Prevention. 2019. "10 Leading Causes of Death, United States 2020, Both Sexes, All Ages, All Races."

Chandra, Amitabh, David Malenka, and Jonathan Skinner. 2014. "The diffusion of new medical technology: The case of drug-eluting stents." In *Discoveries in the Economics of Aging*. 389–403. University of Chicago Press.

Chong, Alexandra, Eileen Witherspoon, Batsheva Honig, Elizabeth Ela, Hillary Cavanagh, and Lara Strawbridge. 2022. "Reflections on the Oncology Care Model and Looking Ahead to the Enhancing Oncology Model." *JCO Oncology Practice*, 18(10): 685–690.

DiMasi, Joseph A., Henry G. Grabowski, and Ronald W. Hansen. 2016. "Innovation in the pharmaceutical industry: New estimates of R&D costs." *Journal of Health Economics*, 47: 20–33.

Dingel, Jonathan I, Joshua D Gottlieb, Maya Lozinski, and Pauline Mourot. 2023. "Market Size and Trade in Medical Services."

Facchini, Gabriel. 2022. "Forgetting-by-not-doing: The Case of Surgeons and Cesarean Sections." *Health Economics*, 31(3): 481–495.

Feng, Josh, and Luca Maini. 2016. "Where Is the Good Stuff? Drug Quality Measures and Economic Research." National Bureau of Economic Research White paper.

Gabaix, Xavier, and Thomas Graeber. 2024. "The Complexity of Economic Decisions." National Bureau of Economic Research Working Paper 33109.

Ganguli, Ina, Jeffrey Lin, Vitaly Meursault, and Nicholas F. Reynolds. 2026. "Spreading Out Across Expanding Idea Space." National Bureau of Economic Research w35499, Cambridge, MA.

Garicano, Luis. 2000. "Hierarchies and the Organization of Knowledge in Production." *Journal of Political Economy*, 108(5): 874–904.

Garicano, Luis, and Tano Santos. 2004. "Referrals." *American Economic Review*, 94(3): 499–525.

Garicano, Luis, and Thomas N. Hubbard. 2009. "Specialization, Firms, and Markets: The Division of Labor within and between Law Firms." *The Journal of Law, Economics, and Organization*, 25(2): 339–371.

Gesme, Dean H, and Marian Wiseman. 2011. "Subspecialization in community oncology: option or necessity?" *Journal of Oncology Practice*, 7(3): 199–201.

Glied, Sherry, and Adriana Lleras-Muney. 2008. "Technological Innovation and Inequality in Health." *Demography*, 45(3): 741–761.

Hockenberry, Jason M, and Lorens A Helmchen. 2014. "The Nature of Surgeon Human Capital Depreciation." *Journal of Health Economics*, 37: 70–80.

Jaffe, Elaine S. 2009. "The 2008 WHO classification of lymphomas: implications for clinical practice and translational research." *Hematology. American Society of Hematology. Education Program*, 523–531.

Jain, Mukesh K., Vivian G. Cheung, Paul J. Utz, Brian K. Kobilka, Tadataka Yamada, and Robert Lefkowitz. 2019. "Saving the Endangered Physician-Scientist — A Plan for Accelerating Medical Breakthroughs." *New England Journal of Medicine*, 381(5): 399–402.

Jena, Anupam B, Andrew R Olenski, and Daniel M Blumenthal. 2016. "Sex Differences in Physician Salary in US Public Medical Schools." *JAMA Internal Medicine*, 176(9): 1294–1304.

Jones, Benjamin F. 2009. "The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder?" *Review of Economic Studies*, 76(1): 283–317.

Kann, Benjamin H, Skyler B Johnson, Hugo JWL Aerts, Raymond H Mak, and Paul L Nguyen. 2020. "Changes in Length and Complexity of Clinical Practice Guidelines in Oncology, 1996-2019." *JAMA Network Open*, 3(3): e200841.

Karadakic, René, Christopher Manz, Arno Cai, David C Chan, Bruce E Landon, Jukka-Pekka Onnela, Nancy L Keating, and Michael L Barnett. 2025. "Geographic Variation in the Utilization of Cancer Care From Subspecialized Medical Oncologists in the United States, 2008 to 2020." *Annals of Internal Medicine*, 178(10): 1420–1428.

Karadakic, René, David C Chan, Nancy L Keating, Bruce E Landon, and Michael L Barnett. 2026. "The Value of (Sub) Specialization: Evidence from Oncology." CESifo Working Paper 12937.

Kim, Sunwoong. 1989. "Labor Specialization and the Extent of the Market." *Journal of Political Economy*, 97(3): 692–705.

McGivney, Bill. 2008. "The NCCN Compendium for Cancer Management." *American Health & Drug Benefits*, 1(5): 40–44. Interview with Bill McGivney, PhD.

Molto, Consolación, Thomas J. Hwang, Maria Borrell, Marta Andres, Ignasi Gich, Agustí Barnadas, Eitan Amir, Aaron S. Kesselheim, and Ariadna Tibau. 2020. "Clinical benefit and cost of breakthrough cancer drugs approved by the US Food and Drug Administration." *Cancer*, 126(19): 4390–4399.

Mulcahy, Nick. 2010. "First-Ever Guideline on Mesothelioma Issued by NCCN." *Medscape*. Accessed: 2024-08-06.

National Institutes of Health, Advisory Committee to the Director. 2014. "Physician-Scientist Workforce Working Group Report." National Institutes of Health, Bethesda, MD.

NCCN. 2011. "NCCN Clinical Practice Guidelines in Oncology." *Web page, Internet Archive capture of 6 April 2011.*

NCCN. 2023*a*. "About the National Comprehensive Cancer Network." https://www.nccn.org/home/about. Accessed January 15, 2024.

NCCN. 2023*b*. "History of the National Comprehensive Cancer Network." https://www.nccn.org/home/about/nccn-history. Accessed January 15, 2024.

NCCN. 2024. "Development and Update of Guidelines." Accessed: 2024-08-08.

NCI. 2022. "Surgical Pathology Reports." Accessed: 2026-08-28.

NCI. 2026. "Metastatic Cancer: When Cancer Spreads." Accessed: 2026-08-28.

Nei, Masatoshi. 1987. *Molecular Evolutionary Genetics.* New York:Columbia University Press.

Olivier, Timothée, Alyson Haslam, and Vinay Prasad. 2021. "Anticancer drugs approved by the US food and drug administration from 2009 to 2020 according to their mechanism of action." *JAMA Network Open*, 4(12): e2138793.

Pantziarka, Pan, Rica Capistrano I, Arno De Potter, Liese Vandeborne, and Gauthier Bouche. 2021. "An open access database of licensed cancer drugs." *Frontiers in Pharmacology*, 12: 627574.

Priem, Jason, Heather Piwowar, and Richard Orr. 2022. "OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts." *arXiv preprint arXiv:2205.01833*.

ResDAC. 2024. "Data on Provider Practice and Specialty (MD-PPAS) User Guide." Centers for Medicare & Medicaid Services User Guide. Chronic Conditions Warehouse Virtual Research Data Center.

Rice, Danielle B, Hana Raffoul, John P A Ioannidis, and David Moher. 2020. "Academic Criteria for Promotion and Tenure in Biomedical Sciences Faculties: Cross Sectional Analysis of International Sample of Universities." *BMJ*, 369: m2081.

Roach, Michael, and Wesley M Cohen. 2013. "Lens or Prism? Patent Citations as a Measure of Knowledge Flows from Public Research." *Management Science*, 59(2): 504–525.

Schlafly, Andrew, and Ronnie Sebro. 2022. "Does NIH funding differ between medical specialties? A longitudinal analysis of NIH grant data by specialty and type of grant, 2011–2020." *BMJ Open*, 12(12): e058191.

Schuhmacher, Alexander, Markus Hinder, Elazar Brief, Oliver Gassmann, and Dominik Hartl. 2025. "Benchmarking R&D success rates of leading pharmaceutical companies: an empirical analysis of FDA approvals (2006-2022)." *Drug Discovery Today*, 30(2): 104291. PMID: 39805539.

Skinner, Jonathan. 2011. "Causes and consequences of regional variations in health care." In *Handbook of Health Economics*. Vol. 2, 45–93. Elsevier.

Skinner, Jonathan, and Douglas Staiger. 2015. "Technology diffusion and productivity growth in health care." *Review of Economics and Statistics*, 97(5): 951–964.

Smith, Adam. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations.* London, United Kingdom:W. Strahan and T. Cadell.

Taylor, H. 2004. "Most physicians do not participate in clinical trials because of lack of opportunity, time, personnel support and resources." *Harris Interactive: Health Care News*, 4(9): 1–8.

Thiels, Cornelius A, Kristine T Hanson, Elizabeth B Habermann, Judy C Boughey, and Travis E Grotz. 2020. "Integrated Cancer Networks Improve Compliance with National Guidelines and Outcomes for Resectable Gastric Cancer." *Cancer*, 126(6): 1283–1294.

U.S. Food and Drug Administration. 2024. "CDER Breakthrough Therapy Designation Approvals." Center for Drug Evaluation and Research. Data as of December 31, 2024.

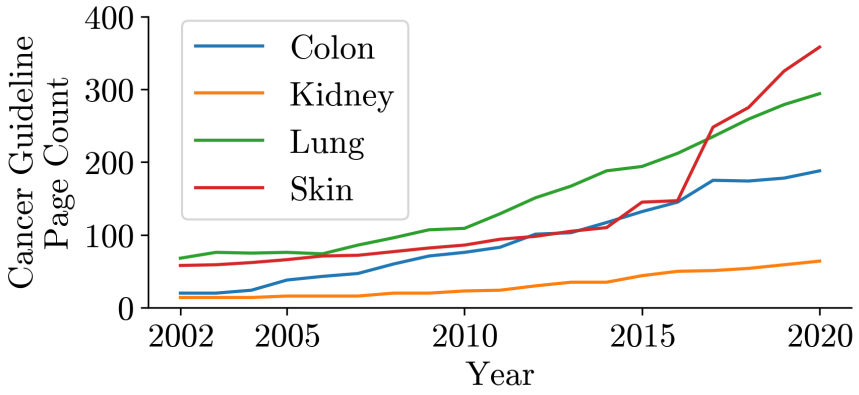
Voss, Rachel K, Yi-Ju Chiang, Keila E Torres, B Ashleigh Guadagnolo, Gary N Mann, Barry W Feig, Janice N Cormier, and Christina L Roland. 2017. "Adherence to National Comprehensive Cancer Network Guidelines is Associated with Improved Survival for Patients with Stage 2A and Stages 2B and 3 Extremity and Superficial Trunk Soft Tissue Sarcoma." *Annals of Surgical Oncology*, 24: 3271–3278.

Wallis, Claudia. 2022. "A Wave of New Cancer Treatments Challenges Community Oncologists to Keep Up." *Scientific American*. Accessed: 2024-06-07.

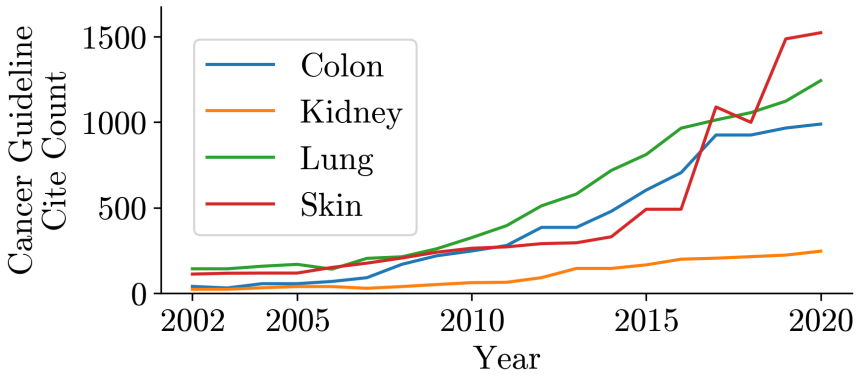
Welch, W Pete, and Andrew B Bindman. 2016. "Town and Gown Differences Among the 100 Largest Medical Groups in the United States." *Academic Medicine*, 91(7): 1007–1014.

Figure 1: Oncology has experienced large increases in available knowledge

(a) Pages in oncology clinical practice guidelines, by cancer type (2002–2020)

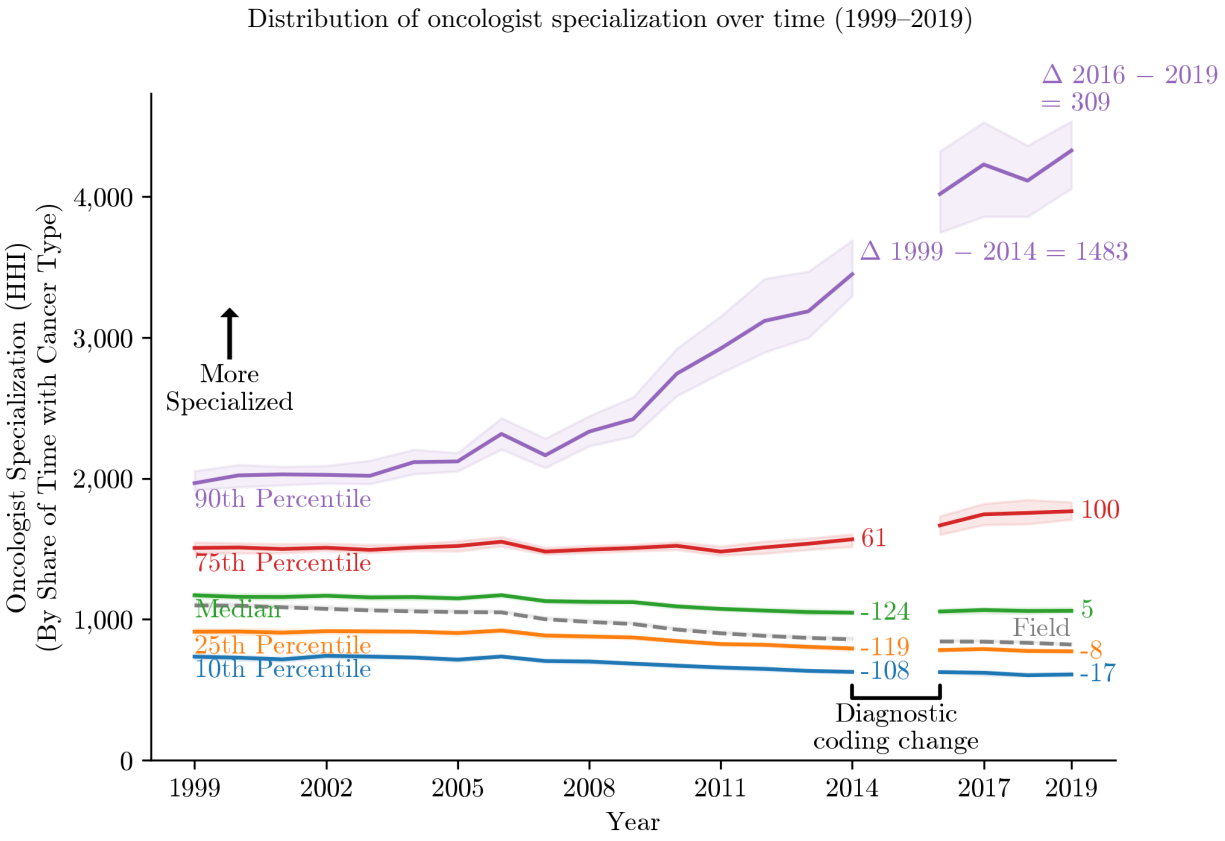



(b) Citations in oncology clinical practice guidelines, by cancer type (2002–2020)




*Notes*: Panels (a) and (b) show cancer guideline length and guideline citations over time. Panels (a) and (b) display the total page count and citation count of the first set of treatment, screening, and risk reduction guidelines in each year from National Comprehensive Cancer Network (NCCN) Guidelines. Each line shows one cancer type: colon, kidney, lung or skin.

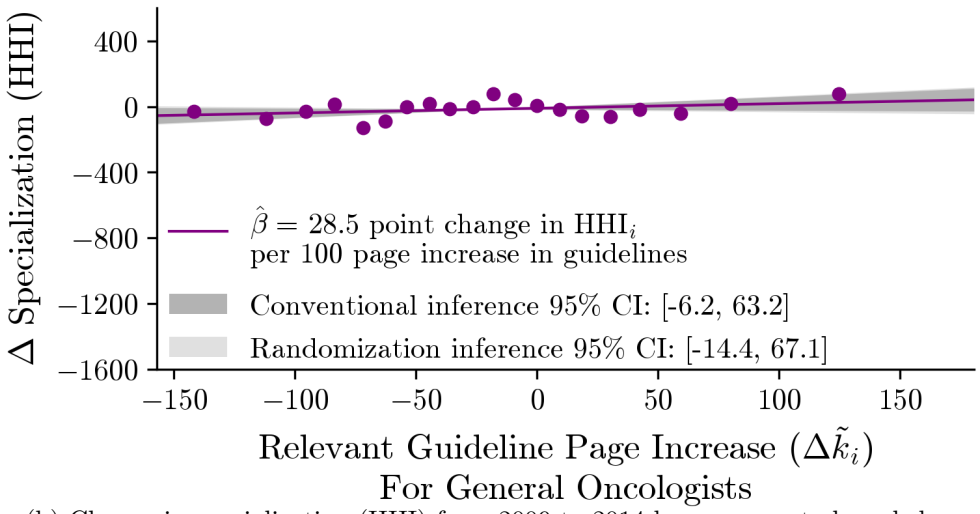
Figure 2: The most specialized oncologists are growing more specialized over time



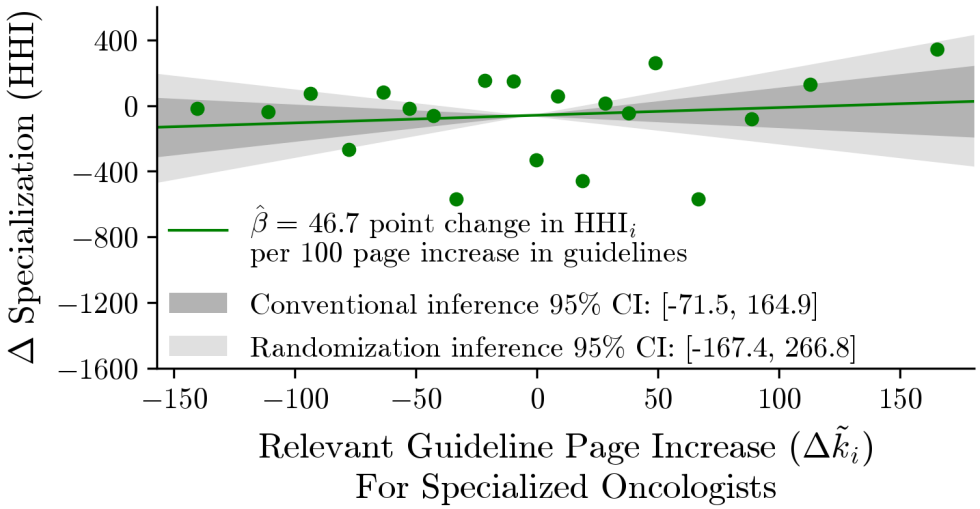

*Notes*: Figure 2 shows the distribution of oncologist specialization from 1999 to 2019. Specialization is measured using a Herfindahl–Hirschman Index (HHI) over the types of cancer that the oncologist evaluated and managed in the year, weighted by outpatient evaluation and management work Relative Value Units (wRVUs). HHI is jackknife-corrected as detailed in Appendix C.3. The 10th, 25th, 50th, 75th, and 90th percentiles are displayed and labeled. The numbers at the line ends give each percentile's change in HHI from 1999 to 2014 and from 2016 to 2019. 95% confidence intervals for each percentile are calculated by bootstrapping oncologists. The dashed grey line, labeled Field, shows the population-level HHI, the HHI of all cancer patients in the sample pooled together, with 95% confidence intervals from bootstrapping patients. All oncologists in the sample are hematologist-oncologists, medical oncologists, or hematologists who saw at least 25 unique cancer patients in the year in a 20% sample of Medicare Part B outpatient evaluation and management claims. The figure omits 2015, the year diagnosis coding switched from ICD-9 to ICD-10.

Figure 3: Generalists do not meaningfully increase their specialization with exposure to knowledge growth, and specialist estimates are imprecise

(a) Change in specialization (HHI) from 2009 to 2014 by exposure to knowledge growth for less specialized oncologists (bottom 75% of initial specialization)



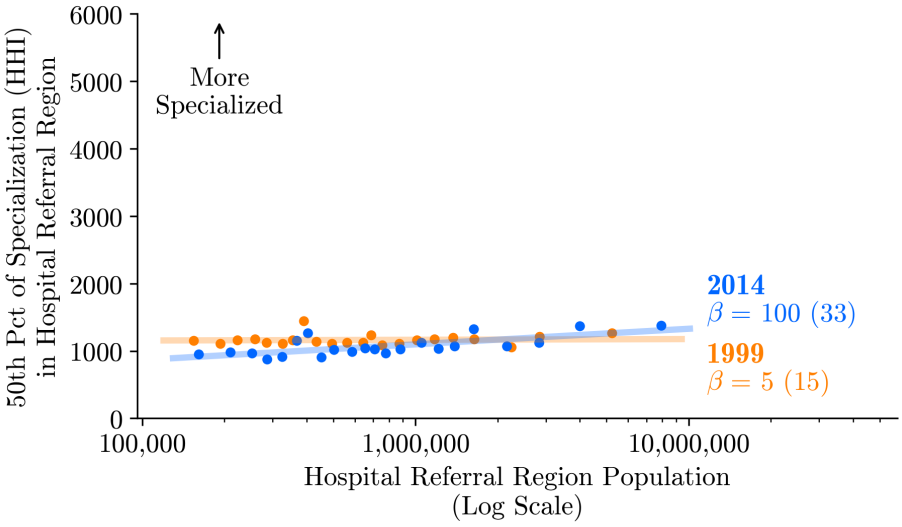

(b) Change in specialization (HHI) from 2009 to 2014 by exposure to knowledge growth for specialist oncologists (top 25% of initial specialization)



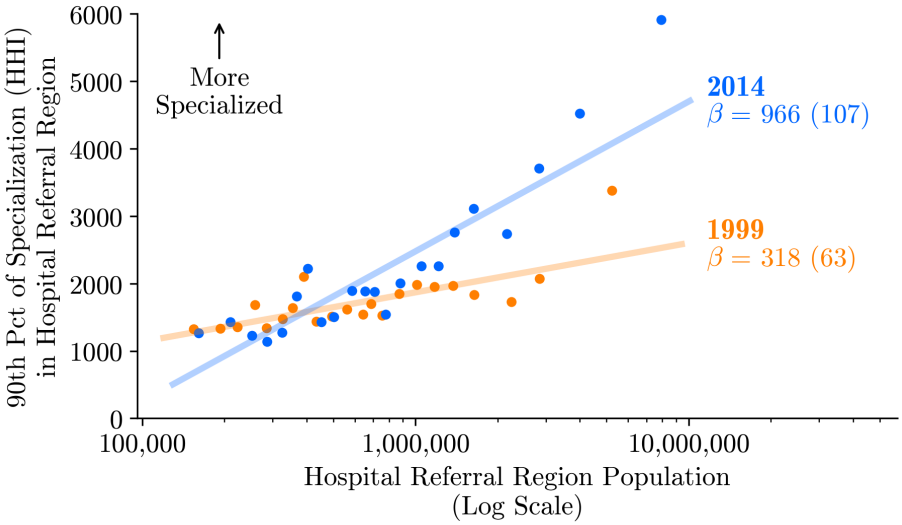

*Notes*: Panels (a) and (b) show binned scatterplots and regression estimates for the relationship between the recentered exposure to knowledge growth and change in specialization. Panel (a) shows this relationship for oncologists who start in the bottom 75% of the initial specialization distribution. Panel (b) shows this relationship for oncologists who start in the top 25% of the initial specialization distribution. Exposure to knowledge growth is recentered as described in Section 5, using the predicted knowledge growth regression, Equation (2). Specialization is measured using a Herfindahl–Hirschman Index (HHI) over the types of cancer that the oncologist evaluated and managed in the year, weighted by outpatient evaluation and management work Relative Value Units (wRVUs). HHI is jackknife-corrected as detailed in Appendix C.3. Shaded bands show 95% confidence intervals from conventional inference, with robust standard errors, and from randomization inference (RI), constructed as in Appendix D.5. All physicians in the sample are hematologist-oncologists, medical oncologists, or hematologists who saw at least 25 unique cancer patients in the year in a 20% sample of Medicare Part B outpatient evaluation and management claims in 2008, 2009, and 2014.

Figure 4: Median specialization in a market is relatively flat over time and across market size, but the 90th percentile of specialization is increasing in the largest markets over time

(a) Median specialization by market size (1999 and 2014)



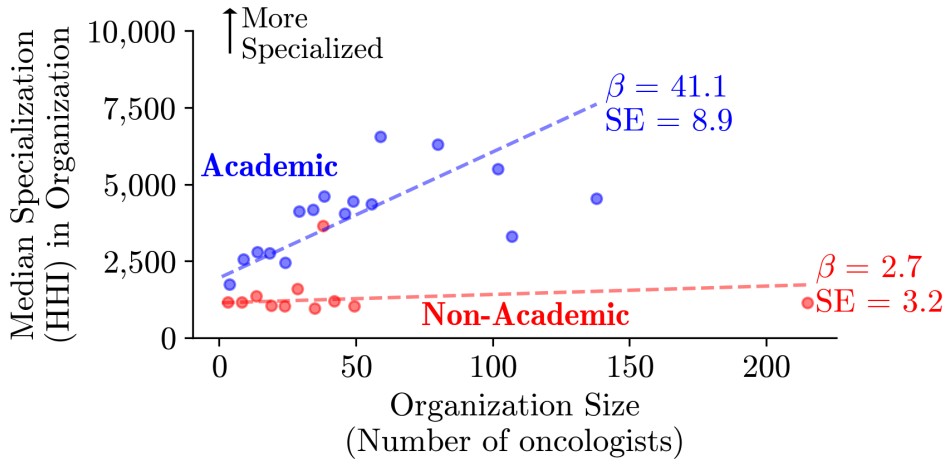

(b) 90th percentile of specialization by market size (1999 and 2014)



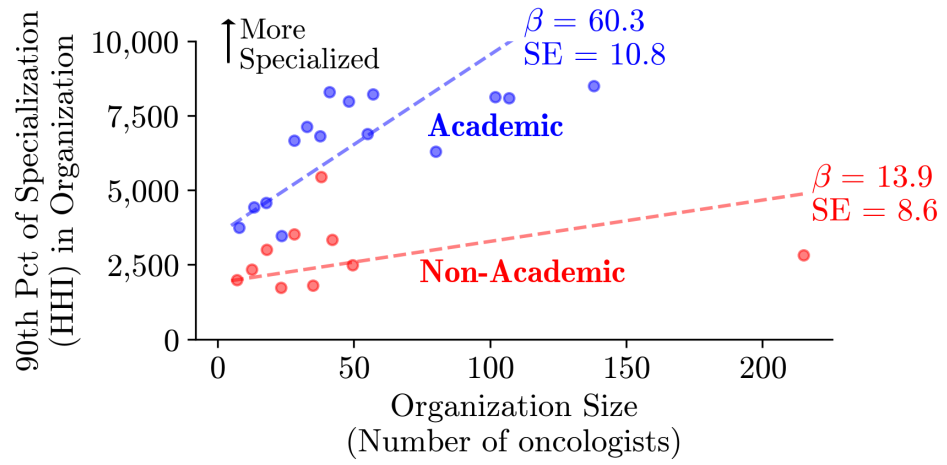

*Notes*: Panels (a) and (b) show binned scatterplots of the 50th and 90th percentile of oncologist specialization in a market by market size in 1999 and 2014. The x-axis shows the market size, measured as population, in the relevant year. Markets are defined using hospital referral regions. In Panel (a), the y-axis shows the 50th percentile of specialization in a market in a year. In Panel (b), the y-axis shows the 90th percentile of specialization in a market in a year. Specialization is measured using a Herfindahl–Hirschman Index (HHI) over the types of cancer that the oncologist evaluated and managed in the year, weighted by outpatient evaluation and management work Relative Value Units (wRVUs). HHI is jackknife-corrected as detailed in Appendix C.3. The lines plotted show the line of best fit between log population and the relevant percentile of specialization in each year. Each line's slope is reported as $\beta$, with the robust standard error in parentheses. All physicians in the sample are hematologist-oncologists, medical oncologists, or hematologists who saw at least 25 unique cancer patients in the year in a 20% sample of Medicare Part B outpatient evaluation and management claims.

Figure 5: Specialization rises with organization size in academic organizations but not in non-academic ones

(a) Median specialization by organization size and type (2014)

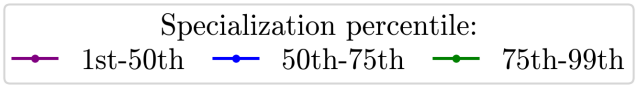



(b) 90th percentile of specialization by organization size and type (2014)

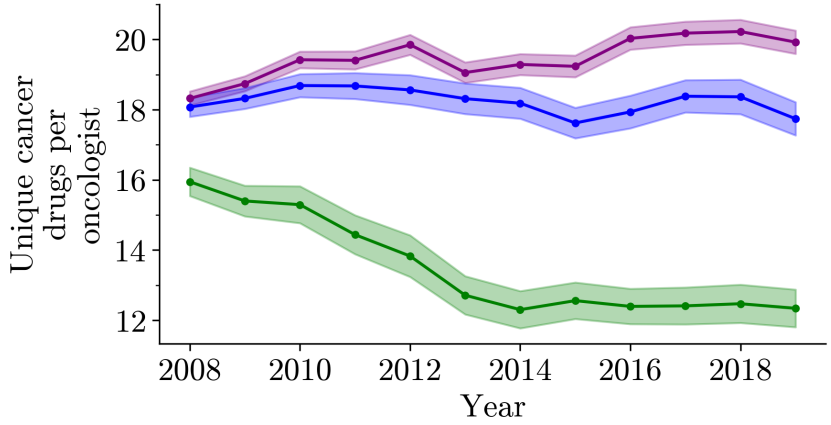



*Notes*: Panels (a) and (b) show binned scatterplots of the 50th and 90th percentile of oncologist specialization in an organization by organization size and type in 2014. Organization size is on the x-axis and is estimated as the number of oncologists at the organization's tax ID in the 2014 Medicare Data on Provider Practice and Specialty (MD-PPAS), using oncologists' primary tax ID. Organization type is determined using tax ID names from the MD-PPAS to categorize organizations as academic or non-academic. See Section 3.2 and Appendix B.3 for discussion of the limitations of these variables from the MD-PPAS. In Panel (a), the y-axis shows the 50th percentile of specialization in an organization. In Panel (b), the y-axis shows the 90th percentile of specialization in an organization, and the sample is restricted to organizations with at least 5 oncologists. Specialization is measured using a Herfindahl–Hirschman Index (HHI) over the types of cancer that the oncologist evaluated and managed in the year, weighted by outpatient evaluation and management work Relative Value Units (wRVUs). HHI is jackknife-corrected as detailed in Appendix C.3. The best fit line is plotted for the relationship between size and the relevant percentile of specialization in an organization. $\beta$ is each line's slope and SE its robust standard error. All physicians in the sample are hematologist-oncologists, medical oncologists, or hematologists who saw at least 25 unique cancer patients in the year in a 20% sample of Medicare Part B outpatient evaluation and management claims.

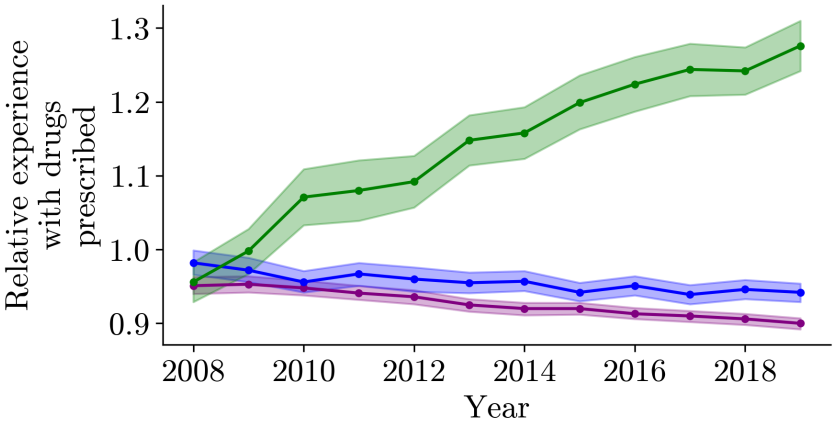
Figure 6: Over time, the most specialized oncologists prescribe fewer unique anti-cancer drugs and become more experienced with them than less specialized oncologists




(a) Count of unique cancer drugs prescribed by year and degree of specialization

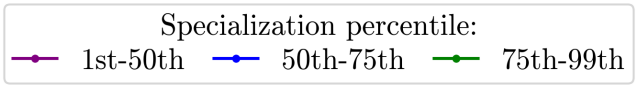



(b) Relative experience with cancer drugs prescribed by year and degree of specialization

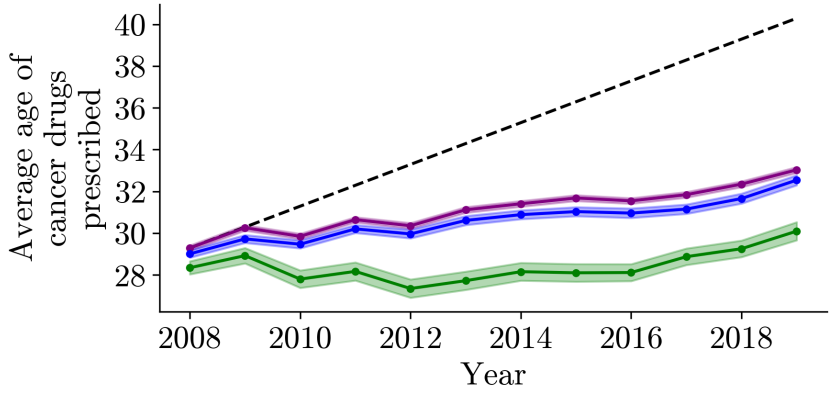



*Notes*: Panel (a) shows the average number of unique anti-cancer drugs that oncologists prescribe, and Panel (b) shows their average relative experience with those drugs, over time by level of specialization. Anti-cancer drugs are defined as unique World Health Organization Anatomical Therapeutic Chemical (ATC) codes that fall into the list of anti-cancer drugs from Pantziarka et al. (2021). Experience with a drug is calculated as the number of patients an oncologist prescribed a drug to in a year. Relative experience normalizes experience to have a mean of one within a drug-year. Because the figure weights oncologists equally rather than oncologist-drug pairs, its averages need not equal one. Specialization is measured using a Herfindahl–Hirschman Index (HHI) over the types of cancer that the oncologist evaluated and managed in the year, weighted by outpatient evaluation and management work Relative Value Units (wRVUs). HHI is jackknife-corrected as detailed in Appendix C.3. Oncologists are grouped into specialization bins by percentile of HHI by year. Shaded bands show 95% confidence intervals from robust standard errors. All physicians in the sample are hematologist-oncologists, medical oncologists, or hematologists who saw at least 11 unique cancer patients in a 20% sample of Medicare Part B outpatient evaluation and management claims and prescribed anti-cancer drugs to at least 11 patients in a 20% sample of Part B and Part D data in the year.

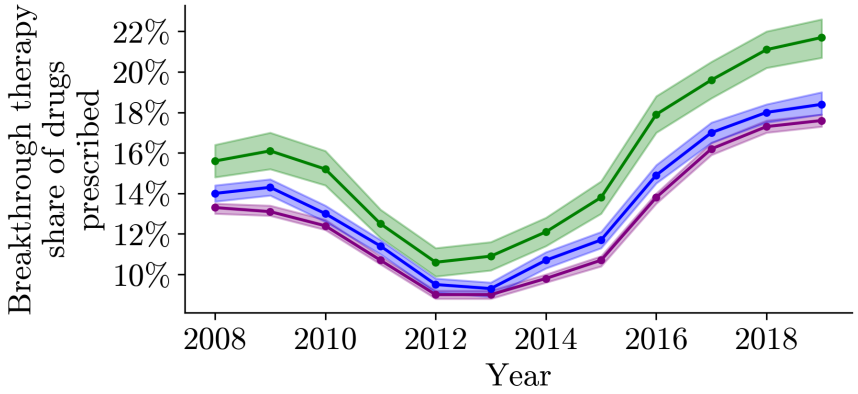
Figure 7: Over time, the most specialized oncologists prescribe newer drugs and a higher share of breakthrough-designated drugs than less specialized oncologists




(a) Average age of cancer drugs prescribed by year and degree of specialization




(b) Average share of breakthrough-designated drugs prescribed by year and degree of specialization




*Notes*: Panel (a) shows the average age of the anti-cancer drugs that oncologists prescribe, and Panel (b) shows the average share of those drugs that are breakthrough-designated, over time by level of specialization. Anti-cancer drugs are defined as unique World Health Organization Anatomical Therapeutic Chemical (ATC) codes that fall into the list of anti-cancer drugs from Pantziarka et al. (2021). Drug age is the number of years since the first U.S. Food and Drug Administration (FDA) approval for the drug's ATC code. In Panel (a), the dashed line rises by one year per year; it shows how average drug age would rise if oncologists kept prescribing the same drugs. A drug is classified as breakthrough-designated if the molecule, as of 2025, has ever received a "Breakthrough Therapy" designation from the FDA for any indication. Drugs are weighted by the number of patients the oncologist prescribed the drug to. Specialization is measured using a Herfindahl–Hirschman Index (HHI) over the types of cancer that the oncologist evaluated and managed in the year, weighted by outpatient evaluation and management work Relative Value Units (wRVUs). HHI is jackknife-corrected as detailed in Appendix C.3. Oncologists are grouped into specialization bins by percentile of HHI by year. Shaded bands show 95% confidence intervals from robust standard errors. All physicians in the sample are hematologist-oncologists, medical oncologists, or hematologists who saw at least 11 unique cancer patients in a 20% sample of Medicare Part B outpatient evaluation and management claims and prescribed anti-cancer drugs to at least 11 patients in a 20% sample of Part B and Part D data in the year.

Table 1: Specialists are more often academic and practice in larger markets and organizations than generalists

Summary statistics for the balanced panel of oncologists (2008, 2009, and 2014)

| | All (N = 3,425) | | | Generalists (N = 2,567) | | Specialists (N = 858) | |
|---|---|---|---|---|---|---|---|
| | Mean | SD | Range | Mean | SD | Mean | SD |
| Unique Patients | 57 | (26) | 25–276 | 59 | (27) | 53 | (24) |
| Cancer Types | 14 | (4) | 1–25 | 15 | (3) | 12 | (4) |
| HHI (2008) | 1,462 | (1,346) | 278–10,000 | 993 | (245) | 2,865 | (2,104) |
| HHI (2009) | 1,474 | (1,362) | 240–9,886 | 1,068 | (372) | 2,688 | (2,243) |
| HHI (2014) | 1,447 | (1,473) | 213–10,000 | 1,052 | (537) | 2,627 | (2,439) |
| Age | 49 | (9) | 32–77 | 49 | (8) | 50 | (9) |
| Female | 0.19 | (0.39) | 0.00–1.00 | 0.15 | (0.36) | 0.30 | (0.46) |
| Academic | 0.09 | (0.29) | 0.00–1.00 | 0.05 | (0.21) | 0.22 | (0.42) |
| In Largest 20% of Markets | 0.49 | (0.50) | 0.00–1.00 | 0.46 | (0.50) | 0.59 | (0.49) |
| Oncologists at Primary Organization | 18 | (40) | 1–219 | 16 | (38) | 23 | (44) |
| Relevant Guideline Pages (2008) | 980 | (197) | 42–1,503 | 1,030 | (142) | 829 | (255) |
| Relevant Guideline Pages (2009) | 1,120 | (221) | 42–1,689 | 1,177 | (156) | 951 | (289) |
| Change in Relevant Guideline Pages (2009–2014) | 1,065 | (221) | 24–1,606 | 1,121 | (162) | 901 | (285) |

*Notes*: Table 1 shows summary statistics for the balanced panel of oncologists observed in 2008, 2009, and 2014. Characteristics are measured in 2008 unless noted otherwise. Specialization is measured using a Herfindahl–Hirschman Index (HHI) over the types of cancer that the oncologist evaluated and managed in the year, weighted by outpatient evaluation and management work Relative Value Units (wRVUs). HHI is jackknife-corrected as detailed in Appendix C.3. Generalists are the oncologists below the 75th percentile of the 2008 HHI distribution, and specialists are those at or above it. Markets are defined using hospital referral regions (HRRs), and market size is measured using the HRR's population in 2008. Age, sex, academic affiliation and organization size come from the 2008 Medicare Data on Provider Practice and Specialty (MD-PPAS), discussed in Section 3.2. All oncologists in the sample are hematologist-oncologists, medical oncologists, or hematologists who saw at least 25 unique cancer patients in a 20% sample of Medicare Part B outpatient evaluation and management claims in each of 2008, 2009, and 2014. More information about each variable is detailed in Section 3.

Table 2: Guidelines grow faster for cancer types with more patients and surrogate endpoints

Poisson regressions of the page count increase (2009–2014) on the log patient count in Medicare (2008) and surrogate endpoints, by cancer type

| | Page Count Increase | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Log Patient Count in Medicare (2008) | 0.279 | 0.298 | 0.248 |
| | (0.0686) | (0.0656) | (0.0524) |
| Surrogate Endpoints | | 1.001 | −4.286 |
| | | (0.368) | (2.040) |
| Surrogate Endpoints × Log Patient Count (2008) | | | 0.566 |
| | | | (0.207) |
| Observations | 35 | 35 | 35 |
| R-squared | 0.166 | 0.484 | 0.714 |
| Estimation Method | PPML | PPML | PPML |

*Notes*: Table 2 reports regressions by cancer type of the page count increase from 2009 to 2014 on the log patient count in Medicare in 2008 (column 1), adding an indicator for whether the cancer type has surrogate endpoints (column 2) and the interaction of the two (column 3). The unit of observation is a cancer type, e.g., breast cancer. The page count increase is the change in page count between the first version of each National Comprehensive Cancer Network (NCCN) Guideline released in 2009 and in 2014. The presence of surrogate endpoints is determined using Budish, Roin, and Williams (2015). The patient count in Medicare is the number of patients whose primary cancer is the cancer type in a 20% sample of 2008 Medicare Fee-For-Service Part B claims, using ICD-9 codes. Cancer types are defined using a custom grouping of ICD-9 codes that mirrors the organization of NCCN Guidelines, as detailed in Appendix B.1. The regressions are estimated by Poisson pseudo-maximum likelihood (PPML), with robust standard errors in parentheses; R-squared is the squared correlation between fitted and actual page count increases.

Table 3: After recentering, physician characteristics no longer jointly predict exposure to knowledge growth (generalists, bottom 75% of initial HHI)

Regressions of generalists' exposure to knowledge growth (2009–2014) on characteristics in 2008

| | (1) Unadjusted exposure ($\Delta k_i$) | (2) Remove physician scope variation | (3) Remove cancer prevalence variation | (4) Remove endpoint type variation ($\Delta \tilde{k}_i$) |
|---|---|---|---|---|
| Market population (millions) | −4.898 | −0.494 | −1.292 | −0.574 |
| Robust SE | (1.377) | (0.855) | (0.924) | (0.753) |
| Placebo interval | [−3.65, 3.45] | [−3.24, 3.26] | [−2.42, 2.56] | [−1.64, 1.56] |
| Oncologists at primary organization | 0.0325 | −0.0218 | −0.0229 | −0.0218 |
| Robust SE | (0.0614) | (0.0348) | (0.0371) | (0.0325) |
| Placebo interval | [−0.174, 0.173] | [−0.178, 0.177] | [−0.126, 0.123] | [−0.0789, 0.0723] |
| Physician age | −0.786 | −0.0420 | −0.0893 | −0.406 |
| Robust SE | (0.287) | (0.166) | (0.185) | (0.156) |
| Placebo interval | [−0.744, 0.715] | [−0.936, 0.758] | [−0.735, 0.625] | [−0.398, 0.427] |
| Patient count | 3.862 | 0.199 | −0.0607 | 0.120 |
| Robust SE | (0.118) | (0.0478) | (0.0490) | (0.0454) |
| Placebo interval | [−0.235, 0.234] | [−0.887, 0.938] | [−0.835, 0.804] | [−0.545, 0.517] |
| Female | −24.53 | −4.108 | 0.605 | 0.758 |
| Robust SE | (7.013) | (4.176) | (4.552) | (3.696) |
| Placebo interval | [−17.5, 17.5] | [−17, 15.6] | [−13.7, 12.6] | [−7.97, 8.39] |
| Academic | −15.01 | −12.60 | −5.308 | 12.60 |
| Robust SE | (14.13) | (9.120) | (8.512) | (6.540) |
| Placebo interval | [−29.9, 29.8] | [−30.6, 30.1] | [−24.5, 23.2] | [−15.7, 16] |
| Joint test $p$-value | 0.001 | 0.595 | 0.980 | 0.708 |
| Observations | 2,561 | 2,561 | 2,561 | 2,561 |
| $R^2$ | 0.423 | 0.009 | 0.002 | 0.006 |

*Notes*: Table 3 shows regression estimates for the relationship between physician characteristics and exposure to knowledge growth for oncologists who start in the bottom 75% of the initial specialization distribution. Specialization is measured using a Herfindahl–Hirschman Index (HHI) over the types of cancer that the oncologist managed in the year, weighted by outpatient evaluation and management work Relative Value Units (wRVUs). HHI is jackknife-corrected as detailed in Appendix C.3. Column 1 uses unadjusted exposure to knowledge growth, and columns 2 to 4 use recentered exposure to knowledge growth, in which predicted exposure is subtracted from unadjusted exposure. In column 2, predicted knowledge growth is the mean across cancer types. In column 3, predicted knowledge growth is estimated in a Poisson regression on the log patient count in Medicare in 2008. In column 4, predicted knowledge growth is estimated in a Poisson regression on the log patient count, an indicator for well-established surrogate endpoints, and their interaction, as in Equation (2). Unadjusted exposure to knowledge growth is calculated as the total increase in the page count of NCCN Guidelines over 2009 to 2014 for the types of cancers a physician treats in 2008. Standard errors are robust. The placebo intervals and the joint test $p$-values come from randomization inference, detailed in Appendix D. A coefficient outside its placebo interval predicts exposure more strongly than 95% of placebo draws (two-sided $p$-value $\leq 0.05$); the joint test asks the same of all six characteristics together. Age, sex, academic affiliation and organization size come from the 2008 Medicare Data on Provider Practice and Specialty (MD-PPAS), discussed in Section 3.2. All physicians in the sample are hematologist-oncologists, medical oncologists, or hematologists who saw at least 25 unique cancer patients in a 20% sample of Medicare Part B outpatient evaluation and management claims in 2008, 2009, and 2014.

Table 4: After recentering, physician characteristics no longer jointly predict exposure to knowledge growth (specialists, top 25% of initial HHI)

Regressions of specialists' exposure to knowledge growth (2009–2014) on characteristics in 2008

| | (1) Unadjusted exposure ($\Delta k_i$) | (2) Remove physician scope variation | (3) Remove cancer prevalence variation | (4) Remove endpoint type variation ($\Delta \tilde{k}_i$) |
|---|---|---|---|---|
| Market population (millions) | −23.26 | −9.305 | −4.518 | 0.0781 |
| Robust SE | (5.067) | (2.006) | (1.874) | (1.210) |
| Placebo interval | [−8.98, 9.06] | [−6.39, 5.81] | [−5.24, 5.46] | [−4.57, 4.6] |
| Oncologists at primary organization | −0.523 | −0.148 | 0.0347 | 0.0342 |
| Robust SE | (0.212) | (0.0992) | (0.0880) | (0.0515) |
| Placebo interval | [−0.455, 0.422] | [−0.244, 0.26] | [−0.22, 0.211] | [−0.131, 0.14] |
| Physician age | −1.249 | 0.152 | −0.639 | −0.513 |
| Robust SE | (0.910) | (0.437) | (0.424) | (0.293) |
| Placebo interval | [−2.24, 2.26] | [−1.47, 1.63] | [−1.33, 1.9] | [−0.774, 0.833] |
| Patient count | 4.780 | 0.345 | −0.352 | −0.329 |
| Robust SE | (0.413) | (0.145) | (0.134) | (0.102) |
| Placebo interval | [−0.823, 0.901] | [−1.03, 1.09] | [−1, 0.992] | [−0.681, 0.679] |
| Female | −56.83 | −18.49 | −19.44 | 0.912 |
| Robust SE | (17.05) | (8.689) | (8.141) | (5.387) |
| Placebo interval | [−46, 41.6] | [−30.2, 32] | [−28.5, 28.8] | [−20.3, 20.2] |
| Academic | −232.2 | −107.8 | −39.73 | 1.577 |
| Robust SE | (23.49) | (11.96) | (10.59) | (6.671) |
| Placebo interval | [−46.8, 47.9] | [−83.8, 58.8] | [−86.4, 60.3] | [−55, 53.7] |
| Joint test $p$-value | 0.001 | 0.015 | 0.530 | 0.857 |
| Observations | 854 | 854 | 854 | 854 |
| $R^2$ | 0.420 | 0.212 | 0.045 | 0.018 |

*Notes*: Table 4 shows regression estimates for the relationship between physician characteristics and exposure to knowledge growth for oncologists who start in the top 25% of the initial specialization distribution. Specialization is measured using a Herfindahl–Hirschman Index (HHI) over the types of cancer that the oncologist managed in the year, weighted by outpatient evaluation and management work Relative Value Units (wRVUs). HHI is jackknife-corrected as detailed in Appendix C.3. Column 1 uses unadjusted exposure to knowledge growth, and columns 2 to 4 use recentered exposure to knowledge growth, in which predicted exposure is subtracted from unadjusted exposure. In column 2, predicted knowledge growth is the mean across cancer types. In column 3, predicted knowledge growth is estimated in a Poisson regression on the log patient count in Medicare in 2008. In column 4, predicted knowledge growth is estimated in a Poisson regression on the log patient count, an indicator for well-established surrogate endpoints, and their interaction, as in Equation (2). Unadjusted exposure to knowledge growth is calculated as the total increase in the page count of NCCN Guidelines over 2009 to 2014 for the types of cancers a physician treats in 2008. Standard errors are robust. The placebo intervals and the joint test $p$-values come from randomization inference, detailed in Appendix D. A coefficient outside its placebo interval predicts exposure more strongly than 95% of placebo draws (two-sided $p$-value $\leq 0.05$); the joint test asks the same of all six characteristics together. Age, sex, academic affiliation and organization size come from the 2008 Medicare Data on Provider Practice and Specialty (MD-PPAS), discussed in Section 3.2. All physicians in the sample are hematologist-oncologists, medical oncologists, or hematologists who saw at least 25 unique cancer patients in a 20% sample of Medicare Part B outpatient evaluation and management claims in 2008, 2009, and 2014.

Table 5: Generalists do not meaningfully increase their specialization in response to knowledge growth, and specialist estimates are imprecise

Regressions of the change in HHI (2009–2014) on recentered exposure to knowledge growth

| | Change in HHI | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| *Panel A. Generalists (bottom 75% of initial specialization)* | | | |
| Knowledge growth (100 pages) | 28.52 | 28.82 | 29.22 |
| Robust SE | (17.69) | (17.88) | (19.57) |
| RI 95% CI | [−14.4, 67.1] | [−16, 68.6] | [−10.9, 69.6] |
| Observations | 2,567 | 2,531 | 2,531 |
| R-squared | 0.001 | 0.001 | 0.123 |
| *Panel B. Specialists (top 25% of initial specialization)* | | | |
| Knowledge growth (100 pages) | 46.70 | 70.26 | 77.96 |
| Robust SE | (60.22) | (63.31) | (75.87) |
| RI 95% CI | [−167, 267] | [−161, 308] | [−171, 386] |
| Observations | 858 | 779 | 779 |
| R-squared | 0.001 | 0.002 | 0.164 |
| Estimation sample | Full | Restricted | Restricted |
| Age FE | | | Yes |
| Market FE | | | Yes |
| Patient count FE | | | Yes |
| Organization size FE | | | Yes |
| Sex FE | | | Yes |
| Academic affiliation FE | | | Yes |

*Notes*: Table 5 shows estimates of the estimating equation in Section 5.1, a regression of the change in specialization from 2009 to 2014 on exposure to knowledge growth. Specialization is measured using a Herfindahl–Hirschman Index (HHI) over the types of cancer that the oncologist evaluated and managed in the year, weighted by outpatient evaluation and management work Relative Value Units (wRVUs). HHI is jackknife-corrected as detailed in Appendix C.3. Generalists are the oncologists below the 75th percentile of the 2008 HHI distribution, and specialists are those at or above it. Robust standard errors are in parentheses. Brackets show 95% confidence intervals from randomization inference (RI), constructed as in Appendix D.5. Exposure to knowledge growth is recentered as described in Section 5, using the predicted knowledge growth regression, Equation (2). Column 1 is the univariate regression on the full sample. Column 2 restricts the sample to physicians with observed, non-singleton fixed effects for the following 2008 variables: age group, sex, market (hospital referral region), decile of patient count, organization size (the number of oncologists at their primary organization), and academic affiliation. Column 3 adds these fixed effects. All physicians in the sample are hematologist-oncologists, medical oncologists, or hematologists who saw at least 25 unique cancer patients in a 20% sample of Medicare Part B outpatient evaluation and management claims in 2008, 2009, and 2014.

Table 6: Generalists do not meaningfully change the count of cancer types they treat in response to knowledge growth, and specialist estimates are imprecise

Regressions of the change in cancer type count (2009–2014) on recentered exposure to knowledge growth

| | Change in cancer type count | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| *Panel A. Generalists (bottom 75% of initial specialization)* | | | |
| Knowledge growth (100 pages) | −0.0849 | −0.116 | −0.165 |
| Robust SE | (0.0966) | (0.0971) | (0.0993) |
| RI 95% CI | [−0.596, 0.333] | [−0.627, 0.299] | [−0.557, 0.172] |
| Observations | 2,567 | 2,531 | 2,531 |
| R-squared | 0.000 | 0.001 | 0.210 |
| *Panel B. Specialists (top 25% of initial specialization)* | | | |
| Knowledge growth (100 pages) | −0.127 | −0.207 | −0.367 |
| Robust SE | (0.147) | (0.153) | (0.170) |
| RI 95% CI | [−0.849, 0.356] | [−1.01, 0.314] | [−1.22, 0.176] |
| Observations | 858 | 779 | 779 |
| R-squared | 0.001 | 0.002 | 0.216 |
| Estimation sample | Full | Restricted | Restricted |
| Age FE | | | Yes |
| Market FE | | | Yes |
| Patient count FE | | | Yes |
| Organization size FE | | | Yes |
| Sex FE | | | Yes |
| Academic affiliation FE | | | Yes |

*Notes*: Table 6 shows estimates of the estimating equation in Section 5.1, with the change from 2009 to 2014 in the count of cancer types a physician treats as the outcome. A cancer type is counted if at least one of the physician's patients in each year's outpatient evaluation and management claims has it as their primary cancer type. Specialization is measured using a Herfindahl–Hirschman Index (HHI) over the types of cancer that the oncologist evaluated and managed in the year, weighted by outpatient evaluation and management work Relative Value Units (wRVUs). HHI is jackknife-corrected as detailed in Appendix C.3. Generalists are the oncologists below the 75th percentile of the 2008 HHI distribution, and specialists are those at or above it. Robust standard errors are in parentheses. Brackets show 95% confidence intervals from randomization inference (RI), constructed as in Appendix D.5. Exposure to knowledge growth is recentered as described in Section 5, using the predicted knowledge growth regression, Equation (2). Column 1 is the univariate regression on the full sample. Column 2 restricts the sample to physicians with observed, non-singleton fixed effects for the following 2008 variables: age group, sex, market (hospital referral region), decile of patient count, organization size (the number of oncologists at their primary organization), and academic affiliation. Column 3 adds these fixed effects. All physicians in the sample are hematologist-oncologists, medical oncologists, or hematologists who saw at least 25 unique cancer patients in a 20% sample of Medicare Part B outpatient evaluation and management claims in 2008, 2009, and 2014.

Table 7: The growing differences in prescribing outcomes between the most and least specialized oncologists are driven by the change in the distribution of specialization (HHI)

Decomposition of 2008–2019 change in prescribing differences between the most and least specialized oncologists

| | Share of change in the difference between the most and least specialized oncologists, by prescribing outcome | | |
|---|---|---|---|
| | Drug age | Relative experience | Unique drugs |
| From the change in the distribution of HHI | 84% | 55% | 67% |
| From the change in outcomes at a given HHI | 10% | 41% | 37% |
| Remainder | 6% | 4% | −4% |

*Notes*: Table 7 shows a decomposition of the 2008–2019 change in the difference in prescribing outcomes between the most and least specialized oncologists (Figures 6–7) into three parts: the change in the distribution of specialization, the change in outcomes at a given specialization, and a remainder. The outcomes are average drug age (Column 1), average relative experience with drugs prescribed (Column 2), and the number of unique anti-cancer drugs prescribed (Column 3), defined as in Figures 7a, 6b, and 6a. Specialization is measured using a Herfindahl–Hirschman Index (HHI) over the types of cancer that the oncologist evaluated and managed in the year, weighted by outpatient evaluation and management work Relative Value Units (wRVUs). HHI is jackknife-corrected as detailed in Appendix C.3. The decomposition is calculated as follows. In each year the most specialized oncologists, denoted $H$, are those at or above the 75th percentile of HHI among all oncologists with at least 11 cancer patients, and the least specialized oncologists, denoted $L$, are those below the 50th percentile, as in Figures 6–7. The index $g \in \{H, L\}$ denotes the most specialized ($H$) or the least specialized ($L$) oncologists. Oncologists are sorted into ten 1,000-point HHI bins indexed by $b$. In year $t$, $\bar{y}_{bt}$ is the mean outcome of all oncologists in bin $b$, $s^g_{bt}$ is the share of $g$'s oncologists in bin $b$, and $\bar{y}^g_t$ is the mean outcome of $g$. The 2008–2019 change in $g$'s mean outcome, $\Delta\bar{y}^g \equiv \bar{y}^g_{2019} - \bar{y}^g_{2008}$, is written as the sum of three terms, $\Delta\bar{y}^g = S^g + O^g + R^g$. $S^g$ captures the change in the distribution of HHI, with bin means held at their 2019 values, and is calculated as $S^g = \sum_b \left(s^g_{b,2019} - s^g_{b,2008}\right) \bar{y}_{b,2019}$. $O^g$ captures the change in outcomes at a given HHI, with $g$'s bin shares held at their 2008 values, and is calculated as $O^g = \sum_b s^g_{b,2008} \left(\bar{y}_{b,2019} - \bar{y}_{b,2008}\right)$. $R^g$ is the remainder, which is not zero because the mean outcome of $g$'s oncologists in a bin can differ from the mean outcome of all oncologists in that bin, and is calculated as $R^g = \Delta\bar{y}^g - S^g - O^g$. The 2008–2019 change in the difference between the most and least specialized oncologists is $D \equiv \Delta\bar{y}^H - \Delta\bar{y}^L$. The first row reports $100\,(S^H - S^L)/D$, the second row $100\,(O^H - O^L)/D$, and the third row $100\,(R^H - R^L)/D$, so the three rows sum to 100%. All physicians in the sample are hematologist-oncologists, medical oncologists, or hematologists who saw at least 11 unique cancer patients in a 20% sample of Medicare Part B outpatient evaluation and management claims and prescribed anti-cancer drugs to at least 11 patients in a 20% sample of Part B and Part D data in the year.

# Appendix – For Online Publication

Knowledge Growth and Specialization: Evidence from Oncologists

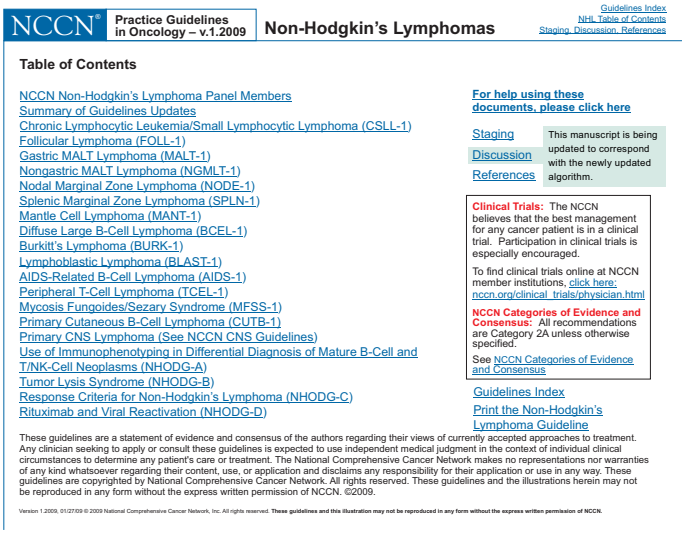
Maya Lozinski

October 2026

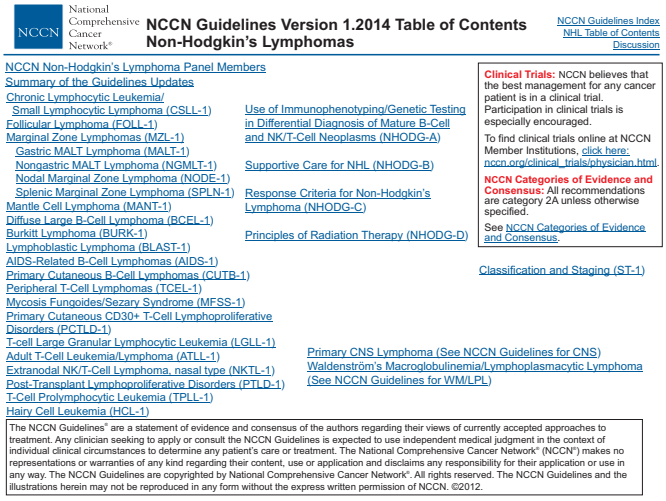
## A Additional exhibits

Figure A.1: The non-Hodgkin lymphoma guideline added sections and subtypes over time

(a) 2009 Table of Contents



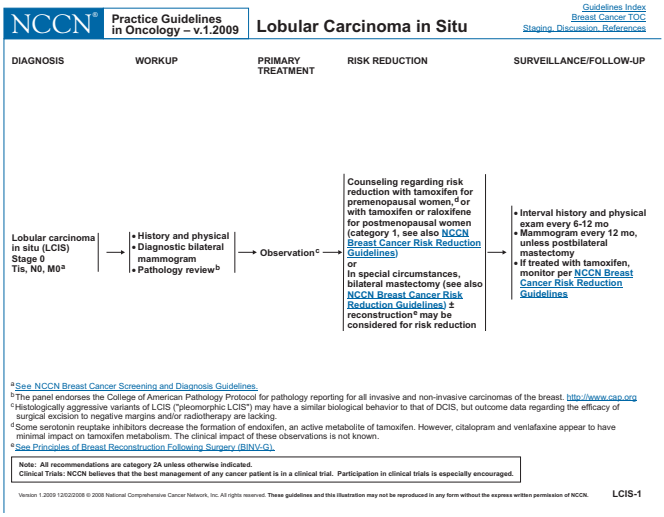

(b) 2014 Table of Contents



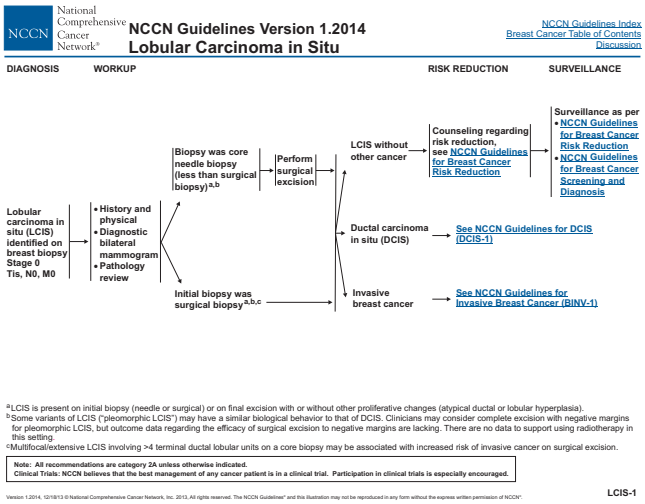

*Notes*: Figure A.1 shows example changes from National Comprehensive Cancer Network (NCCN) Guidelines. Panels (a) and (b) show the table of contents for non-Hodgkin lymphoma (NHL) in 2009 and 2014, respectively. Figure A.1a reproduced with permission from the NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®) for Non-Hodgkin's Lymphomas V.1.2009 © 2009 National Comprehensive Cancer Network, Inc. All rights reserved. The NCCN Guidelines® and illustrations herein may not be reproduced in any form for any purpose without the express written permission of NCCN. To view the most recent and complete version of the NCCN Guidelines, go online to NCCN.org. The NCCN Guidelines are a work in progress that may be refined as often as new significant data becomes available. Figure A.1b reproduced with permission from the NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®) for Non-Hodgkin's Lymphomas V.1.2014. © 2013 National Comprehensive Cancer Network, Inc. All rights reserved. The NCCN Guidelines® and illustrations herein may not be reproduced in any form for any purpose without the express written permission of NCCN. To view the most recent and complete version of the NCCN Guidelines, go online to NCCN.org. The NCCN Guidelines are a work in progress that may be refined as often as new significant data becomes available. NCCN makes no warranties of any kind whatsoever regarding their content, use or application and disclaims any responsibility for their application or use in any way.

Figure A.2: The lobular carcinoma in situ workup flowchart added decision branches

(a) 2009 Flowchart

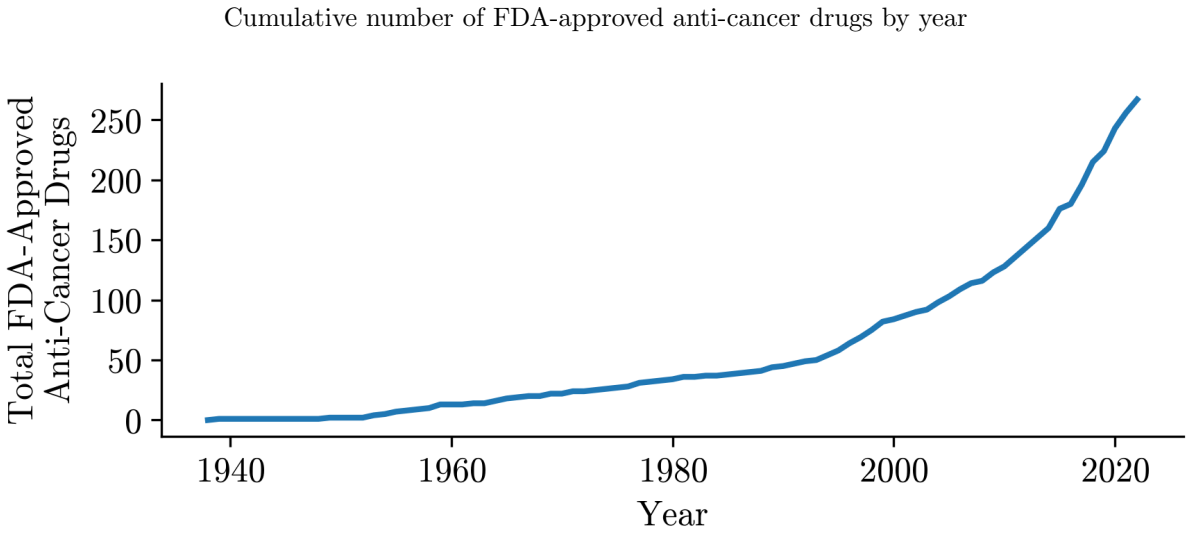



(b) 2014 Flowchart




*Notes*: Figure A.2 shows example changes from National Comprehensive Cancer Network (NCCN) Guidelines. Panels (a) and (b) show a flowchart with the diagnostic workup for lobular carcinoma in situ, a subtype of breast cancer, in 2009 and 2014, respectively. Figure A.2a reproduced with permission from the NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®) for Breast Cancer V.1.2009. © 2008 National Comprehensive Cancer Network, Inc. All rights reserved. The NCCN Guidelines® and illustrations herein may not be reproduced in any form for any purpose without the express written permission of NCCN. To view the most recent and complete version of the NCCN Guidelines, go online to NCCN.org. The NCCN Guidelines are a work in progress that may be refined as often as new significant data becomes available. Figure A.2b reproduced with permission from the NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®) for Breast Cancer V.1.2014. © 2013 National Comprehensive Cancer Network, Inc. All rights reserved. The NCCN Guidelines® and illustrations herein may not be reproduced in any form for any purpose without the express written permission of NCCN. To view the most recent and complete version of the NCCN Guidelines, go online to NCCN.org. The NCCN Guidelines are a work in progress that may be refined as often as new significant data becomes available. NCCN makes no warranties of any kind whatsoever regarding their content, use or application and disclaims any responsibility for their application or use in any way.

Figure A.3: The number of FDA-approved anti-cancer drugs has increased substantially




*Notes*: Figure A.3 shows the cumulative number of unique anti-cancer drugs by year in the open-access database of licensed cancer drugs of Pantziarka et al. (2021). The sample is limited to drugs with direct anti-cancer uses that are approved by the U.S. Food and Drug Administration (FDA). It does not include combinations of drugs, drugs used for cancer symptom relief or diagnosis, or unapproved drugs that are being investigated.

Table A.1: The length of cancer treatment guidelines has increased substantially over time for many types of cancer

Page count of treatment guidelines by cancer type (2002, 2009, 2014, and 2020)

| Cancer type | 2002 | 2009 | 2014 | 2020 |
|---|---|---|---|---|
| Anus | 9 | 22 | 32 | 55 |
| Bladder | 40 | 47 | 65 | 113 |
| Brain and Spinal Cord | 51 | 70 | 118 | 165 |
| Breast | 107 | 200 | 301 | 364 |
| Cervix | 24 | 31 | 67 | 86 |
| Colorectal | 38 | 145 | 217 | 355 |
| Esophagus | 19 | 48 | 125 | 154 |
| Gastric and GE Junction | 18 | 35 | 103 | 125 |
| Head and Neck | 86 | 116 | 182 | 233 |
| Hodgkin Lymphoma | 30 | 39 | 74 | 85 |
| Kidney | 14 | 20 | 35 | 64 |
| Liver and Biliary Tree | 32 | 35 | 110 | 151 |
| Lung and Bronchus | 68 | 107 | 188 | 294 |
| Lymphocytic Leukemia | 0 | 0 | 96 | 141 |
| Mast Cell | 0 | 0 | 0 | 67 |
| Mediastinum | 0 | 0 | 30 | 41 |
| Mesothelioma | 0 | 0 | 30 | 46 |
| Multiple Myeloma | 27 | 42 | 66 | 95 |
| Myelodysplastic and Myeloproliferative Disorders | 22 | 32 | 53 | 192 |
| Myeloid Leukemia | 47 | 83 | 168 | 218 |
| Neuroendocrine | 46 | 64 | 98 | 141 |
| Non-Hodgkin Lymphoma | 49 | 155 | 471 | 658 |
| Occult | 30 | 43 | 55 | 65 |
| Ovary and Adnexa | 26 | 44 | 83 | 127 |
| Pancreas | 23 | 55 | 109 | 160 |
| Placenta | 0 | 0 | 0 | 40 |
| Prostate | 40 | 77 | 120 | 230 |
| Sarcoma | 56 | 92 | 196 | 281 |
| Skin | 58 | 82 | 144 | 465 |
| Small Intestine | 0 | 0 | 0 | 48 |
| Testes | 34 | 30 | 55 | 78 |
| Thyroid | 69 | 73 | 117 | 136 |
| Uterus | 42 | 45 | 81 | 108 |
| All | **1,105** | **1,832** | **3,589** | **5,581** |

*Notes*: Table A.1 reports the page count of the first set of treatment guidelines in 2002, 2009, 2014, and 2020 from National Comprehensive Cancer Network (NCCN) Guidelines.

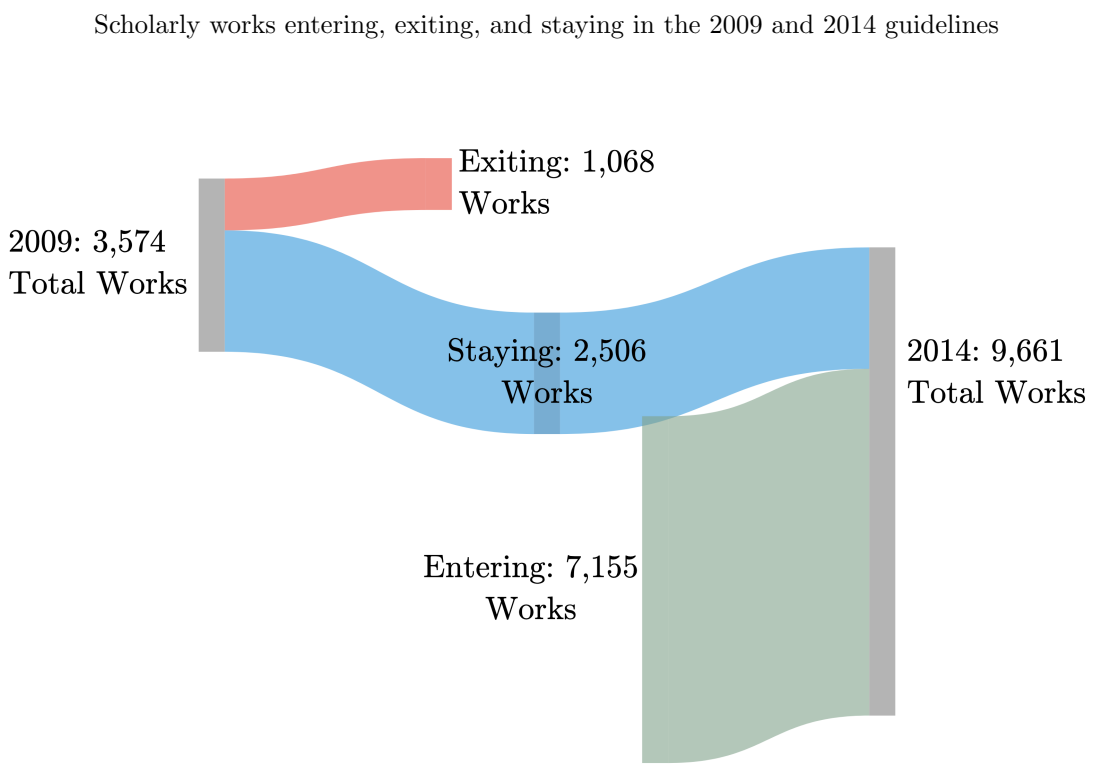
Table A.2: Guideline length is closely associated with the number of citations in guidelines

Regressions of page count on citation count by cancer type (2009, 2014 and 2009–2014)

| | (1) | (2) | (3) |
|---|---|---|---|
| | Page Count (2009) | Page Count (2014) | Change in Page Count (2009–2014) |
| Citation count (2009) | 0.357 | | |
| | (0.0249) | | |
| Citation count (2014) | | 0.274 | |
| | | (0.00935) | |
| Change in citation count (2009–2014) | | | 0.213 |
| | | | (0.00414) |
| Observations | 36 | 36 | 36 |
| R-squared | 0.885 | 0.954 | 0.941 |
| Estimator | OLS | OLS | OLS |
| Standard errors | Robust | Robust | Robust |

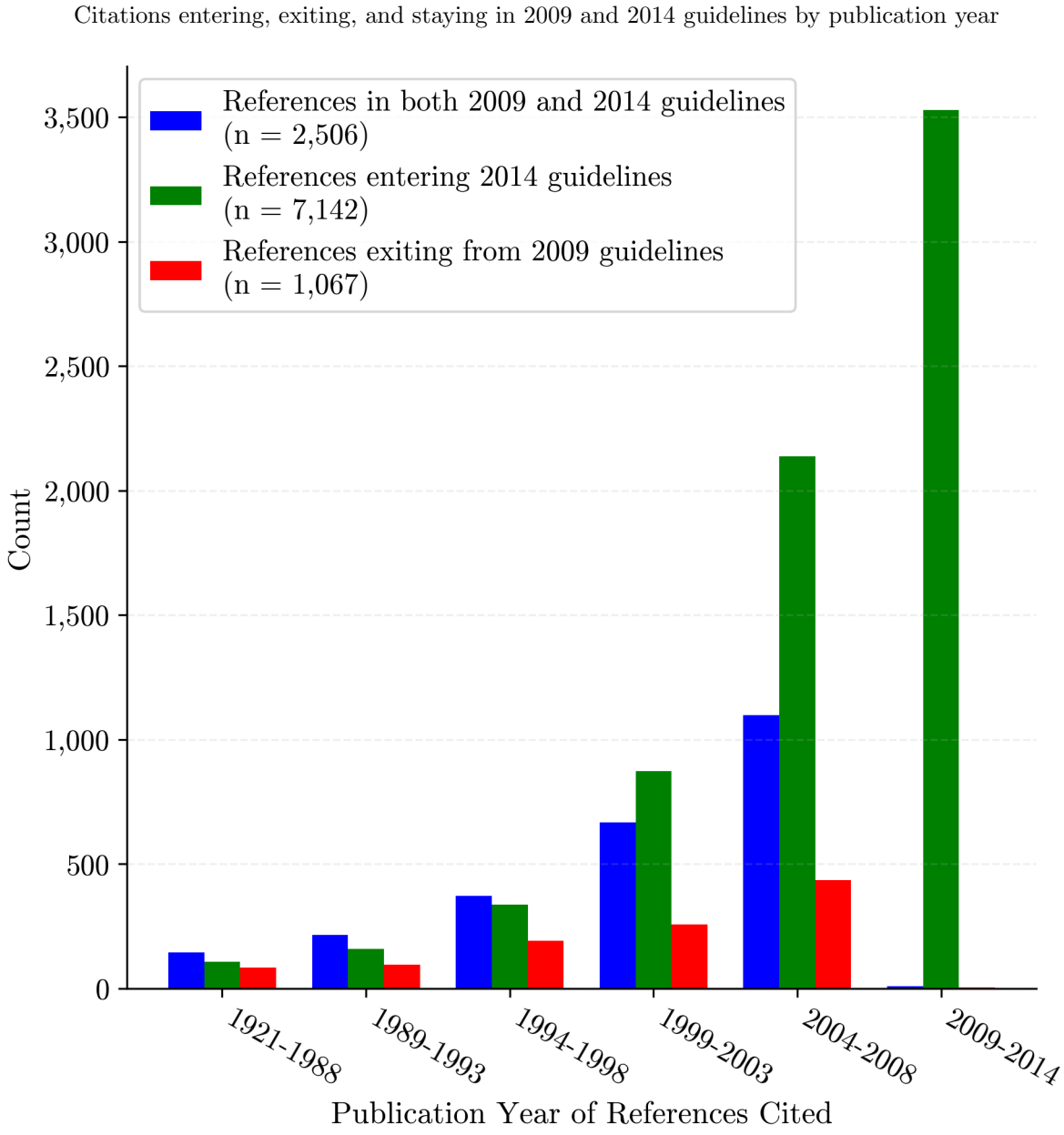
*Notes*: Table A.2 shows the relationships between page count and citation count of guidelines in cross-section and over time. Column 1 contains a cross-sectional regression of page count on citation count in 2009. Column 2 contains the same regression for 2014. Column 3 regresses the change in page count on the change in citation count between 2009 and 2014. All page counts and citation counts are for the first set of guidelines in each year from National Comprehensive Cancer Network (NCCN) Guidelines. Guidelines are grouped using the cancer grouping discussed in Section 3.3 and Appendix B.1 to maintain a stable sample of guidelines over time that match diagnosis code groupings. The regressions are estimated by ordinary least squares (OLS), with robust standard errors in parentheses.

Figure A.4: Guidelines primarily added scholarly works rather than replacing them



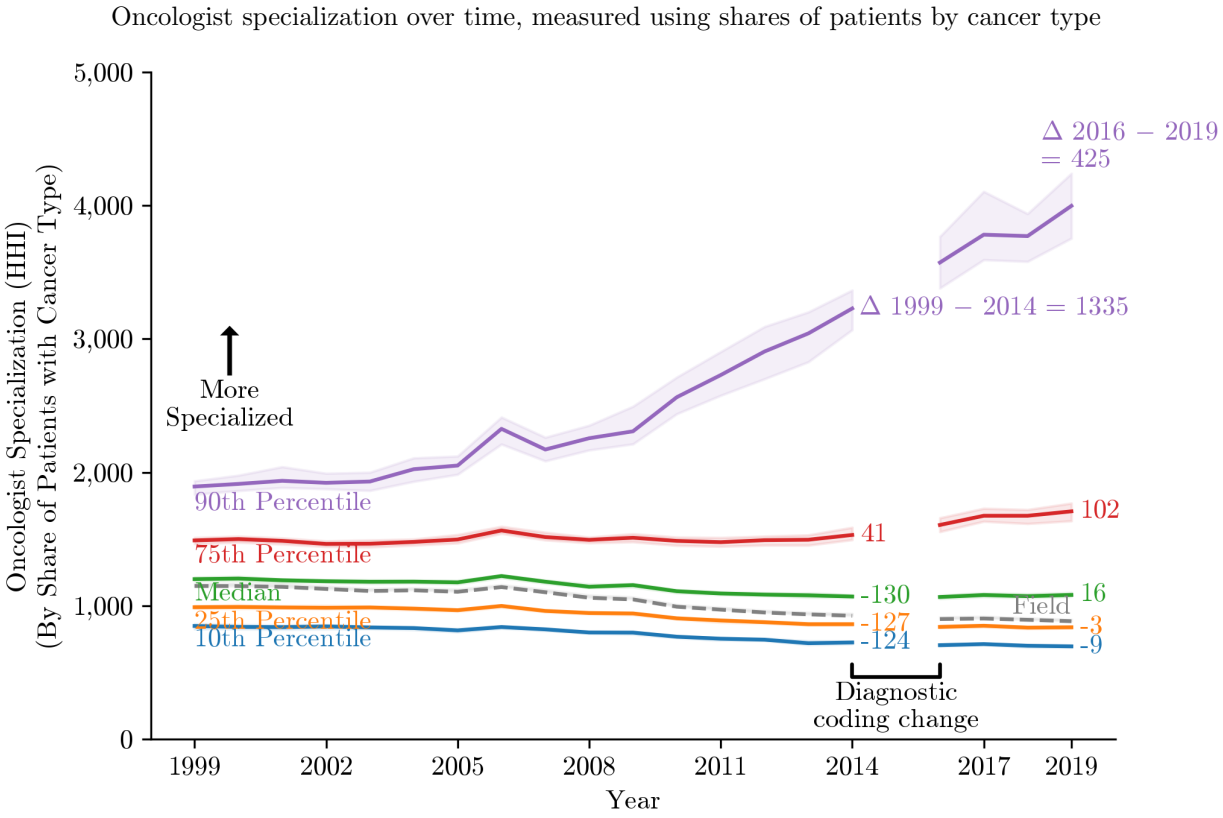

*Notes*: Figure A.4 shows changes in the references cited by National Comprehensive Cancer Network (NCCN) Clinical Practice Guidelines in Oncology for treatment, screening, and risk reduction of cancer between the 2009 and 2014 guidelines. Red marks exiting works, which are cited in the 2009 guidelines but not in the 2014 guidelines. Blue marks staying works, which are cited in both. Green marks entering works, which are cited in the 2014 guidelines but not in the 2009 guidelines. The sample includes all unique references in the first version of the 2009 and 2014 guidelines that link to scholarly works in the OpenAlex database (Priem et al., 2022).

Figure A.5: Growth in oncology clinical practice guideline citations between 2009 and 2014 comes primarily from adding references to new and recent scholarly works



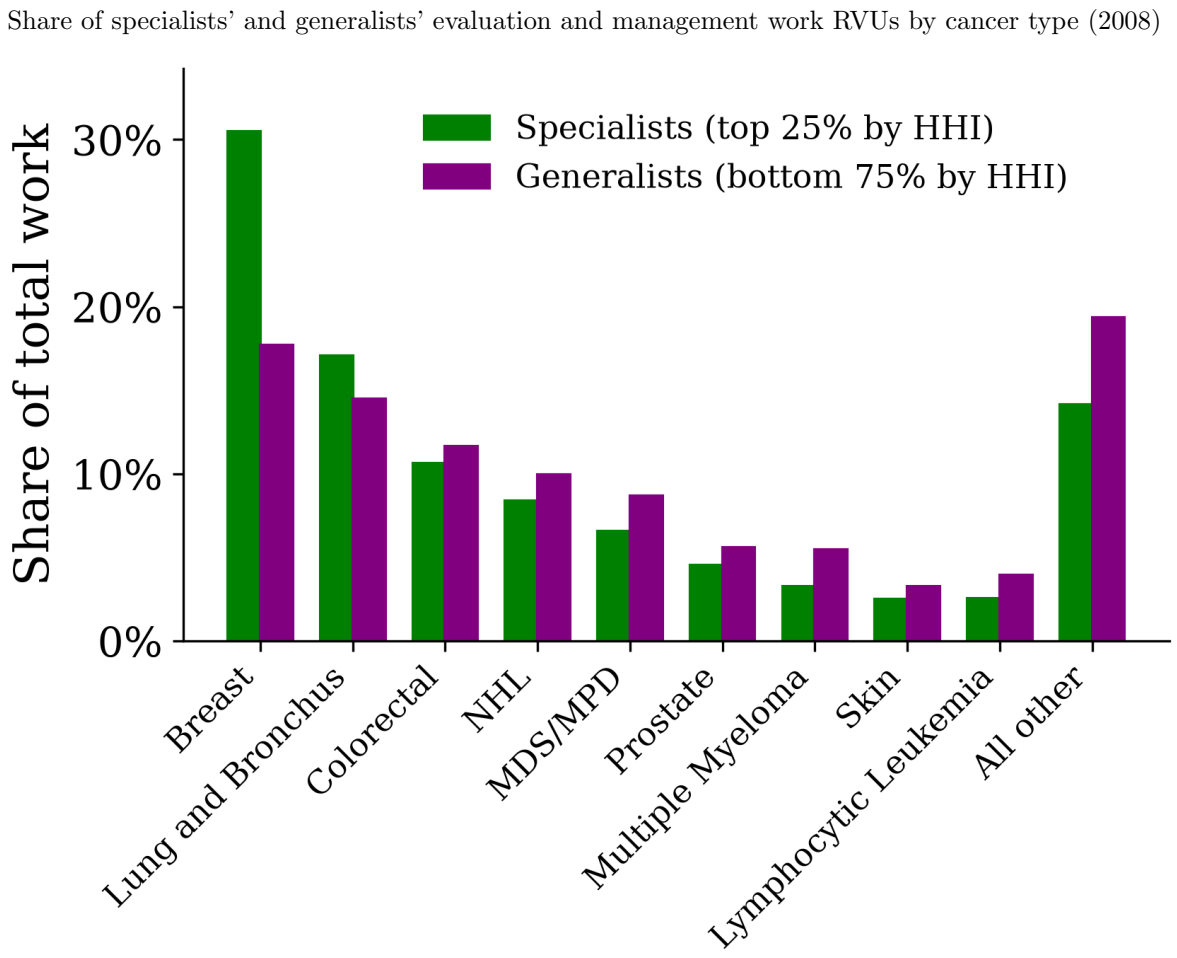

*Notes*: Figure A.5 shows changes in the publication year of references cited by National Comprehensive Cancer Network (NCCN) Clinical Practice Guidelines in Oncology between the 2009 and 2014 guidelines for screening, risk reduction, and treatment. Red marks exiting works, which are cited in the 2009 guidelines but not in the 2014 guidelines. Blue marks staying works, which are cited in both. Green marks entering works, which are cited in the 2014 guidelines but not in the 2009 guidelines. The sample includes all unique references in the first version of the 2009 and 2014 guidelines that link to scholarly works in the OpenAlex database (Priem et al., 2022) and for which a publication year is identified.

Figure A.6: Specialized oncologists are growing more specialized over time




*Notes*: Figure A.6 shows the distribution of oncologist specialization from 1999 to 2019, measured as in Figure 2 by a Herfindahl–Hirschman Index (HHI) over cancer types, but weighted by unique patients rather than by outpatient evaluation and management work Relative Value Units (wRVUs). HHI is jackknife-corrected as detailed in Appendix C.3. The 10th, 25th, 50th, 75th, and 90th percentiles are displayed and labeled. The numbers at the line ends give each percentile's change in HHI from 1999 to 2014 and from 2016 to 2019. 95% confidence intervals for each percentile are calculated by bootstrapping oncologists. The dashed grey line, labeled Field, shows the population-level HHI, the HHI of all cancer patients in the sample pooled together, with 95% confidence intervals from bootstrapping patients. All oncologists in the sample are hematologist-oncologists, medical oncologists, or hematologists who saw at least 25 unique cancer patients in the year in a 20% sample of Medicare Part B outpatient evaluation and management claims. The figure omits 2015, the year diagnosis coding switched from ICD-9 to ICD-10.

Figure A.7: Breast cancer accounts for a larger share of specialists' work than of generalists' work; other cancer types differ less




*Notes*: Figure A.7 shows specialists' and generalists' shares of outpatient evaluation and management work Relative Value Units (wRVUs) in each of the nine largest cancer types and in all other types pooled, summed over all specialists or all generalists. Cancer types are ordered left to right by total wRVUs over 2008–2019. NHL is non-Hodgkin lymphoma, and MDS/MPD are myelodysplastic and myeloproliferative disorders. Specialization is measured using a Herfindahl–Hirschman Index (HHI) over the types of cancer that the oncologist evaluated and managed in the year, weighted by wRVUs. HHI is jackknife-corrected as detailed in Appendix C.3. Specialists are oncologists at or above the 75th percentile of HHI in 2008, and generalists are those below it. All physicians in the sample are hematologist-oncologists, medical oncologists, or hematologists who saw at least 25 unique cancer patients in the year in a 20% sample of Medicare Part B outpatient evaluation and management claims, and who prescribed anti-cancer drugs to at least 11 patients in the year across a 20% sample of Part B and Part D data.

Table A.3: The 90th percentile of specialization rises with market size, more steeply in 2014, while the median is relatively flat

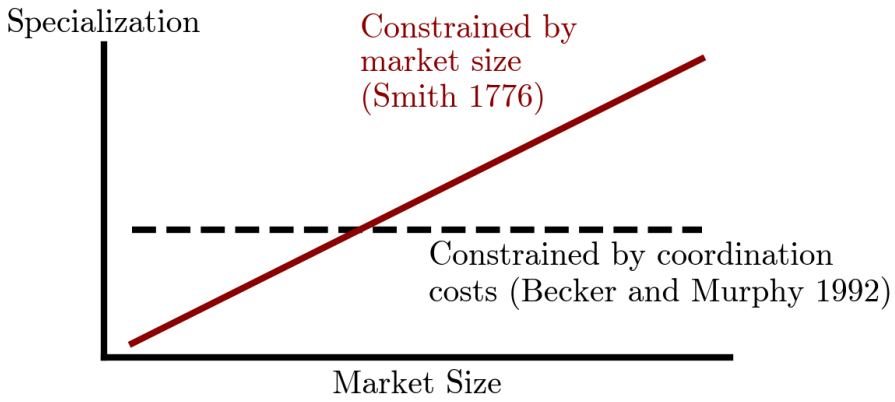
Regressions of the median and 90th percentile of specialization (HHI) in a market on log population (1999 and 2014)

| | (1) Median HHI | (2) 90th percentile HHI |
|---|---|---|
| Log population | 4.614 | 318.2 |
| | (14.89) | (63.01) |
| Log population × 2014 | 95.88 | 647.3 |
| | (36.00) | (124.6) |
| Observations | 599 | 599 |
| R-squared | 0.050 | 0.288 |

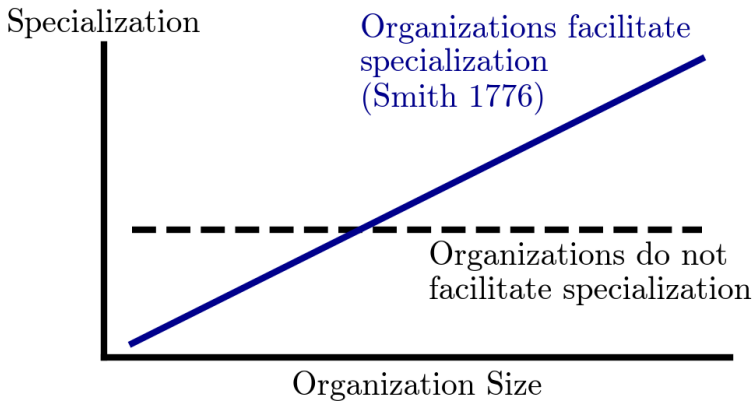
*Notes*: Table A.3 shows a regression of oncologist specialization on market size, year, and an interaction of market size and year. The regressions also include a 2014 indicator and a constant, which are not shown. In Column 1, the outcome is the median of oncologist specialization in a market. In Column 2, the outcome is the 90th percentile of oncologist specialization in a market. Markets are defined using hospital referral regions. Market size is measured as population in the relevant year. Multiplying a coefficient by $\ln 2$ gives the change in specialization per doubling of market population. Specialization is measured using a Herfindahl–Hirschman Index (HHI) over the types of cancer that the oncologist evaluated and managed in the year, weighted by outpatient evaluation and management work Relative Value Units (wRVUs). HHI is jackknife-corrected as detailed in Appendix C.3. The regressions are estimated by ordinary least squares (OLS), with robust standard errors in parentheses. All physicians in the sample are hematologist-oncologists, medical oncologists, or hematologists who saw at least 25 unique cancer patients in the year in a 20% sample of Medicare Part B outpatient evaluation and management claims.

Figure A.8: Predicted relationship between specialization and market size or organization size under competing theories

(a) Predicted specialization by market size



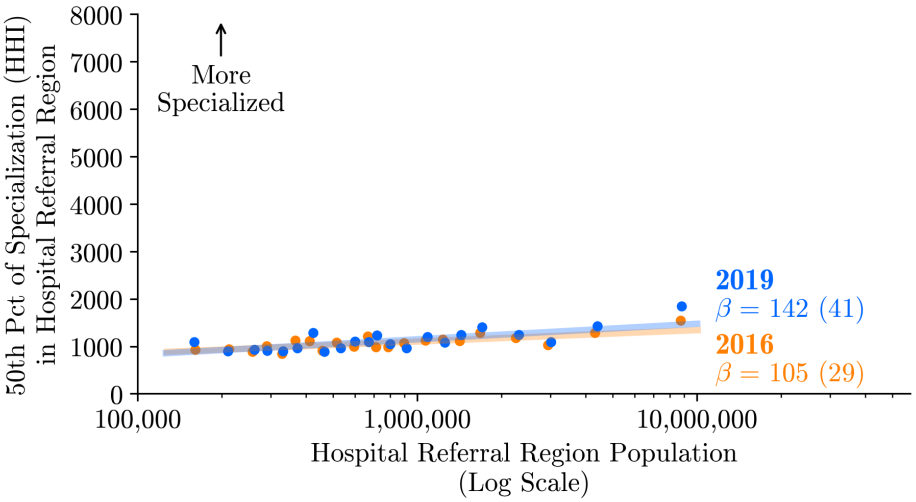

(b) Predicted specialization by organization size



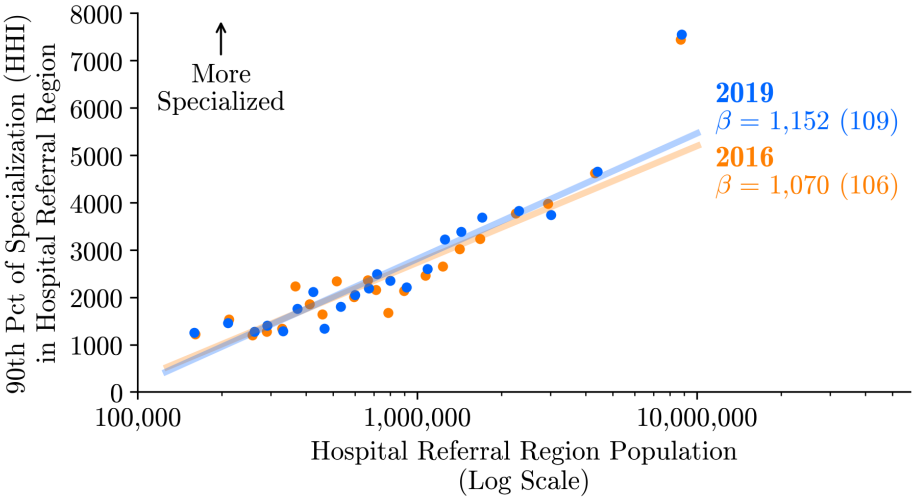

*Notes*: Panel (a) shows the predicted relationship between specialization and market size if specialization is constrained by market size (Smith, 1776) versus coordination costs (Becker and Murphy, 1992). The x-axis shows market size and the y-axis shows oncologist specialization. Panel (b) shows the predicted relationship between specialization and organization size if organizations do or do not facilitate specialization. The x-axis shows organization size and the y-axis shows oncologist specialization.

Figure A.9: Median specialization in a market is relatively flat over time and across market size; the 90th percentile rises with market size but does not change significantly between 2016 and 2019

(a) Median specialization by market size (2016 and 2019)

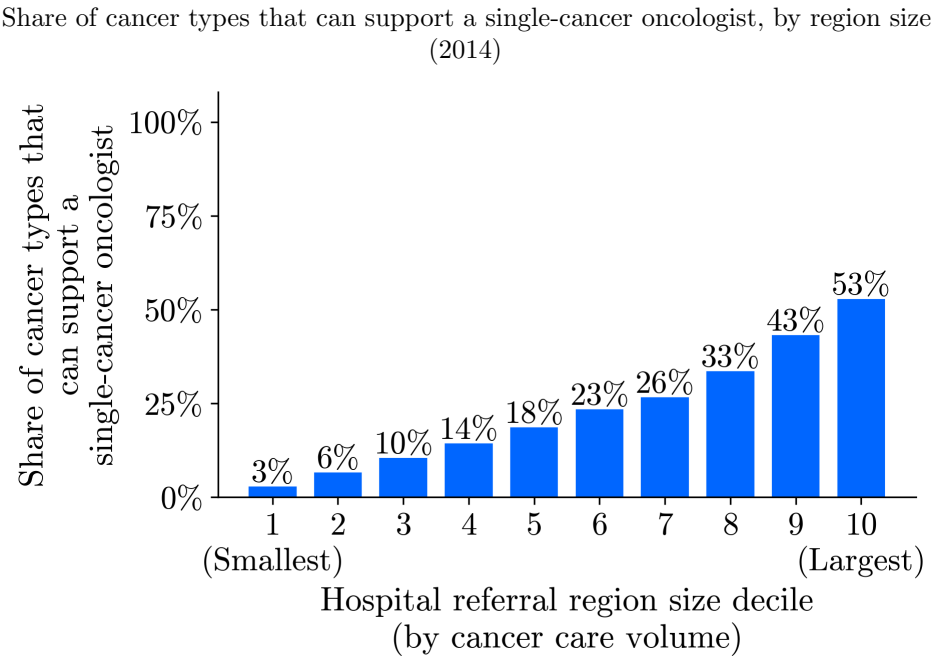



(b) 90th percentile of specialization by market size (2016 and 2019)



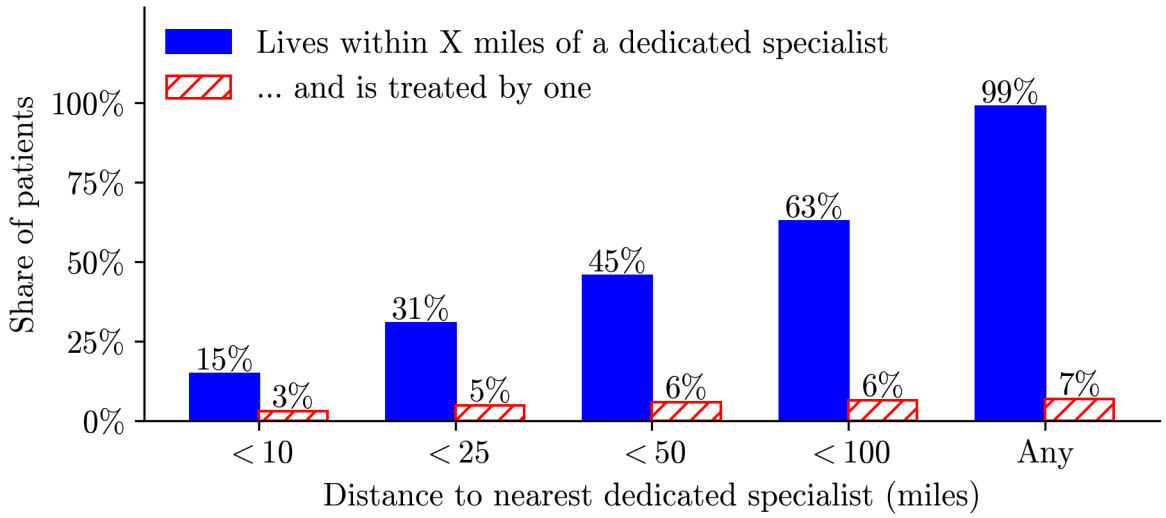

*Notes*: Figure A.9 shows binned scatterplots of the 50th and 90th percentile of oncologist specialization in a market by market size in 2016 and 2019, as Figure 4 does for 1999 and 2014. The x-axis shows the market size, measured as population, in the relevant year. Markets are defined using hospital referral regions. In Panel (a), the y-axis shows the 50th percentile of specialization in a market in a year. In Panel (b), the y-axis shows the 90th percentile of specialization in a market in a year. Specialization is measured using a Herfindahl–Hirschman Index (HHI) over the types of cancer that the oncologist evaluated and managed in the year, weighted by outpatient evaluation and management work Relative Value Units (wRVUs). HHI is jackknife-corrected as detailed in Appendix C.3. The lines plotted show the line of best fit between log population and the relevant percentile of specialization in each year. Each line's slope is reported as $\beta$, with the robust standard error in parentheses. All physicians in the sample are hematologist-oncologists, medical oncologists, or hematologists who saw at least 25 unique cancer patients in the year in a 20% sample of Medicare Part B outpatient evaluation and management claims.

Figure A.10: Outside the largest regions, most cancer types cannot support a single-cancer oncologist




*Notes*: Figure A.10 shows the average share of cancer types that can support a single-cancer oncologist in a region, by region size, in 2014. Regions are hospital referral regions. Region size is measured as cancer care volume: the total work Relative Value Units (wRVUs) on outpatient evaluation and management claims billed by oncologists for cancer patients who live in the region. The x-axis groups regions into deciles of cancer care volume. A cancer type is considered able to support a single-cancer oncologist in a region when the wRVUs for the region's patients with that cancer type are at least the median of oncologists' annual wRVUs for cancer patients. The median is taken across oncologists who saw at least 11 cancer patients in 2014. The y-axis shows the share of cancer types that can support a single-cancer oncologist, averaged over the regions in each decile, with each region weighted equally. All data are from a 20% sample of Medicare Part B outpatient evaluation and management claims in 2014.

Figure A.11: Few patients live near a dedicated specialist in their cancer type, and fewer are treated by one




*Notes*: Figure A.11 shows the share of cancer patients in 2014 who live within each distance of the nearest dedicated specialist in their cancer type, and the share who live within that distance and are treated by a dedicated specialist in their cancer type. Bars are cumulative: each distance includes the patients at all shorter distances, and "Any" includes every patient with a dedicated specialist in their cancer type anywhere in the country. A dedicated specialist is an oncologist who devotes more than half of their work to one cancer type. Work is measured in work Relative Value Units (wRVUs) on evaluation and management claims. Distances are straight-line miles between the centers of the patient's residence ZIP code and the specialist's practice ZIP code. A patient is treated by a specialist when that oncologist billed the most of the patient's evaluation and management work. Specialist status is measured among oncologists who saw at least 11 cancer patients. The sample is cancer patients who live in a ZIP code with income data and whose treating oncologist saw at least 11 cancer patients. All data are from a 20% sample of all Medicare Part B evaluation and management claims.

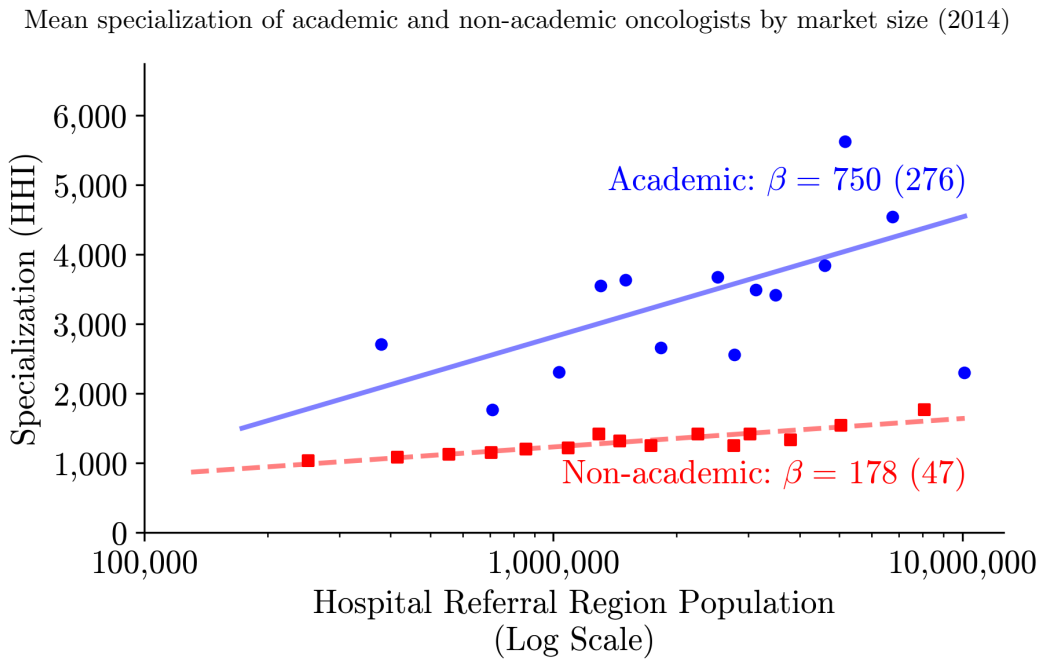
Table A.4: Specialization rises with organization size in academic organizations but not in non-academic ones

Regressions of the median and 90th percentile of specialization (HHI) in an organization on organization size, by organization type (2014)

| | (1) Median HHI Academic | (2) 90th percentile HHI Academic | (3) Median HHI Non-academic | (4) 90th percentile HHI Non-academic |
|---|---|---|---|---|
| Organization Size | 41.07 | 60.28 | 2.732 | 13.86 |
| | (8.864) | (10.77) | (3.209) | (8.598) |
| Observations | 155 | 115 | 1,361 | 335 |
| R-squared | 0.183 | 0.229 | 0.001 | 0.016 |
| Minimum Org Size | 1 | 5 | 1 | 5 |
| Standard Errors | Robust | Robust | Robust | Robust |

*Notes*: Table A.4 shows regressions of oncologist specialization on organization size by organization type in 2014. In Column 1, the outcome is the median of oncologist specialization in academic organizations. In Column 2, the outcome is the 90th percentile of oncologist specialization in academic organizations with at least 5 oncologists. In Column 3, the outcome is the median of oncologist specialization in non-academic organizations. In Column 4, the outcome is the 90th percentile of oncologist specialization in non-academic organizations with at least 5 oncologists. Organization size is estimated as the number of oncologists at the organization's tax ID in the 2014 Medicare Data on Provider Practice and Specialty (MD-PPAS), using oncologists' primary tax ID. Organization size enters in levels, so coefficients are per additional oncologist. Organization type is determined using tax ID names from the MD-PPAS. See Section 3.2 and Appendix B.3 for discussion of the limitations of variables from the MD-PPAS. Specialization is measured using a Herfindahl–Hirschman Index (HHI) over the types of cancer that the oncologist evaluated and managed in the year, weighted by outpatient evaluation and management work Relative Value Units (wRVUs). HHI is jackknife-corrected as detailed in Appendix C.3. The regressions are estimated by ordinary least squares (OLS), with robust standard errors in parentheses. All physicians in the sample are hematologist-oncologists, medical oncologists, or hematologists who saw at least 25 unique cancer patients in the year in a 20% sample of Medicare Part B outpatient evaluation and management claims.

Figure A.12: Academic oncologists are more specialized than non-academic oncologists across the distribution of market sizes




*Notes*: Figure A.12 shows a physician-level binned scatterplot of specialization by market size in 2014, separately for academic and non-academic physicians. The x-axis shows market size, measured as population in 2014, on a log scale. Markets are defined using hospital referral regions (HRRs). The y-axis shows the mean specialization of the physicians in each bin. Physician academic status is determined using tax ID names from the 2014 Medicare Data on Provider Practice and Specialty (MD-PPAS). See Section 3.2 and Appendix B.3 for discussion of the limitations of variables from the MD-PPAS. Specialization is measured using a Herfindahl–Hirschman Index (HHI) over the types of cancer that the oncologist evaluated and managed in the year, weighted by outpatient evaluation and management work Relative Value Units (wRVUs). HHI is jackknife-corrected as detailed in Appendix C.3. The lines plotted show the line of best fit between log population and specialization in each series. The estimating equation for each series is $HHI_i = \beta_0 + \beta_1 \log\left(pop_{m(i)}\right) + \epsilon_i$, where $i$ indexes physicians and $m(i)$ is physician $i$'s market. The displayed slopes are per unit of log population, with standard errors clustered by HRR in parentheses. Multiplying a slope by $\ln 2$ gives the change in specialization per doubling of market population. The sample is not restricted to the balanced panel. All physicians in the sample are hematologist-oncologists, medical oncologists, or hematologists who saw at least 25 unique cancer patients in the year in a 20% sample of Medicare Part B outpatient evaluation and management claims.

Table A.5: Academic oncologists are more specialized than non-academic oncologists at the same market size

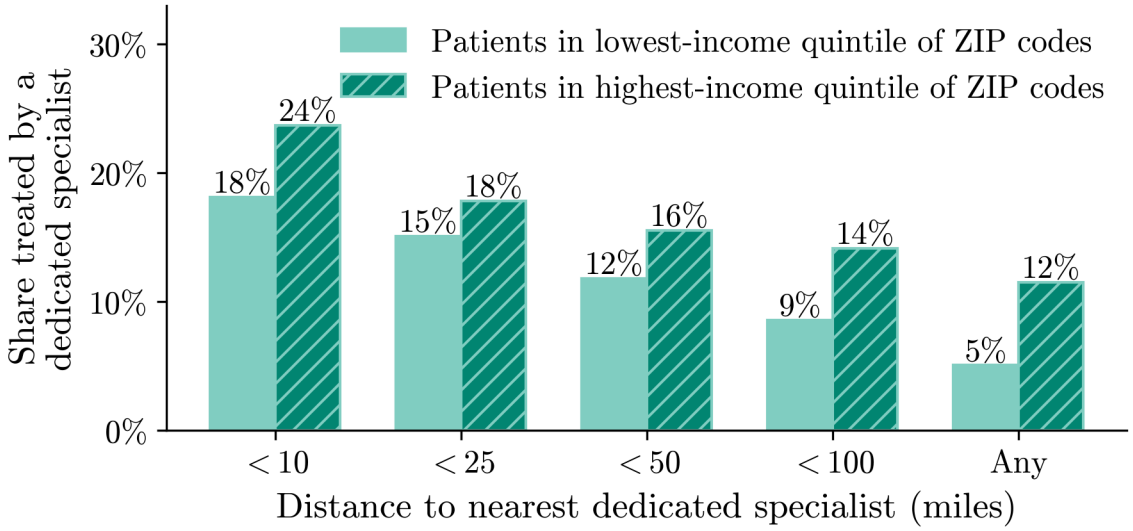
Regressions of specialization (HHI) on academic status and market size (2008 and 2014)

| | Dependent variable: HHI | |
|---|---|---|
| | (1) | (2) |
| | 2008 | 2014 |
| Academic | 1,624 | 1,845 |
| | (245.9) | (189.5) |
| Log market population | | |
| (year-matched, centered at mean) | 128.3 | 177.7 |
| | (36.73) | (46.90) |
| Academic × log market population | 584.5 | 571.9 |
| | (384.3) | (297.4) |
| Observations | 4,698 | 4,713 |
| R-squared | 0.216 | 0.245 |

*Notes*: Table A.5 shows regressions of oncologist specialization on academic status (indicator), log market population, and their interaction, separately for 2008 (Column 1) and 2014 (Column 2). The estimating equation is $HHI_i = \beta_0 + \beta_1 Academic_i + \beta_2 \log\big(pop_{m(i)}\big) + \beta_3 (Academic_i \times \log\big(pop_{m(i)}\big)) + \epsilon_i$, where $i$ indexes physicians and $m(i)$ is physician $i$'s market. Markets are defined using hospital referral regions (HRRs). Market size is measured as population in the relevant year. Log market population is centered at its mean in the estimation sample. Multiplying the log population or interaction coefficient by $\ln 2$ gives the change in specialization per doubling of market population. Physician academic status in each column is determined using tax ID names from that year's Medicare Data on Provider Practice and Specialty (MD-PPAS). See Section 3.2 and Appendix B.3 for discussion of the limitations of variables from the MD-PPAS. Specialization is measured using a Herfindahl–Hirschman Index (HHI) over the types of cancer that the oncologist evaluated and managed in the year, weighted by outpatient evaluation and management work Relative Value Units (wRVUs). HHI is jackknife-corrected as detailed in Appendix C.3. The regressions are estimated by ordinary least squares (OLS), with standard errors clustered by HRR in parentheses. The sample in each column is not restricted to the balanced panel. All physicians in the sample are hematologist-oncologists, medical oncologists, or hematologists who saw at least 25 unique cancer patients in the year in a 20% sample of Medicare Part B outpatient evaluation and management claims.

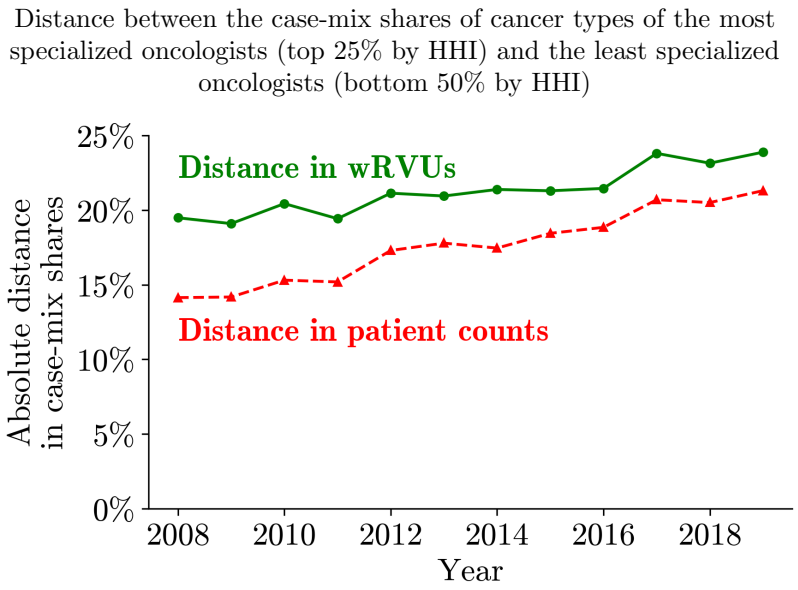
Figure A.13: Among patients who live near a dedicated specialist, higher-income patients are more often treated by a dedicated specialist

Share of patients treated by a dedicated specialist, by distance to the nearest dedicated specialist and ZIP-code income, 2014



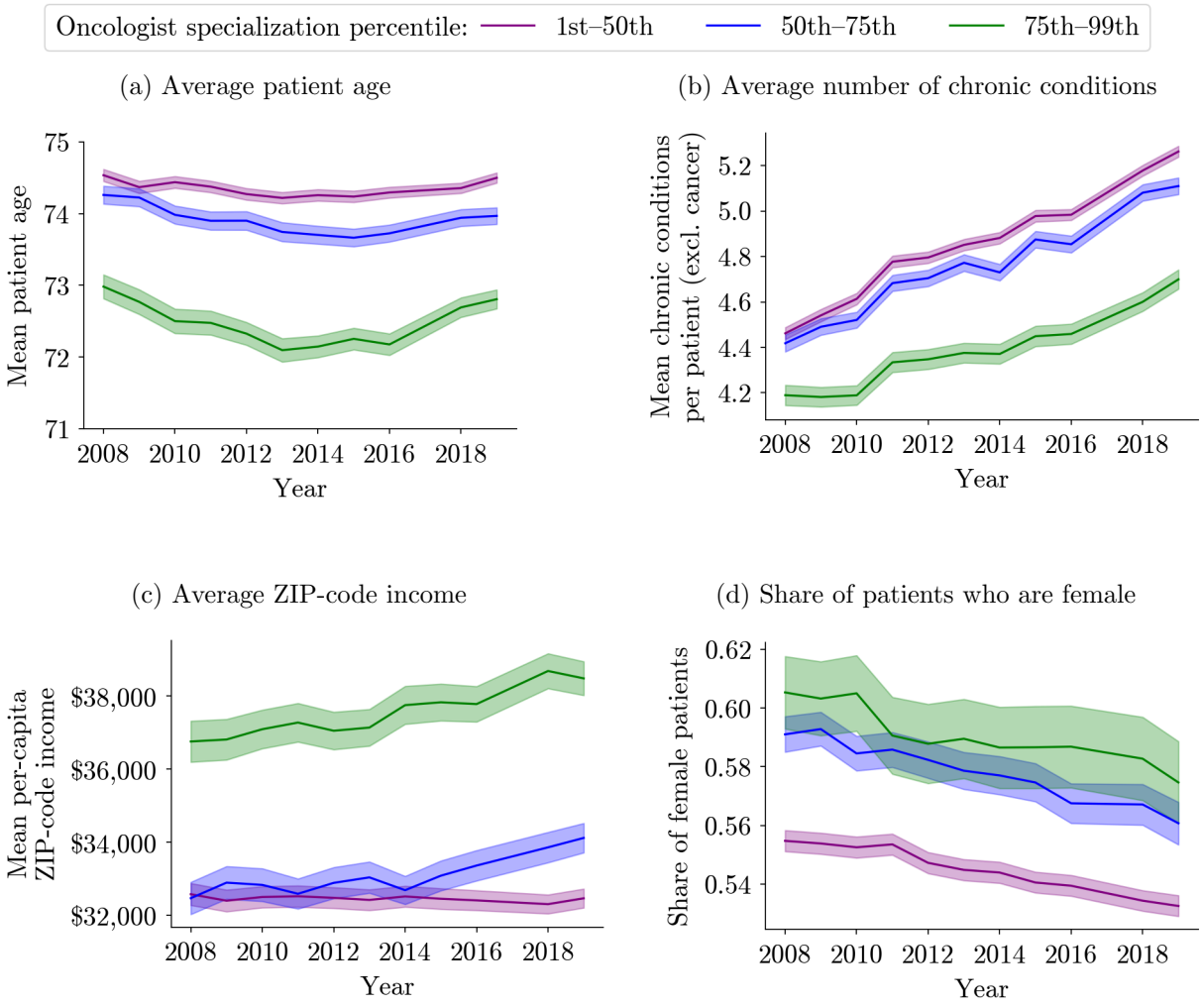

*Notes*: Figure A.13 shows the share of cancer patients treated by a dedicated specialist in their cancer type, by distance to the nearest dedicated specialist, for patients in the lowest and highest quintiles of ZIP-code per-capita income in 2014. Distance categories are cumulative, and "Any" includes all patients with a dedicated specialist in their cancer type anywhere in the country. A dedicated specialist is an oncologist who devotes more than half of their evaluation and management work Relative Value Units (wRVUs) to one cancer type. A patient is considered treated by the oncologist who billed the largest share of the patient's evaluation and management wRVUs. Distance is measured in straight-line miles between ZIP-code centers. Specialist status is measured among oncologists who saw at least 11 cancer patients. The sample is cancer patients who live in a ZIP code with income data and whose treating oncologist saw at least 11 cancer patients. Data are from a 20% sample of all Medicare Part B evaluation and management claims.

Figure A.14: The most and least specialized oncologists treat increasingly different cancers



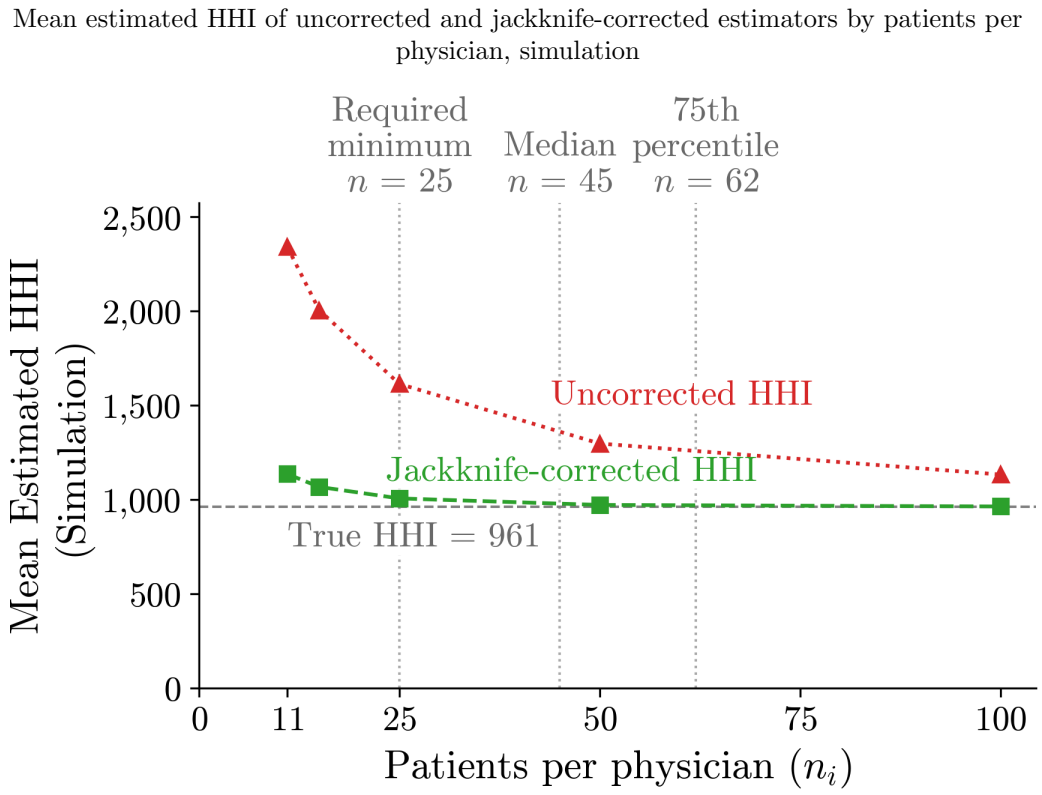

*Notes*: Figure A.14 shows the distance between the average case-mix shares of the most specialized and the least specialized oncologists over time. The distance in each year is half the L1 distance, $\frac{1}{2}\sum_c |s_c^{\text{spe}} - s_c^{\text{gen}}|$, where $s_c^{\text{spe}}$ and $s_c^{\text{gen}}$ are the average case-mix shares of type of cancer $c$ of the most and the least specialized oncologists. An oncologist's case-mix share of a type of cancer is calculated as the share of the oncologist's outpatient evaluation and management work Relative Value Units, or of the oncologist's unique cancer patients, in that type of cancer in a year. Oncologists are weighted equally in each average. Specialization is measured using a Herfindahl–Hirschman Index (HHI) over the types of cancer that the oncologist evaluated and managed in the year, weighted by outpatient evaluation and management work Relative Value Units (wRVUs). HHI is jackknife-corrected as detailed in Appendix C.3. Oncologists are grouped into specialization bins by percentile of HHI by year. The most specialized oncologists are defined as those at or above the 75th percentile, and the least specialized oncologists as those below the 50th percentile. All physicians in the sample are hematologist-oncologists, medical oncologists, or hematologists who saw at least 25 unique cancer patients in the year in a 20% sample of Medicare Part B outpatient evaluation and management claims, and who prescribed anti-cancer drugs to at least 11 patients in the year across a 20% sample of Part B and Part D data.

Figure A.15: Patients of the most specialized oncologists are younger, more female, have fewer chronic conditions, and live in higher-income ZIP codes




*Notes*: This figure shows the characteristics of the cancer patients attributed to oncologists over time by level of specialization. Each cancer patient is attributed to the oncologist with the most of the patient's evaluation and management work Relative Value Units in the year. Chronic conditions are counted from the Medicare Beneficiary Summary File's chronic-condition flags, excluding cancer. ZIP-code income is the per-capita income of the patient's residence ZIP code. Oncologists are weighted equally to estimate each average by degree of specialization. Specialization is measured using a Herfindahl–Hirschman Index (HHI) over the types of cancer that the oncologist evaluated and managed in the year, weighted by outpatient evaluation and management work Relative Value Units (wRVUs). HHI is jackknife-corrected as detailed in Appendix C.3. Oncologists are grouped into specialization bins by percentile of HHI by year, where a higher HHI indicates a more specialized oncologist. The sample is restricted to oncologists with at least 11 attributed patients in the year in a 20% sample of Medicare Part B outpatient evaluation and management claims.

Figure A.16: The jackknife correction removes most of the finite-sample bias in measured specialization




*Notes*: Figure A.16 shows a Monte Carlo simulation of two estimators of physician specialization: the uncorrected and the jackknife-corrected Herfindahl–Hirschman Index (HHI), on the 0–10,000 scale. The uncorrected HHI is computed over each cancer type's share of the physician's total time, measured in evaluation and management work Relative Value Units (wRVUs). The jackknife-corrected HHI is $\widehat{\text{HHI}}^{\text{jack}} = n_i \widehat{\text{HHI}} - \frac{n_i-1}{n_i} \sum_{j=1}^{n_i} \widehat{\text{HHI}}_{(j)}$, where $n_i$ is the physician's number of patients, $\widehat{\text{HHI}}$ is the uncorrected HHI, and $\widehat{\text{HHI}}_{(j)}$ is the HHI recomputed leaving out patient $j$. The simulation draws patients independently, drawing each patient's cancer type from the aggregate cancer-type distribution and each patient's wRVUs from a lognormal distribution. The simulation is calibrated to the 2009 claims data using the aggregate cancer-type distribution, the mean wRVUs per patient by cancer type, and the median, across oncologists with at least 11 patients, of the coefficient of variation of each oncologist's wRVUs per patient. The simulation parameters imply a true HHI of 961, marked by the horizontal dashed line and used as the benchmark against which to measure bias. The figure plots the mean uncorrected HHI and the mean jackknife-corrected HHI across simulated physicians, against the number of patients per simulated physician ($n_i$). Each plotted point averages 50,000 simulated physicians. The vertical dotted lines mark the sample's required minimum of 25 patients per physician, and the median and 75th percentile of patients per physician in 2009 among physicians with at least 25 patients that year.

Table A.6: The jackknife correction removes most of the finite-sample bias in measured specialization for a general oncologist and for a specialist

Mean bias and SD of the HHI estimators by observed patients, simulation

| Observed patients ($n$) | Mean bias in uncorrected HHI (standard error) | SD of uncorrected HHI | Mean bias in jackknife-corrected HHI (standard error) | SD of jackknife-corrected HHI |
|---|---|---|---|---|
| *Panel A. General oncologist (true HHI 961)* | | | | |
| 11 | 1,381.3 | 718 | 175.4 | 914 |
| | (3.2) | | (4.1) | |
| 25 | 653.2 | 412 | 45.8 | 462 |
| | (1.8) | | (2.1) | |
| 50 | 336.9 | 264 | 11.2 | 281 |
| | (1.2) | | (1.3) | |
| 100 | 172.6 | 176 | 3.1 | 183 |
| | (0.8) | | (0.8) | |
| 200 | 86.7 | 120 | 0.1 | 123 |
| | (0.5) | | (0.6) | |
| *Panel B. 80% breast cancer specialist (true HHI 5,629)* | | | | |
| 11 | 869.0 | 2,058 | 164.7 | 2,714 |
| | (9.2) | | (12.1) | |
| 25 | 436.4 | 1,591 | 56.2 | 1,871 |
| | (7.1) | | (8.4) | |
| 50 | 221.1 | 1,210 | 7.6 | 1,333 |
| | (5.4) | | (6.0) | |
| 100 | 117.7 | 892 | 4.5 | 942 |
| | (4.0) | | (4.2) | |
| 200 | 59.5 | 649 | 1.0 | 668 |
| | (2.9) | | (3.0) | |

*Notes*: Table A.6 reports the mean bias and the standard deviation of two estimators of physician specialization in a Monte Carlo simulation: the uncorrected and the jackknife-corrected Herfindahl–Hirschman Index (HHI), on the 0–10,000 scale. The uncorrected HHI is computed over each cancer type's share of the physician's total time, measured in evaluation and management work Relative Value Units (wRVUs). The jackknife-corrected HHI is $\widehat{\text{HHI}}^{\text{jack}} = n\,\widehat{\text{HHI}} - \frac{n-1}{n}\sum_{j=1}^{n}\widehat{\text{HHI}}_{(j)}$, where $n$ is the physician's number of patients, $\widehat{\text{HHI}}$ is the uncorrected HHI, and $\widehat{\text{HHI}}_{(j)}$ is the HHI recomputed leaving out patient $j$. The simulation draws patients independently, drawing each patient's cancer type from the physician's cancer-type distribution and each patient's wRVUs from a lognormal distribution. The simulation is calibrated to the 2009 claims data using the aggregate cancer-type distribution, the mean wRVUs per patient by cancer type, and the median, across oncologists with at least 11 patients, of the coefficient of variation of each oncologist's wRVUs per patient. In Panel A, each simulated physician draws patients from the aggregate cancer-type distribution, for a true HHI of 961. In Panel B, each simulated physician draws 80% of patients from breast cancer and 20% from the other types in proportion to their aggregate shares, for a true HHI of 5,629. Each row uses 50,000 simulated physicians. Mean bias and SD are the mean and the standard deviation, across simulated physicians, of the estimated HHI minus the true HHI. Values in parentheses are Monte Carlo standard errors of the mean bias.

Table A.7: Other measures of knowledge growth lead to the same conclusions as guideline pages

Regressions of the change in HHI (2009–2014) on recentered exposure to knowledge growth, using alternative measures or recentering controls

| | Change in HHI | | | |
|---|---|---|---|---|
| | (1) Guideline citations (100 citations) | (2) Net scholarly works (100 works) | (3) Guideline pages, first guideline as baseline (100 pages) | (4) Guideline pages, recentered with 2009 length (100 pages) |
| *Panel A. Generalists (bottom 75% of initial specialization)* | | | | |
| Knowledge growth | 9.659 | 11.32 | 27.39 | 30.24 |
| Robust SE | (4.226) | (5.079) | (15.92) | (21.40) |
| RI 95% CI | [−0.298, 18.8] | [−0.169, 21.9] | [−13.5, 65.1] | [−18.5, 73.4] |
| Observations | 2,531 | 2,531 | 2,531 | 2,531 |
| R-squared | 0.002 | 0.002 | 0.001 | 0.001 |
| *Panel B. Specialists (top 25% of initial specialization)* | | | | |
| Knowledge growth | 12.27 | 16.30 | 14.50 | 70.16 |
| Robust SE | (14.60) | (18.03) | (56.71) | (81.21) |
| RI 95% CI | [−45.2, 69.7] | [−48.6, 88.1] | [−210, 252] | [−167, 330] |
| Observations | 779 | 779 | 779 | 779 |
| R-squared | 0.001 | 0.001 | 0.000 | 0.001 |

*Notes*: Table A.7 shows estimates of the estimating equation in Section 5.1, a regression of the change in specialization from 2009 to 2014 on exposure to knowledge growth, using alternative measures or recentering controls. Specialization is measured using a Herfindahl–Hirschman Index (HHI) over the types of cancer that the oncologist evaluated and managed in the year, weighted by outpatient evaluation and management work Relative Value Units (wRVUs). HHI is jackknife-corrected as detailed in Appendix C.3. Generalists are the oncologists below the 75th percentile of the 2008 HHI distribution, and specialists are those at or above it. Robust standard errors are in parentheses. Brackets show 95% confidence intervals from randomization inference (RI), constructed as in Appendix D.5. Exposure to knowledge growth is recentered as described in Section 5, using the predicted knowledge growth regression, Equation (2), applied to each column's measure of knowledge growth. Column 1 measures knowledge growth by the increase in guideline citations of any type. Column 2 measures it by the net increase in scholarly works cited. Column 3 measures it by guideline pages, using the page count of a cancer type's first guideline as its baseline when that guideline appears during the study period. Column 4 measures knowledge growth by guideline pages, as in Table 5, but adds each guideline's 2009 length to the predicted knowledge growth regression. All columns are univariate regressions with no fixed effects, estimated on physicians with observed, non-singleton fixed effects for the following 2008 variables: age group, sex, market (hospital referral region), decile of patient count, organization size (the number of oncologists at their primary organization), and academic affiliation. All physicians in the sample are hematologist-oncologists, medical oncologists, or hematologists who saw at least 25 unique cancer patients in a 20% sample of Medicare Part B outpatient evaluation and management claims in 2008, 2009, and 2014.

Table A.8: In robustness checks, generalists still do not meaningfully increase their specialization, and specialist estimates remain imprecise

Regressions of the change in HHI (2009–2014) under alternative patient count thresholds, visit types or recentering

| | Change in HHI | | | | |
|---|---|---|---|---|---|
| | (1) Baseline (100 pages) | (2) Two patients per cancer type (100 pages) | (3) Three patients per cancer type (100 pages) | (4) New patient visits only (100 pages) | (5) Recentered by incidence (100 pages) |
| *Panel A. Generalists (bottom 75% of initial specialization)* | | | | | |
| Knowledge growth | 28.52 | −3.246 | 1.166 | 31.07 | 30.93 |
| Robust SE | (17.69) | (16.78) | (16.71) | (38.24) | (18.13) |
| RI 95% CI | [−14.4, 67.1] | [−68, 58.3] | [−98.2, 96.2] | [−16.8, 84.3] | [−13.4, 70.1] |
| Mean cancer types per physician | 15.3 | 9.7 | 6.9 | 10.2 | 15.3 |
| Observations | 2,567 | 2,567 | 2,567 | 981 | 2,567 |
| R-squared | 0.001 | 0.000 | 0.000 | 0.001 | 0.001 |
| *Panel B. Specialists (top 25% of initial specialization)* | | | | | |
| Knowledge growth | 46.70 | 132.0 | 100.7 | −102.7 | 32.51 |
| Robust SE | (60.22) | (64.19) | (79.35) | (110.4) | (62.03) |
| RI 95% CI | [−167, 267] | [−122, 381] | [−151, 321] | [−504, 334] | [−177, 259] |
| Mean cancer types per physician | 11.8 | 7.0 | 4.8 | 6.6 | 11.8 |
| Observations | 858 | 858 | 858 | 328 | 858 |
| R-squared | 0.001 | 0.004 | 0.002 | 0.002 | 0.000 |

*Notes*: Table A.8 shows estimates of the estimating equation in Section 5.1, a regression of the change in specialization from 2009 to 2014 on exposure to knowledge growth, using alternative constructions of specialization, exposure or recentering. See Table 5 for the definitions of HHI, generalists, specialists, recentered exposure, robust standard errors (in parentheses) and randomization-inference (RI) confidence intervals (in brackets). All columns are univariate regressions with no fixed effects. Columns 2 and 3 raise to two and three the number of patients of a cancer type a physician must see before that cancer type enters their exposure measure. Column 4 builds both specialization and exposure from new patient evaluation and management claims rather than all outpatient evaluation and management claims, and correspondingly lowers to 11 the number of patients a physician must see to enter the sample. Column 5 replaces $p_c$, the log of the number of Medicare patients with cancer type $c$ in 2008, with the log of the share of new Medicare cancer patients in 2008 who have cancer type $c$ in the predicted knowledge growth regression, Equation (2). Unless otherwise specified, all physicians in the sample are hematologist-oncologists, medical oncologists, or hematologists who saw at least 25 unique cancer patients in a 20% sample of Medicare Part B outpatient evaluation and management claims in 2008, 2009, and 2014.

# B Data Appendix

This appendix provides additional detail and discussion about data construction. Section B.1 provides details and examples about the construction of the custom groupings for cancer diagnosis codes. Section B.2 provides details about how each patient is assigned a primary cancer. Section B.3 provides details and examples about the construction of the indicator for academic status for physicians. Section B.4 describes how I extract citations from the guidelines and link them to scholarly works. Section B.5 describes how I wrote and checked the analysis code.

## B.1 Cancer Groupings

I create a set of custom diagnosis code groupings that align with how cancers are grouped in NCCN Guidelines. This section discusses why and how I create these custom groupings. The code and mappings to create these groupings can be freely shared upon request.

I create this grouping because ICD-9 diagnosis codes are too granular to be used directly as types of cancer treated. For example, they contain separate codes for breast cancer in the right breast versus the left breast. I wish to group these into the category of "breast cancer" to align with the NCCN Guidelines for breast cancer treatment.

To create the grouping, I start with a list of cancer diagnosis codes from the Clinical Classification Software (CCS). The CCS also has groupings to the level of cancer type, but these groupings have several challenges that I rectify with the custom grouping. First, I drop the CCS diagnosis codes for cancer treatment and consider only those for specific types of cancer. Second, I make the groupings more specific. For example, the CCS groupings classify colon cancer of "unknown behavior" as an "unknown" cancer type. I re-classify this as colon cancer. Third, I make the custom grouping more granular to align with the NCCN Guidelines. For example, I divide the CCS category of "Leukemia" into "Myeloid Leukemia" and "Lymphocytic Leukemia" (NCCN Guidelines are even further split by acute and chronic

within each subtype of leukemia). Lastly, I rearrange groupings that are misaligned with clinical groupings. For example, the CCS groupings group rectal cancer with anal cancer. However, NCCN Guidelines group rectal cancer with colon cancer. The custom grouping follows the NCCN Guidelines' approach, with a single category for colon and rectal cancer.

Notably, cancer groupings in NCCN Guidelines have become more granular over time. I group NCCN Guidelines together to form cancer type groups, defaulting to the broadest (and typically oldest) grouping seen in NCCN Guidelines. For example, the NCCN Guidelines contain a single set of guidelines for "Non-Hodgkin's Lymphoma" through 2016. After 2016, these guidelines are divided into separate guidelines for B-cell lymphomas, T-cell lymphomas, lymphoplasmacytic lymphomas, and others. I group all of these into a single category of non-Hodgkin lymphoma in order to have a stable grouping over time. Similarly, the NCCN Guidelines for basal and squamous cell cancer (both forms of skin cancer) were presented in a single document until 2014, after which they were issued separately. I treat these jointly under the broader category of "skin cancer guidelines."

I also group screening and diagnosis NCCN Guidelines with treatment NCCN Guidelines where relevant when counting guideline lengths and citation counts. For most cancers, all types of recommendations are in a single guideline. However, breast and prostate cancer have NCCN Guidelines for screening and diagnosis that are separate from those for treatment. I group all types of NCCN Guidelines for breast cancer and all types of NCCN Guidelines for prostate cancer together. This grouping ensures breast and prostate cancer measures of length are analogous to the measures for other types of cancers.

These mappings are based on ICD-9 codes. In mid-2015, the claims data switch from ICD-9 to ICD-10 codes. From 2015 onwards, I map from ICD-10 to ICD-9 codes, then to the custom cancer grouping.

Notably, the ICD-based groupings are based on the anatomical origin of the tumor, which mirrors the first-order dimension of cancer subtyping. For example, breast cancers are cancers where the origin tumor site is the breast, regardless of where the cancer has

spread. Within each anatomical origin, there are further subclassifications along numerous other dimensions, e.g., stage and biomarkers. I cannot observe these other dimensions in the diagnosis codes. However, I do not need to observe them, because NCCN Guidelines are also organized by the anatomical origin of tumors. Although recent years have seen the emergence of biomarker-driven, tissue-agnostic cancer therapies, the first such FDA approval occurred in 2017 (Olivier, Haslam, and Prasad, 2021), which falls outside the primary panel period analyzed here.

The anatomical organization of NCCN Guidelines suggests that the grouping of cancer types aligns with how the field of oncology initially categorizes cancer (there are many dimensions of additional subtyping by size, spread, biomarkers, and other features). This anatomical approach also aligns with how sub-specialized oncologists at the University of Chicago describe their focus area, i.e., using anatomic designations such as "breast cancer oncologist", "GI oncologist", or "thoracic oncologist".

## B.2 Assigning Patients to a Primary Cancer

Cancer patients are defined as those with at least $250 (in 1999 inflation-adjusted dollars) of Medicare Part B spending associated with cancer diagnoses. A patient's primary cancer is the most specific cancer type for which they have the highest number of diagnosis codes in a given year. For example, a patient with 10 diagnosis codes for lung cancer and 30 for breast cancer is classified as having breast cancer. If a patient has codes for both "Unspecified" cancer and breast cancer, they are classified as having breast cancer, regardless of the number of Unspecified codes. Patients with only unspecified or non-specific cancer diagnoses are excluded from the sample (1.0% of patients in 2009).

Most patients (67%) are diagnosed with a single cancer type, while only 33% have multiple types in 2009. Among patients with multiple diagnoses, the primary cancer accounts for the majority of diagnosis codes (76%). In 2008, excluding patients coded only by their most-visited oncologist, other providers contributed 57% of primary-cancer codes on average,

and dropping the most-visited oncologist's codes left the assigned primary cancer unchanged for 92%.

## B.3 Labeling Academic Physicians

This section discusses how I label physicians as academic or non-academic. The resulting measure of academic status is a broadly inclusive proxy of academic status rather than a definitive measure. The label falls short of a definitive measure because of the inherent challenges of binarizing the diversity of organizational arrangements in oncology and the limitations of the data. Below I detail the process, which broadly emulates the approach of Welch and Bindman (2016).

I use the Medicare Data on Provider Practice and Specialty (MD-PPAS) to proxy academic status. For each year, the data contain the name associated with the tax ID number from which the physician billed the most services in that year. I assign academic status based on the name of this tax ID number. The tax ID reflects the organization or individual receiving Medicare payments, and multiple physicians sharing the same primary tax ID are interpreted as belonging to the same organization or group practice (ResDAC, 2024). In the balanced panel, each physician's academic status is fixed at its 2008 value. Exhibits that describe a single year, such as Figure A.12 and Table A.5, instead use the MD-PPAS from that year.

A tax ID name must meet at least one of several possible criteria to be categorized as academic. The tax ID is considered academic if its name contains the name of a recognizable academic institution or other words that signal academic status (e.g., "university", "college"). I also consider a tax ID to be academic if its name and state match the name and state of record for an academic institution's physician group. I also consider a tax ID to be academic if the institution named listed an affiliation with an academic center or university on its website as of May 2024, the time when this determination of academic status was made. I use affiliation in May 2024 rather than the year of the data as I have more limited tools to determine past affiliation status. Finally, I also consider a tax ID number to be academic if

the institution named has a Hematology-Oncology fellowship training program, even if it is not affiliated with a university.

The resulting measure of academic status is broadly inclusive. For example, satellite locations of the Mayo Clinic in Florida and Arizona are counted as academic, as are regional affiliates of Memorial Sloan Kettering Cancer Center. Similarly, the NorthShore University HealthSystem is labeled as academic since its name contains the word "University" and it was affiliated with the University of Chicago as of May 2024. This measure of academic status prioritizes sensitivity (including all plausibly academic physicians) over specificity (definitively excluding all non-academic physicians from the label of "academic").

Some healthcare organizations may operate under multiple tax ID numbers, and providers within the same organization may bill under individual tax IDs. In such cases, the number of oncologists linked to a tax ID may understate the size of the organization's medical oncology practice.

## B.4 Guideline Citations

I build the citation counts in several steps. First, I extract the citations from each guideline using Claude Haiku 4.5 (Anthropic), a large language model. The model reads an image of every page and lists each citation on the page with its authors, title, journal, and publication year. Second, I clean each title by removing punctuation and standardizing spelling. Third, I search for each cleaned title in OpenAlex (Priem, Piwowar, and Orr, 2022). The search returns candidate works published within one year of the citation's publication year, when the citation lists one. It ranks the candidates by a relevance score, which measures how closely each candidate fits the cleaned title. Fourth, I accept a candidate as the citation's match if its score clears a cutoff, or if its score clears a lower cutoff and the candidate's first author also appears among the citation's authors. I set the cutoffs by inspecting candidates just above and below them by hand.

## B.5 Code

I wrote and checked the analysis code with the help of Claude Code (Anthropic), an AI coding tool. The research design and all analysis choices are my own. No AI tool had access to the confidential Medicare claims data. I developed and tested all claims code on simulated claims data, then ran it myself on the secure server that holds the claims. The code includes automated checks on the data. For example, many merges declare the expected key structure (one-to-one or many-to-one) and halt if the data violate it, and assertions halt the run when conditions or assumptions fail. Where possible, results in the paper are read directly from pipeline output files rather than typed by hand.

# C Measurement Error in Physician Specialization

This appendix characterizes the measurement error in estimated physician specialization and describes the approach used to correct it. Physician specialization is estimated from a finite panel of observed patients, and finite sampling biases the estimate upward. The uncorrected HHI overstates true specialization on average, and the overstatement grows as the number of observed patients falls, most severely for generalists. The bias is not classical measurement error, so left uncorrected it can bias the regression estimates rather than only widening their confidence intervals. Every HHI reported in this paper is bias-corrected with a delete-one-patient jackknife, which removes most of the bias. Section C.1 lays out the sampling model and derives the bias in the simple case where each patient counts equally. Section C.2 presents an exact correction for that case and explains why the paper's time-weighted measure requires a different one. Section C.3 presents the jackknife correction that the paper applies. Sections C.4 and C.5 measure in simulation how much bias the correction removes and how much variance it adds. Section C.6 describes the bias that remains at the sample's patient counts and how the research design limits it, and Section C.7 summarizes.

## C.1 The Sampling Model and the Bias

Recall that the main text defines specialization as an HHI over each cancer type's share of the oncologist's time, measured in evaluation and management work relative value units (wRVUs). Write $\tau_{ic}$ for the population share of oncologist $i$'s time spent on cancer type $c$ and $\hat{t}_{ic}$ for the share observed in the data. The estimand is the time-weighted index $\text{HHI}_i = 10{,}000 \sum_c \tau_{ic}^2$, and the uncorrected estimate is $\widehat{\text{HHI}}_i = 10{,}000 \sum_c \hat{t}_{ic}^2$.

The statistical model treats the patient as the unit of random sampling. Physician $i$'s observed patients are indexed $j = 1, \ldots, n_i$. Each patient carries one cancer type, drawn from the physician's true mix with probability $\pi_{ic}$ and $\sum_c \pi_{ic} = 1$, and a weight $w_{ij} > 0$, her total wRVUs for the year. The observed time share $\hat{t}_{ic}$ is then the sum of $w_{ij}$ over observed

patients with cancer type $c$ divided by the sum over all observed patients.

One assumption is maintained throughout: conditional on the physician's true mix, patients' cancer types are independent draws. The data support treating the observed patients as random draws, since the 20% Medicare sample is itself drawn at the beneficiary level, so the observed patients are a random subsample of the physician's Medicare patients.

Consider first the simple case where each patient counts equally, so that $\tau_{ic} = \pi_{ic}$ and $\hat{t}_{ic} = x_{ic}/n_i$, where $x_{ic}$ is the number of observed patients with cancer type $c$, and $\widehat{\text{HHI}}_i = 10{,}000 \sum_c (x_{ic}/n_i)^2$. A physician who treats a single cancer type is estimated without error at any patient count, since every possible draw of her patients yields $\widehat{\text{HHI}}_i = 10{,}000$. For any less specialized physician, the patient count matters.

Suppose instead the physician splits her practice evenly between two cancer types, so that $\text{HHI}_i = 5{,}000$. From one observed patient, I will mechanically estimate $\widehat{\text{HHI}}_i = 10{,}000$. From two observed patients, the smallest value I can estimate is $\widehat{\text{HHI}}_i = 5{,}000$, and the draws in which both patients share a cancer type still yield 10,000. As the number of observed patients grows, the expected estimate approaches the truth from above.

The exact bias follows from three short steps. Each observed count is marginally binomial, $x_{ic} \sim \text{Binomial}(n_i, \pi_{ic})$, so $\mathbb{E}[x_{ic}] = n_i \pi_{ic}$ and $\text{Var}(x_{ic}) = n_i \pi_{ic}(1 - \pi_{ic})$. Because the second raw moment of any random variable is its variance plus its squared mean, the expected squared share is

$$\mathbb{E}\left[\left(\frac{x_{ic}}{n_i}\right)^2\right] = \pi_{ic}^2 + \frac{\pi_{ic}(1 - \pi_{ic})}{n_i}. \tag{C.1}$$

Each observed share is an unbiased estimate of $\pi_{ic}$, but squaring is convex, so the expected square exceeds the squared truth by exactly the variance of the share. Summing over cancer types and applying the identities $\sum_c \pi_{ic} = 1$ and $10{,}000 \sum_c \pi_{ic}^2 = \text{HHI}_i$ gives the bias equation

$$\mathbb{E}[\widehat{\text{HHI}}_i] = \text{HHI}_i + \frac{10{,}000 - \text{HHI}_i}{n_i}. \tag{C.2}$$

The bias is positive whenever $\text{HHI}_i < 10{,}000$ and shrinks in proportion to $1/n_i$, that is, as

the patient count grows. It matches both examples: it is zero at every patient count for the single-type physician and 5,000 at one observed patient for the two-type physician.

The bias has a useful interpretation, borrowed from population genetics, where a similar measure to the HHI estimates the diversity of genes (Nei, 1987). Computing $\widehat{\text{HHI}}_i$ amounts to drawing two patients at random from the observed sample with replacement and asking whether they share a cancer type. With replacement, the same patient can be drawn twice, which happens with probability $1/n_i$, and a patient always shares a cancer type with herself. The guaranteed self-match is the entire bias.

The bias is not classical measurement error. Its mean is not zero, and it varies with patient volume, the number of observed patients $n_i$, which differs across oncologists, so no constant absorbs it. Left uncorrected, the error can therefore bias the regression estimates rather than only widening their confidence intervals.

## C.2 An Exact Correction and Why Time Weighting Breaks It

Because Equation C.2 is linear in $\text{HHI}_i$, it can be inverted exactly. Solving for $\text{HHI}_i$ and replacing the expectation with the observed value defines the closed-form corrected estimator

$$\widehat{\text{HHI}}_i^{\text{cf}} = \frac{n_i \widehat{\text{HHI}}_i - 10{,}000}{n_i - 1}, \tag{C.3}$$

which is unbiased for every $n_i \geq 2$. The same correction gives the unbiased estimate of gene diversity in population genetics (Nei, 1987).

Recall that the paper's measure weights each patient by her share of the physician's time rather than counting patients equally. The weights are related to cancer types: certain cancers absorb more evaluation and management time per patient than others. In the 2009 claims data, a patient-level chi-squared test rejects that a patient's wRVUs are unrelated to her cancer type, among patients of physicians with at least 11 evaluation and management patients ($\chi^2(34) = 15{,}226.2$, $p < 0.001$).

The closed-form correction of Equation C.3 is no longer exact under this dependence. Extending it to weighted shares is exact only if patients' weights are unrelated to their cancer types, because only then can the patient weights be held fixed while the cancer types remain independent draws. The data reject that condition. Rather than assume independence, the paper uses a correction that does not require it.

## C.3 The Jackknife Correction

I compute each physician's HHI $n_i + 1$ times: once from all $n_i$ observed patients, and $n_i$ more times leaving out one patient at a time. Writing $\widehat{\text{HHI}}_{(j)}$ for the HHI computed without patient $j$, the jackknife-corrected HHI is

$$\widehat{\text{HHI}}_i^{\text{jack}} = n_i \, \widehat{\text{HHI}}_i - \frac{n_i - 1}{n_i} \sum_{j=1}^{n_i} \widehat{\text{HHI}}_{(j)}. \tag{C.4}$$

The delete-one-patient jackknife is a standard method for reducing finite-sample bias (Cameron and Trivedi, 2005, section 11.5.5). It removes the part of the bias that shrinks in proportion to $1/n_i$. It requires no assumption about how wRVUs relate to cancer types: the correction is computed entirely from the physician's own observed patients, weights included.

In the equal-weight case, the jackknife reproduces the exact correction of Equation C.3 exactly. Write $Q_i = \sum_c x_{ic}^2$, so that $\widehat{\text{HHI}}_i = 10{,}000\, Q_i / n_i^2$. Removing one patient of type $c$ gives $\widehat{\text{HHI}}_{(j)} = 10{,}000\,(Q_i - 2x_{ic} + 1)/(n_i - 1)^2$. Summing over the $n_i$ patients, where type $c$ appears $x_{ic}$ times and $\sum_c x_{ic} \cdot 2x_{ic} = 2Q_i$, yields $\sum_{j=1}^{n_i} \widehat{\text{HHI}}_{(j)} = 10{,}000\,[(n_i - 2)Q_i + n_i]/(n_i - 1)^2$. Substituting into Equation C.4 and simplifying gives

$$\widehat{\text{HHI}}_i^{\text{jack}} = 10{,}000 \cdot \frac{Q_i - n_i}{n_i(n_i - 1)} = \widehat{\text{HHI}}_i^{\text{cf}}, \tag{C.5}$$

so in the one case where the exact answer is known, the jackknife recovers it.

A smaller bias can remain after the correction. How much remains, at the sample's patient

counts, is an empirical question. The next section measures it in simulation.

## C.4 Evidence on Bias

I measure the remaining bias in a Monte Carlo simulation. Each simulated physician draws patients at random from the aggregate cancer-type distribution of patients in the claims data, which yields a general oncologist with a true HHI of 961 points. This is a conservative choice: by Equation C.2, bias grows as the true HHI falls, so the simulation studies the case with the most bias.

The simulation's remaining inputs are also estimated from the claims data: mean time per patient by cancer type, the coefficient of variation of time per patient within a cancer type, set to its median across types, and the distribution of patients per physician. Each patient-count cell simulates 50,000 physicians and evaluates the uncorrected and the jackknife-corrected HHI against the true time-weighted HHI implied by the parameters.

Figure A.16 plots the mean uncorrected and jackknife-corrected HHI against the number of patients, and Table A.6 reports the estimated bias of each with simulation standard errors. The uncorrected HHI has a mean bias of 1,381 points at 11 patients, falling to 653 points at 25 (the sample minimum), 337 points at 50 (approximately the sample median), and 173 points at 100. As a share of the true HHI, the uncorrected estimator overstates specialization by 68% at 25 patients, 35% at 50, and 18% at 100. A general oncologist with a true HHI of 961 measures about 1,610 on average at the sample's 25-patient minimum. Nearly half of the measured value is bias. The bias is both substantial and uneven across the sample's patient counts.

The jackknife removes most of this bias. Mean bias falls to 175 points at 11 patients, 46 points at 25, 11 points at 50, and 3 points at 100, or 4.8%, 1.2%, and 0.3% of the true HHI at 25, 50, and 100 patients. At the sample minimum and above, the remaining bias is under 10% of the uncorrected bias, and it is much flatter across patient counts.

The second panel of Table A.6 repeats the exercise for a specialist. The 80% breast

cancer specialist draws 80% of her patients from breast cancer and spreads the remaining 20% over the other types in proportion to the aggregate mix, for a true time-weighted HHI of 5,629 points. Her uncorrected HHI has a mean bias of 869 points at 11 patients, 436 points at 25 patients, and 221 points at 50 patients. The jackknife reduces her mean bias to 165, 56, and 8 points, respectively. At the sample minimum and above, the remaining bias is at most 1.0% of her true HHI.

I check the simulation against the claims data. For every physician with at least 50 observed patients in 2009, I compute the jackknife-corrected HHI twice: once from 25 of her patients drawn at random, and once from all of her observed patients. Both estimate the same true HHI, so any average gap between them is measurement error, and the simulation predicts its size. Across 2,191 physicians, the gap averages 38 HHI points (s.e. = 12). The simulation, run through the same exercise for the general oncologist, predicts a gap of 39 points. The claims data show the gap the simulation predicts.

## C.5 Evidence on Noise

The correction buys its bias reduction with slightly more noise. The standard-deviation columns of Table A.6 show that the jackknife-corrected HHI is more dispersed than the uncorrected HHI at every patient count. For the general oncologist at 25 patients, the standard deviation is 462 points against 412, or 12% higher.

Noise falls with patient count. For the general oncologist, the standard deviation of the corrected HHI is 462 points at 25 patients, 281 points at 50, and 183 points at 100. Estimates for physicians observed with more patients are therefore more precise.

Noise is not monotone in the true HHI. With the correction at 25 patients, the standard deviation is zero for a single-cancer specialist (true HHI 10,000 points), 1,871 points for the 80% breast cancer specialist (5,629 points), and 462 points for the general oncologist (961 points). A physician treating a single cancer type is measured with zero error at any patient count, and physicians between the extremes carry the most noise.

Higher variance makes differences and changes in measured HHI harder to detect. Since the variance is not monotone in the true HHI, whether measurement variance is in fact higher for specialists in the claims data is an empirical question. Measurement noise that varies with patient count and with the level of the HHI is heteroskedasticity, which the main regressions' robust standard errors absorb. The noise that remains after the bias correction is near mean zero, so it widens confidence intervals rather than biasing coefficients.

## C.6 What the Paper Does

The paper's baseline measure combines the 25-patient sample minimum with the jackknife correction. Every HHI reported in this paper is the jackknife-corrected value, computed from all of a physician's observed patients in the year.

At the sample's patient counts, the bias that remains is small. Among 2009 physicians with at least 25 patients, the median patient count is 45, and at that count the simulation puts the remaining mean bias at 17 HHI points, against a mean HHI of 1,474. Most of the upward bias in measured specialization is therefore removed before any regression is run.

The research design limits what the remaining bias can do. The main estimates use within-physician changes in HHI over time, and the remaining bias is positive in both years, so it partially cancels in the difference, particularly for physicians whose patient volume is stable. The main results also include a specification with patient-volume fixed effects, which can absorb systematic variation in the remaining bias associated with patient count.

## C.7 Summary

The finite-sample error in measured specialization is large, and it is not classical measurement error. The uncorrected estimator overstates specialization by 68% at the 25-patient sample minimum, and the overstatement varies with patient volume, so the error can bias the regression estimates. In the simulation calibrated to the claims data, the jackknife correction removes more than 90% of this bias at the sample minimum and above, and the remaining bias

partially cancels in within-physician changes. The correction slightly increases the variance of the estimated HHI, which widens confidence intervals, and the variance differs by patient count and by true HHI, so estimates are less precise where measurement variance is higher. The correction and the research design together limit what the remaining measurement error can do to the estimated change in specialization.

# D Randomization Inference

This appendix gives the full construction of the randomization inference summarized in Section 5.4. Every $p$-value and placebo interval in the balance tables, and every $p$-value and confidence interval for the main estimates, is computed by randomization inference, which treats the cancer-level knowledge shocks, not physician-level knowledge shocks, as the source of randomness. The one exception is balance table column (1), whose randomization inference reassigns unadjusted exposure to knowledge growth at random across physicians.

Section D.1 explains the randomization inference test, which compares the coefficient estimated on the observed data with coefficients computed after reassigning the knowledge shocks at random across cancer types. Section D.2 computes one-sided and two-sided $p$-values from the rank of the coefficient estimated on the observed data among the coefficients from the reassignments of the knowledge shocks, rather than assuming a formula for the distribution of a coefficient or test statistic. Section D.3 builds each physician's placebo exposure to knowledge growth from knowledge shocks reassigned in standard-deviation units. Section D.4 applies randomization inference to the balance tables, which check whether physician characteristics measured in 2008 predict exposure to knowledge growth. Section D.5 builds confidence intervals for the main estimates from every effect size the randomization inference test cannot reject. Section D.6 allows for estimation error in the coefficients of the predicted knowledge growth regression.

## D.1 Random Reassignment of Knowledge Shocks

**Exposure to knowledge growth and knowledge shocks.** As defined in Section 3.3, physician $i$'s exposure to knowledge growth $\Delta k_i$ sums knowledge growth $\Delta k_c$ over $C_i$, the set of cancer types physician $i$ treated. The predicted knowledge growth regression, equation (2) in Section 5.3, is a Poisson regression that predicts each cancer type's knowledge growth from observable cancer characteristics. Write $\delta$ for the vector of its four coefficients, $\hat{\delta}$ for the

estimate of $\delta$, and $\widehat{\mathbb{E}}[\Delta k_c]$ for the predicted knowledge growth of cancer type $c$. Cancer type $c$'s knowledge shock is realized minus predicted knowledge growth, $\Delta k_c - \widehat{\mathbb{E}}[\Delta k_c]$. Physician $i$'s recentered exposure to knowledge growth $\Delta \tilde{k}_i$ sums the knowledge shocks over $C_i$.

**Logic of the randomization inference test.** The realized coefficient $\hat{\beta}$ is the estimate, on the observed data, of the coefficient of interest, such as $\beta$ in equation (1). Every randomization inference test in this appendix asks whether a coefficient, or, for the joint test of Section D.4, an $F$-statistic, as extreme as the realized one would be unlikely under a null hypothesis. Rather than assuming a formula for the distribution of a coefficient or test statistic under the null hypothesis, randomization inference builds that distribution from placebo draws, defined below.

**Placebo draws.** The tests for the main estimates and for balance table columns (2) to (4) of Section D.4 reassign the knowledge shocks across cancer types. For the main estimates, the null hypothesis is that knowledge growth has no effect, and the maintained assumption is that the knowledge shocks are as good as randomly assigned. For balance table columns (2) to (4), the null hypothesis is that the knowledge shocks are as good as randomly assigned.

These tests follow three steps. First, reassign the knowledge shocks across the 35 cancer types. Second, rebuild each physician's recentered exposure to knowledge growth from the reassigned knowledge shocks. Third, recompute the regression coefficient. Each reassignment of the knowledge shocks is one placebo draw (a placebo shock realization in Section 5.4). The test for balance table column (1) instead reassigns unadjusted exposure to knowledge growth ($\Delta k_i$) across physicians.

I use 999 placebo draws, indexed by $b = 1, \ldots, 999$, which give 999 placebo coefficients $\hat{\beta}^{(b)}$.

## D.2 $p$-Values

Each $p$-value is based on the realized coefficient's rank among 1,000 coefficients, the 999 placebo coefficients of Section D.1 plus the realized coefficient itself.

**Upper one-sided $p$.** The upper one-sided $p$ measures how far into the upper tail of the coefficient's distribution under the null hypothesis the realized coefficient sits. Count the coefficients at or above the realized coefficient, out of the 1,000, with the realized coefficient counted among them (the +1 below),

$$p^{+} = \frac{1 + \#\{b : \hat{\beta}^{(b)} \geq \hat{\beta}\}}{1{,}000}. \tag{D.1}$$

Its smallest possible value is 0.001, since the count includes the realized coefficient itself.

The joint test, defined in Section D.4, applies the same formula to the $F$-statistic in place of the coefficient, since only a large $F$-statistic is evidence against the null hypothesis.

**Two-sided $p$.** The two-sided $p$ is used for the regression coefficients in the balance tables and the main estimates. In both, a coefficient in either tail is evidence against the null hypothesis. The lower one-sided $p$ mirrors the upper one-sided $p$,

$$p^{-} = \frac{1 + \#\{b : \hat{\beta}^{(b)} \leq \hat{\beta}\}}{1{,}000}. \tag{D.2}$$

Take the smaller of the two one-sided $p$-values and double it, with a ceiling of 1,

$$p = \min\left(1,\ 2\min[p^{-},\ p^{+}]\right). \tag{D.3}$$

Each one-sided $p$ is at least 0.001, so the smallest two-sided $p$ is 0.002.

## D.3 Placebo Exposure to Knowledge Growth

**Standardized knowledge shocks.** For the main estimates and balance table column (4), each placebo draw of Section D.1 reassigns the knowledge shocks across cancer types as standardized Poisson residuals.

The standardized Poisson residual is

$$\Delta\tilde{k}_c^{\mathrm{P}} = \frac{\Delta k_c - \widehat{\mathbb{E}}[\Delta k_c]}{\sqrt{\widehat{\mathbb{E}}[\Delta k_c]}}. \tag{D.4}$$

In words, it is cancer type $c$'s knowledge shock (realized knowledge growth minus predicted knowledge growth), divided by the Poisson-implied standard deviation of cancer type $c$'s knowledge growth, $\sqrt{\widehat{\mathbb{E}}[\Delta k_c]}$. A Poisson variance equals its mean, so the division puts every cancer type's knowledge shock in standard-deviation units.

**Placebo knowledge growth.** In placebo draw $b$, the standardized Poisson residuals are reassigned at random across the 35 cancer types. Cancer type $c$ inherits the standardized Poisson residual of cancer type $\sigma_b(c)$, where $\sigma_b$ is a random permutation of the cancer types. Cancer type $c$'s placebo knowledge growth in placebo draw $b$ is

$$\Delta k_c^{(b)} = \widehat{\mathbb{E}}[\Delta k_c] + \sqrt{\widehat{\mathbb{E}}[\Delta k_c]}\ \Delta\tilde{k}^{\mathrm{P}}_{\sigma_b(c)}. \tag{D.5}$$

In words, cancer type $c$ keeps its own predicted knowledge growth and adds a random residual, scaled by cancer type $c$'s own Poisson-implied standard deviation.

In placebo draw $b$, physician $i$'s placebo exposure to knowledge growth sums the placebo knowledge shocks over the cancer types physician $i$ treated,

$$\Delta\tilde{k}_i^{(b)} = \sum_{C_i} \left(\Delta k_c^{(b)} - \widehat{\mathbb{E}}[\Delta k_c]\right). \tag{D.6}$$

Each physician's cancer types stay the same in every draw; only the knowledge growth of

those cancer types changes.

**Coefficients held fixed.** The predicted knowledge growth regression is estimated once, on the observed data, and its coefficients $\hat{\delta}$ are held fixed across all placebo draws. I call randomization inference with $\hat{\delta}$ held fixed the baseline version of randomization inference. Borusyak and Hull (2021) note that holding fixed the estimated parameters of the process that assigns the shocks, here $\hat{\delta}$, is a simpler alternative to their worst-case treatment of estimation error. Section D.6 builds versions of randomization inference that do not hold $\hat{\delta}$ fixed.

## D.4 Balance Tables

In each of four columns, the balance tables report a placebo interval for each physician characteristic and a joint test of all the characteristics together.

**Balance on physician characteristics.** If the knowledge shocks are as good as randomly assigned, physician characteristics measured in 2008 should not predict the knowledge shocks of the cancer types a physician treats, and so should not predict recentered exposure to knowledge growth. The balance tables in Section 5.5 test that prediction by regressing exposure to knowledge growth on those characteristics. A characteristic that predicts exposure to knowledge growth more strongly than it does in the placebo draws is evidence that the knowledge shocks are not as good as randomly assigned.

**Balance table columns.**

- *Column (1), unadjusted exposure.* Column (1) reassigns unadjusted exposure $\Delta k_i$ itself at random across physicians, under the null hypothesis that unadjusted exposure is randomly assigned across physicians. Each such reassignment is one placebo draw.

- *Column (2), recentered on the mean.* Column (2) subtracts the mean growth across all cancer types, $\overline{\Delta k}$, in place of predicted knowledge growth. Cancer type $c$'s knowledge

shock is then realized knowledge growth minus $\overline{\Delta k}$,

$$\Delta k_c - \overline{\Delta k}. \tag{D.7}$$

Each placebo draw reassigns these knowledge shocks across the 35 cancer types. Physician $i$'s recentered exposure is rebuilt by summing the reassigned knowledge shocks over $C_i$, the cancer types physician $i$ treated.

- *Columns (3) and (4), recentered with a Poisson regression.* Column (3) reassigns the standardized Poisson residuals from a Poisson regression on prevalence alone. Column (4) reassigns the standardized Poisson residuals from the predicted knowledge growth regression, which adds an indicator for high-quality surrogate endpoints and its interaction with prevalence. The tests for the main estimates reassign the same residuals.

**Reported statistics.** Like the balance table in Borusyak and Hull (2023), the balance tables report standard errors and the joint test's $p$-value, and no confidence interval. I also report, for each row, the placebo interval.

**Placebo interval.** The placebo interval runs from the 2.5th to the 97.5th percentile of the placebo coefficients across the 999 placebo draws,

$$\big[\, q_{2.5}\big(\hat{\beta}^{(b)}\big),\ q_{97.5}\big(\hat{\beta}^{(b)}\big)\,\big]. \tag{D.8}$$

The placebo interval is not a confidence interval but the middle 95% of the coefficient's distribution under the null hypothesis. Each balance table column's placebo interval comes from that column's own placebo draws. In balance table column (4), these are the draws of the baseline version of randomization inference, which hold the coefficients $\hat{\delta}$ fixed.

The placebo interval lies around zero, where the placebo coefficients fall. The realized coefficient lies outside the placebo interval exactly when the two-sided $p$ is at or below 0.05.

**Joint test.** Each row tests one characteristic. The joint test asks whether the physician characteristics, taken together, predict a balance table column's exposure measure better than they predict the placebo version of that exposure measure in that column's placebo draws. The statistic is the standard $F$-statistic for joint significance in the regression of that exposure measure on all the physician characteristics. I compute it on the observed data and on each placebo draw. The joint test's $p$-value is the upper one-sided $p^+$ of Section D.2 applied to the $F$-statistic.

## D.5 Test-Inversion Confidence Intervals

For the main estimates, the randomization inference test of Section D.1 has the null hypothesis that knowledge growth has no effect. Each confidence interval comes from repeating that test for each candidate value of $\beta$, in six steps. A computational shortcut speeds up those steps.

**Comparison with conventional confidence intervals.** The main estimates are point estimates of $\beta$, the effect of knowledge growth on specialization. A confidence interval gives the range of effect sizes consistent with the data. A conventional confidence interval treats physicians as independent observations and relies on asymptotics in the number of physicians. Randomization inference makes neither assumption.

Here the randomness comes from the 35 knowledge shocks. For each candidate effect size, I apply the randomization inference test of Section D.1, with the same placebo draws and the two-sided $p$-value of Section D.2, to an adjusted outcome that subtracts the candidate effect times recentered exposure to knowledge growth. The confidence interval collects every candidate effect size this test cannot reject. The procedure is the test-inversion construction of Borusyak and Hull (2021).

**Sharp null hypothesis.** Fix a candidate effect $\beta_0$ (the subscript indicates the sharp null hypothesis being tested). The sharp null hypothesis is that the true effect is exactly $\beta_0$ for every physician. Write $y_i$ for the physician's outcome (the change in specialization or in

the count of cancer types), and recall that $\Delta\tilde{k}_i$ is the physician's recentered exposure. The adjusted outcome is

$$y_i - \beta_0 \, \Delta\tilde{k}_i. \tag{D.9}$$

If $\beta_0$ is the true effect, subtracting $\beta_0 \, \Delta\tilde{k}_i$ removes exactly the effect of the knowledge shocks. The adjusted outcome then does not depend on the knowledge shocks, so the randomization inference test of a zero effect from Section D.1 applies to it exactly. The confidence interval for the true effect $\beta$ is every candidate $\beta_0$ that test fails to reject at 5%, using the two-sided $p$.

The test-inversion intervals are confidence intervals in the usual sense, under the maintained assumption that the knowledge shocks are as good as randomly assigned, with the coefficients $\hat{\delta}$ held fixed. They contain the true effect with probability at least 95% over reassignments of the knowledge shocks, and zero lies outside exactly when the test of a zero effect rejects at 5%.

**Mechanics.** Before the steps below, I regress the outcome, observed recentered exposure and each placebo draw's placebo exposure to knowledge growth on the controls $X_i$ in equation (1), and keep the residuals. By the Frisch–Waugh–Lovell theorem, this gives the same coefficient on exposure as including the controls in every regression. The procedure has six steps.

1. At candidate $\beta_0$, form the adjusted outcome.

2. Regress the adjusted outcome on observed recentered exposure $\Delta\tilde{k}_i$. The coefficient is $\hat{\beta} - \beta_0$.

3. Regress the adjusted outcome on each placebo draw's placebo exposure $\Delta\tilde{k}_i^{(b)}$ to get the 999 placebo coefficients.

4. The $p$-value at $\beta_0$ is the two-sided $p$ of Section D.2 over these 1,000 coefficients.

5. The confidence interval for $\beta$ is every candidate $\beta_0$ with $p$ above 0.05. Each endpoint is found by bisection, halving the gap between a candidate the test does not reject and

one it rejects. The search stops when the gap is below $10^{-6}$ times the larger of 1 and the endpoint's absolute value.

6. If no candidate above $\hat{\beta}$ is rejected, the confidence interval is unbounded above. If no candidate below $\hat{\beta}$ is rejected, the confidence interval is unbounded below.

**A computational shortcut.** The shortcut replaces the regressions in steps 2 and 3 and leaves every coefficient and $p$-value unchanged. The shortcut computes the placebo coefficient $\hat{\beta}^{(b)}$ at candidate $\beta_0$, the slope from regressing the adjusted outcome on the placebo exposure. A regression slope is a covariance ratio,

$$\hat{\beta}^{(b)}(\beta_0) = \frac{\operatorname{cov}\left(y_i - \beta_0\,\Delta\tilde{k}_i,\ \Delta\tilde{k}_i^{(b)}\right)}{\operatorname{var}\left(\Delta\tilde{k}_i^{(b)}\right)}. \tag{D.10}$$

Covariance is linear in each argument,

$$\operatorname{cov}\left(y_i - \beta_0\,\Delta\tilde{k}_i,\ \Delta\tilde{k}_i^{(b)}\right) = \operatorname{cov}\left(y_i,\ \Delta\tilde{k}_i^{(b)}\right) - \beta_0\,\operatorname{cov}\left(\Delta\tilde{k}_i,\ \Delta\tilde{k}_i^{(b)}\right). \tag{D.11}$$

Divide through by $\operatorname{var}\left(\Delta\tilde{k}_i^{(b)}\right)$ and name the two slopes that appear,

$$c_b = \frac{\operatorname{cov}\left(y_i,\ \Delta\tilde{k}_i^{(b)}\right)}{\operatorname{var}\left(\Delta\tilde{k}_i^{(b)}\right)} \qquad d_b = \frac{\operatorname{cov}\left(\Delta\tilde{k}_i,\ \Delta\tilde{k}_i^{(b)}\right)}{\operatorname{var}\left(\Delta\tilde{k}_i^{(b)}\right)}. \tag{D.12}$$

So the placebo coefficient at $\beta_0$ is

$$\hat{\beta}^{(b)}(\beta_0) = c_b - \beta_0\, d_b. \tag{D.13}$$

The slopes $c_b$ and $d_b$ are computed once per placebo draw and reused for every candidate $\beta_0$ to compute the $p$-value at that candidate. For step 2, the same algebra with $\Delta\tilde{k}_i$ in place of $\Delta\tilde{k}_i^{(b)}$ gives $\hat{\beta} - \beta_0$.

## D.6 Estimation Error in Predicted Knowledge Growth

Two extensions of randomization inference allow for estimation error in the coefficients $\hat{\delta}$ of the predicted knowledge growth regression.

**Source of the estimation error.** Sections D.1 through D.5 use the baseline version of randomization inference defined in Section D.3, which holds the coefficients $\hat{\delta}$ fixed. The coefficients $\hat{\delta}$ are estimates, so they contain estimation error. That error carries into predicted knowledge growth, the knowledge shocks, and the standardized Poisson residuals that are reassigned.

The coefficients $\hat{\delta}$ are estimated from the same 35 cancer types whose standardized Poisson residuals are reassigned. Borusyak and Hull write that in some cases confidence intervals computed with the estimated parameters held fixed (here $\hat{\delta}$) "are asymptotically correct [. . . ], although general conditions for this are unknown" (Borusyak and Hull, 2021, appendix C.5).

**Two extensions of randomization inference.** One extension of randomization inference assumes a normal distribution for $\hat{\delta}$ and draws $\delta$ from it. I call this the normal-draws version of randomization inference. The other extension of randomization inference takes the worst case, that is, the largest $p$-value over the values of $\delta$ that the data cannot rule out, following Berger and Boos (1994), as Borusyak and Hull (2021) propose when the process that assigns the shocks has parameters that must be estimated. I call this the Berger–Boos version of randomization inference.

**Normal-draws version of randomization inference.** The normal-draws version of randomization inference runs in five steps.

1. Draw $\delta$ 200 times from the estimated normal distribution of $\hat{\delta}$, with mean $\hat{\delta}$ and the robust covariance matrix of $\hat{\delta}$.

2. At each normal draw of $\delta$, recompute predicted knowledge growth, then the standardized Poisson residuals against that predicted knowledge growth.

3. Reassign those standardized Poisson residuals and rebuild the placebo exposures to knowledge growth as in Section D.3.

4. Recompute the coefficient on observed recentered exposure to knowledge growth with that normal draw's predicted knowledge growth, then the 999 placebo coefficients and the two-sided $p$ exactly as in Section D.2.

5. The reported $p$ is the mean of the 200 $p$-values, one per normal draw.

The confidence interval comes from test inversion as in Section D.5, using the mean $p$ and the same bisection.

The normal-draws version of randomization inference accounts for estimation error on average, not in the worst case, and assumes $\hat{\delta}$ is normally distributed. Normality is a large-sample approximation, and $\hat{\delta}$ is estimated from only 35 cancer types.

**Berger–Boos version of randomization inference.** It runs in five steps.

1. Take a confidence set for $\delta$, a region that contains the true $\delta$ with approximately 99.9% probability in large samples. I use the ellipsoid implied by the robust covariance matrix of $\hat{\delta}$.

2. Cover the confidence set with a fixed grid of 1,009 possible values of $\delta$:

   - $\hat{\delta}$ itself, the center of the confidence set (1 point).
   - Two points on the boundary in each of four directions, one on each side of $\hat{\delta}$ (8 points). The directions are the columns of the Cholesky factor of the robust covariance matrix of $\hat{\delta}$.
   - 600 randomly drawn points on the boundary of the confidence set (600 points).
   - Two points inside the confidence set toward each of 200 of those boundary points, one halfway from $\hat{\delta}$ and one three-quarters of the way (400 points total).

3. At each grid point, recompute predicted knowledge growth and the standardized Poisson residuals using the grid point's value of $\delta$ instead of $\hat{\delta}$. Then rerun the randomization inference test of Section D.1, recomputing the coefficient on observed recentered exposure with that grid point's predicted knowledge growth. The coefficient reported in the tables, $\hat{\beta}$, is still the one computed at $\hat{\delta}$. Each grid point's test uses level 4.9% instead of 5%. The remaining 0.1% allows for the chance that the confidence set misses the true $\delta$, so the overall level is at most 5% in large samples.

4. The reported $p$ is the largest $p$ across the 1,009 grid points, plus 0.001 for that 0.1% chance, capped at 1 because a $p$-value cannot exceed 1.

5. Each grid point also gives its own confidence interval, by the test inversion of Section D.5 at level 4.9%. The Berger–Boos confidence interval runs from the smallest lower endpoint to the largest upper endpoint across those 1,009 intervals.

Ideally, step 4 would take the largest $p$ over the entire confidence set, not only over the grid. The largest $p$ can sit between grid points, so the reported $p$ can understate, but never overstate, the $p$ that step 4 would give over the entire confidence set.

**Berger–Boos versus baseline version of randomization inference.** The Berger–Boos version of randomization inference is conservative by construction. Its $p$ is never below the $p$ from the baseline version of randomization inference, and its confidence interval always contains the baseline confidence interval. Both hold because $\hat{\delta}$ is one of the 1,009 grid points. The largest $p$ over the grid is at least the $p$ at $\hat{\delta}$. The Berger–Boos interval also contains the interval at $\hat{\delta}$, which uses level 4.9% rather than 5% and so contains the baseline interval.

**Baseline version of randomization inference in the balance tables.** In a balance test, a small $p$ flags imbalance. The Berger–Boos $p$ is never smaller than the baseline $p$, so it would flag imbalance less often. The baseline version of randomization inference is therefore the stricter test of balance, and the balance tables report it.

**Versions of randomization inference reported.** For the main estimates, the paper's $p$-values and confidence intervals come from the baseline version of randomization inference. Appendix E shows that, for the specialization estimates, the normal-draws confidence intervals are wider than the baseline confidence intervals but lead to the same conclusions. For the main estimates, the Berger–Boos confidence intervals are unbounded, so they rule out nothing. Appendix E notes this, and no table prints them, because they are uninformative for every estimate.

# E Robustness Checks

The main findings rest on several design choices. Each robustness check below changes one of these choices and compares its estimates with the main estimates. The choices the robustness checks vary include how the confidence intervals are built, how knowledge growth is measured and recentered, when a physician counts as exposed to a cancer's guideline growth, and which visits build the HHI.

The robustness checks below omit the fixed effects for physician characteristics. Identification does not rest on these fixed effects. The balance tests in Section 5.5 find no evidence that the six physician characteristics are jointly related to recentered exposure to knowledge growth, so there is little sign of the omitted variable bias that the fixed effects would correct. In addition, across most regressions, adding the fixed effects leaves the confidence intervals the same or wider, reducing precision in many estimates. In the main results, including the fixed effects changes no conclusion.

The main conclusions hold when the confidence intervals allow for error in the estimate of predicted knowledge growth, except under the worst-case confidence intervals. The baseline version of randomization inference holds the coefficients that predict knowledge growth from cancer characteristics (equation (2)) fixed at their estimated values (Borusyak and Hull, 2023), which could overstate precision. One extension instead draws these coefficients from their estimated distribution, assuming the estimation error is normally distributed. The generalist confidence interval then widens by about 5% and still rules out increases above 68 HHI points per 100 pages. The specialist confidence interval widens by about 15% and still includes zero. The other extension builds worst-case confidence intervals, which use the least favorable coefficients the data allow (Berger and Boos, 1994), in line with the approach of Borusyak and Hull (2023) (Appendix D.6). The worst-case confidence intervals are unbounded for every estimate and so rule out nothing.

Three other measures of knowledge growth lead to the same conclusions as guideline pages, for generalists and specialists (Table A.7). Two of the measures track the scientific

literature behind the guidelines more directly than page counts: the increase in the number of guideline citations and the net increase in scholarly works cited. For generalists, the confidence intervals with these two measures rule out rises above 19 HHI points per 100 citations and 22 HHI points per 100 scholarly works. The third measure treats a cancer type's first guideline as documentation of knowledge that already existed rather than as knowledge growth. This approach uses the page count of the first guideline as the baseline page count for the three cancer types whose first guideline appears during the study period. With this approach, the generalist confidence interval still has an upper bound of 66 HHI points per 100 pages, essentially the same bound as with guideline pages. For specialists, the confidence intervals remain wide under all three measures, so the data still neither confirm nor rule out a meaningful specialization response to knowledge growth.

Scaling page growth by each physician's baseline pages also leaves the conclusions unchanged. Relative knowledge growth is measured by dividing recentered exposure to knowledge growth by a physician's total baseline pages in 2009, counting a first guideline that appears during the study period as baseline pages. The same absolute page growth is a larger relative growth for a physician with few baseline pages, so those physicians contribute more variation and thus carry more weight than in the main specification. For generalists, the confidence interval rules out rises above 714 HHI points for page growth equal to a physician's baseline pages. For a generalist with the average number of baseline pages, this result reflects at most a rise of 61 HHI points per 100 pages, essentially the same as without scaling. For specialists, the confidence interval is about eight times as wide, so the data again leave open whether they respond meaningfully to knowledge growth.

Recentering exposure to knowledge growth in two other ways leads to the same conclusions for generalists and specialists. Recentering by each cancer's 2008 incidence instead of its prevalence changes little (Table A.8). The generalist confidence interval still has an upper bound of 71 HHI points per 100 pages, and the specialist confidence interval remains about as wide as in the main specification. The other modification of recentering adds each guideline's

2009 length to the prediction of knowledge growth, because initial length could also predict future growth (Table A.7, Column 4). The generalist confidence interval then rules out rises above 74 HHI points per 100 pages, and the specialist confidence interval remains about as wide as in the main specification. The main results therefore hold whether the prediction of knowledge growth uses incidence or prevalence and whether it includes 2009 length.

The conclusions also hold when a physician must have more observed patients with a cancer type to count as exposed to its guideline growth (Table A.8). The main specification counts a physician as exposed after a single observed patient, so one-off cases could drive the result. However, patient counts come from a 20% sample of Traditional Medicare claims, so one observed patient corresponds to roughly five Traditional Medicare patients, and more across the physician's entire practice. A higher threshold may remove one-off noise, but it may also discard cancer types that are genuinely part of the physician's practice. Ultimately, requiring two or three patients leaves the generalist result unchanged: the estimate is small and not distinguishable from zero. The generalist confidence intervals rule out rises above 59 HHI points per 100 pages at two patients and 97 HHI points at three. The specialist estimates are larger than in the main specification at both thresholds, but their confidence intervals remain wide and include zero.

Building the HHI only from new patient visits also supports the main conclusions (Table A.8). Physicians could respond to knowledge growth only in which new patients they take on, a response that including established patients could hide. However, generalists show no such response: the estimate is almost unchanged, and the confidence interval reaches at most 85 HHI points per 100 pages. The specialist confidence interval is about twice as wide as in the main specification and includes zero, in part because the sample is only 38% of the main sample.