Knowledge Growth and Specialization: Evidence from Oncologists *

Maya Lozinski
Massachusetts General Hospital
Harvard Medical School

August 2025

Abstract

How does the expansion of general knowledge influence worker specialization and the adoption of new technologies? This study empirically addresses this question within the field of medical oncology, which has experienced rapid knowledge growth, as evidenced by the evolution of cancer treatment guidelines. A small proportion of initially specialized oncologists respond to this growth by further narrowing their focus. In contrast, the majority remain generalists and do not increase their level of specialization over time, even in the phase of expanding knowledge. These divergent responses appear to reflect differing constraints: extremes of specialization are significantly influenced by scale, with specialization increasing substantially with organization size —particularly in academic institutions—and the size of the local healthcare market. However, generalists encounter coordination costs that persist regardless of organizational or market scale. Importantly, specialization is increasingly associated with differences in patient care. Over time, specialists are more likely to prescribe recently approved anti-cancer drugs and have greater clinical experience with their medications compared with generalists. These findings suggest that, in response to rapid knowledge growth, some professionals may not specialize further but instead risk lagging behind the technological frontier.

*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.org.

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

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 operating 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 plausible 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, market, or structural constraints. In such cases, experts are faced with a dilemma: either bear the growing costs of maintaining expertise across multiple domains or risk falling behind the frontier of knowledge.

This study investigates how expert workers respond to knowledge growth in the context of medical oncology. Specifically, 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: 4 years of medical school, followed by 3 years of internal medicine residency and at least 2 years of fellowship training in oncology. Their practice centers on complex cognitive and communicative tasks, and they are primarily consumers—not producers—of scientific knowledge (Jain et al., 2019; Schlafly & Sebro, 2022; Taylor, 2004).

Despite these significant initial investments in human capital, oncologists must engage in continual learning to remain at the forefront of their rapidly evolving field. Cancer affects a vast portion of the population—the lifetime risk of developing cancer is approximately 40% (American Cancer Society, 2024)[1]. This high disease burden fuels both innovation

[1]Fortunately, many cancer treatments are life-extending or curative; the lifetime risk of dying from cancer is significantly lower at 20%.

and rapid knowledge growth. Indeed, one article in a popular science magazine observed that "incorporating this torrential evolution of knowledge is an impossible, Sisyphean task" (Wallis, 2023). 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 examine the consequences of accelerating knowledge accumulation, with clear implications for clinical care and population 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 study, changes in guideline length and citation counts serve as proxies for the expansion of clinically relevant knowledge. A principal advantage of these guideline-based measures is that they reflect the operational body of knowledge that oncologists must engage with in their daily practice.

The analysis confirms substantial growth in oncology-related knowledge over time. Between 2002 and 2020, the total number of pages in NCCN screening and treatment guidelines increased from 1,129 pages to 5,825 pages. This expansion corresponds closely with an increase in the relevant bio-medical literature and coincides with the approval of a significantly greater number of 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 construct a 21-year panel of Medicare claims data and measure annual specialization at the physician level. Interestingly, while highly specialized oncologists have become increasingly specialized the median oncologist has not. In fact, median specialization has declined slightly, despite the marked expansion of the clinical knowledge base.

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

However, this measure may suffer from omitted variable bias, as physicians' choice of subfields is not random. To address this, I apply a recentering procedure adapted from Borusyak and Hull (2023), which removes predictable variation in exposure due to the number of cancer types treated, cancer prevalence, and the presence of surrogate endpoints. The residual variation—interpreted as exposure to "unpredictable" knowledge shocks (e.g., success in clinical trials or discovery of novel molecular targets)—is used in subsequent analyses. Reassuringly, this procedure eliminates correlations between the measure and observable confounders such as patient volume and market size.

Using the adjusted measure, I find that exposure to knowledge growth affects specialists and generalists differently. Oncologists who are already specialized respond to knowledge growth, increasing their specialization by approximately 196 Herfindahl—Hirschman Index (HHI) points per 100-page increase in guideline length—a 4% increase in HHI for a 13% increase in relevant knowledge. In contrast, generalists exhibit a muted response: they treat slightly fewer unique cancers and with limited increases in overall specialization.

Several constraints appear to underlie these divergent responses. Specialist oncologists are strongly constrained by organizational scale. Their degree of specialization increases with institution size, particularly within academic settings, which house a disproportionate share of specialists. Moreover, the maximum level of specialization observed within a given market scales with market size, and this relationship has strengthened over time. Consequently,

geographic disparities in access to highly specialized care have grown, pointing to increasing market-level economies of scale in oncology.

In contrast, general oncologists—who typically work outside academic institutions—face persistent coordination costs that are not alleviated by larger organizations or markets. In non-academic settings, specialization does not increase with organizational size, and the median market-level specialization remains relatively flat with respect to market size. These findings suggest that, for generalists, the returns to specialization may be low or the associated costs prohibitively high, regardless of local market size.

Given these disparities, an important question arises: does specialization affect patient care? Analysis of prescribing behavior indicates that it does. Over time, specialists have prescribed fewer unique cancer drugs, suggesting greater familiarity and depth of experience with each. Furthermore, generalists have lagged in adopting new cancer therapies, especially breakthrough drugs. These practice differences have widened over time, implying that specialization is increasingly associated with higher-quality care.

This study makes several contributions to the literature on healthcare economics and labor specialization. First, it offers a novel empirical measure of knowledge growth based on guideline evolution—a more clinically grounded 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. I extend this line of injury by quantifying the growth in content across all cancer types and documenting the turnover in citations over time. In the economics literature, Glied and Lleras-Muney (2008) measures technological change in health care using stocks and flows of drug approvals. However, drug approval–based approaches face certain limitations that guideline-based measures can overcome. Specifically, clinical guidelines reflect the full 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. Importantly, guidelines are curated to include only clinically relevant knowledge, excluding outdated information—such

as therapies no longer in use—thus capturing the current knowledge set most pertinent to clinical decision-making.

Patents and drug approvals have also served as proxies for innovation in an extensive literature (Alcacer & Gittelman, 2006; Roach & Cohen, 2013). A key advantage of using clinical guidelines over patents or drug approvals in this setting is that guidelines better measure the volume of knowledge directly relevant to the day-to-day work of clinicians rather than the historical accumulation of knowledge. For example, the chemotherapy agent sirolimus has a compelling and extensive history, encompassing numerous patents and scholarly articles—rich enough to warrant a dedicated volume (Duffin, 2019). However, this historical narrative holds little relevance for the routine clinical responsibilities of oncologists and is not included in clinical guidelines.

Notably, this study examines a research question similar to that posed by Jones (2009), namely how the growth of knowledge influences specialization, but it arrives at different conclusions. Jones (2009) empirically demonstrated that inventors increasingly face a greater "depth of knowledge" over time—quantified through patent citations)—which, in turn, drives increased specialization. Similarly, a small number of highly specialized oncologists respond to expanding knowledge by becoming even more specialized. However, the majority of oncologists remain generalists and do not exhibit increased specialization over time, nor do they respond to knowledge growth in the same way. A key explanation for these contrasting results may lie in the fundamental differences between inventors and physicians. While inventors are primarily generators of new knowledge, most oncologists—particularly generalists—serve as users of existing knowledge. This distinction suggests that users may experience the burden of knowledge accumulation more lightly than creators, as applying existing knowledge is arguably less demanding than producing it. Furthermore, as documented in this study, knowledge users may opt to lag behind the technological frontier when the cost of remaining up becomes prohibitly high—an option that may not be available to inventors. In medicine, as in many other domains, the majority of professionals are knowledge users rather than

creators. Consequently, the implications of knowledge growth for such workers may differ significantly from what previous studies on knowledge creators have suggested.

This research also contributes to the broader literature on labor specialization. Consistent with findings by Baumgardner (1988) and Baicker and Chandra (2004), this work finds that physicians tend to be more specialized in larger markets and within larger healthcare organizations, even when controlling for narrowly defined specialties. Importantly, the study extends these findings by demonstrating that such specialization has increased over time. Additionally, the results replicate a key insight from (Atalay, Sotelo, and Tannenbaum, 2024) , who find that the task composition of work differs significantly by market size, even within tightly defined occupational categories. Notably, this study provides suggestive evidence that physician specialization is associated with improvements in care quality—a contrast to the findings of Baicker and Chandra (2004). The difference may partly be explained by the long time period analyzed here, during which the returns to specialization appear to have increased.

Moreover, this study adds to a well-established empirical literature on the effects of technological change in health care. Prior research has often focused on the adoption of specific technologies and their varying economic impacts on healthcare providers and patients (e.g., Arrow et al., 2020; Chandra et al., 2014; Skinner, 2011; Skinner & Staiger, 2015). In contrast, this study considers the aggregate effects of technological accumulation, using the cumulative stock of drug approvals as a proxy for overall knowledge growth. Existing evidence suggests that such cumulative technological change exacerbates educational disparities in health outcomes (Glied & Lleras-Muney, 2008). Building on this, the current analysis investigates how knowledge growth affects healthcare workers and contributes to geographic disparities in healthcare provision. The findings align with those of Dingel et al. (2023), who identify significant market-level economies of scale in health care. This study advances their work by demonstrating that these economies of scale have likely intensified over time, with knowledge growth emerging as a central explanatory factor.

2 Conceptual Framework

This section presents a theoretical framework to consider how oncologists respond to knowledge growth. Medical oncologists, who treat cancer primarily with pharmaceutical interventions rather than surgery or radiation, can specialize horizontally (within disease types) and vertically (within knowledge hierarchies). Although traditional economic theory posits that market size restricts specialization, coordination costs can also limit the extent of labor division. As clinical guidelines become more complex, both the costs of acquiring knowledge and the potential returns to specialization increase, which may further incentivize specialization. Consequently, knowledge growth could exacerbate differences in specialization between large and small healthcare markets: large markets can support more specialized oncologists, while small markets may only support generalists, contributing to geographic inequalities in access to specialized cancer care.

2.1 Medical Oncology

This research focuses specifically 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. With respect to modality, they are based primarily on drug-based therapies, unlike their surgical or radiation-oncology counterparts. Their training pathways also differ significantly to reflect this clinical focus[2]. Thus, medical oncology constitutes a distinct specialty with its own professional identity and set of clinical responsibilities.

2.2 Horizontal Specialization

The costs and benefits of horizontal specialization among oncologists can be analyzed using classic economic models of division of labor. Medical oncologists may specialize by cancer

[2]While surgical oncologists first specialize in a specific anatomical region and then in cancer surgery, medical oncologists complete internal medicine training followed by a fellowship in medical oncology.

type—for example, focusing exclusively on breast cancer or lung cancer—thereby increasing productivity and clinical efficacy. Such specialization reduces redundant learning efforts, as each oncologist needs to stay updated only on advancements within their specific domain. This focus also facilitates "learning by doing," enabling, for instance, a breast cancer specialist to become highly adept at managing drug regimens and associated side effects, ultimately improving patient outcomes.

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

However, Becker and Murphy (1992) argue that the division of labor—particularly in the context of cancer care—distributes responsibilities across a greater number of workers, thereby incurring substantial coordination costs. These costs can manifest in various forms, including communication inefficiencies, incentive misalignment, and hold-up problems. For instance, when treating cancer patients, financial incentives may encourage oncologists to retain patients rather than refer them to specialists with greater expertise, ultimately hindering optimal care delivery. The optimal level of specialization is determined by weighing its productivity benefits against these coordination costs. In some cases, these coordination challenges may substantially limit the extent of specialization, resulting in a relatively flat relationship between specialization and market size. Nevertheless, advances in knowledge increase the benefits of specialization within a field, thereby promoting greater specialization in equilibrium.

2.3 Vertical Specialization

The organization of labor within medical oncology can also be examined through the framework of vertical knowledge hierarchies, as described by Garicano (2000) and Garicano and Santos (2004). Oncologists occupy the apex of a hierarchical structure that includes oncology nurses, technicians, care coordinators, and other non-physician healthcare providers. All on-

cologists are formally licensed to manage a broad range of cancers. However, within medical oncology, informal knowledge hierarchies persist despite formal equivalence in licensure. In theory, general community oncologists are equipped to manage routine cancer cases, while more complex presentations are ideally referred to subspecialists with deeper expertise.

However, several frictions hinder the efficiency of the referral market. Anecdotal evidence suggests that healthcare organizations are often reluctant to refer patients externally, largely due to the financial incentives associated with retaining cancer care in-house. Furthermore, anti-kickback regulations prohibit compensation for referrals, meaning physicians are not reimbursed for the opportunity cost of lost revenue when referring patients to other providers. These dynamics can result in systematic under-referrral, consistent with the predictions of Garicano and Santos (2004), who argue that in a spot market with zero-priced referrals, the equilibrium level of referral falls short of the social optimum. To address this inefficiency, they propose the establishment of ex-ante agreements—often implemented through firms, partnerships, or similar organizational structures—–that formalize referral and compensation arrangements.

This type of vertical specialization is influenced by the cost of knowledge acquisition. In this setting, clinical guidelines can be conceptualized as a technology that impacts these costs. Initially, the introduction of guidelines lowers these costs by offering a structured, accessible knowledge base. However, as guidelines become increasingly detailed and complex, the cost of extracting and applying relevant knowledge also rises. These guidelines function as compressed summaries of external expertise, and their length and complexity serve as proxies for the volume of underlying knowledge required to deliver comprehensive cancer care. As the burden of knowledge acquisition increases, oncologists may be more likely to specialize further, thereby deepening vertical knowledge hierarchies within oncology (Garicano, 2000; Garicano and Santos, 2004).

2.4 Knowledge Growth and Market Size

A natural corollary to the preceding arguments is that the growth of medical knowledge may exacerbate differences in specialization across markets of varying sizes. For instance, the viability of a breast cancer specialist is contingent upon a sufficient patient population—a threshold that may only be met in large healthcare markets. As the complexity of breast cancer care increases, these specialists may further narrow their focus to specific subtypes of breast cancer, but again, only where patient volume can sustain it. In contrast, smaller markets may lack the population density necessary to support such specialized roles. As a result, general oncologists may remain the only viable providers in these settings. This disparity may contribute to a widening gap in the degree of specialization——and potentially in the quality and scope of care——between small and large markets.

3 Data and Measurement

The empirical strategy integrates multiple data sources to examine trends in knowledge growth and specialization within medical oncology over time. Knowledge growth is proxied through changes in page length and reference counts in cancer treatment guidelines, categorized by cancer type. Physician specialization is assessed using Medicare claims and provider data, which capture longitudinal changes in physician practice patterns. These datasets are combined to generate physician-level estimates of exposure to knowledge growth across a 7-year panel.

3.1 Cancer Guidelines Data

To measure knowledge growth, I constructed a novel dataset based on historical cancer treatment guidelines issued by the National Comprehensive Cancer Network (NCCN). The NCCN is a prominent not-for-profit alliance comprising 33 leading cancer centers dedicated to patient care, research, and education (NCCN, 2023*a*). Since 1996, the NCCN has published

treatment guidelines for common cancers, culminating in the release of a comprehensive "library" of guidelines in 2001 (NCCN, 2023*b*). These guidelines are widely used in clinical practice—according to one survey, 96% of oncologists reported that the NCCN Guidelines are important for making treatment decisions (McGivney, 2008).

The NCCN Guidelines offer recommendations spanning the full spectrum of cancer care, including "diagnosis, imaging, pharmacotherapy, radiation, and surgical interventions" (McGivney, 2008). These recommendations are particularly relevant to medical oncologists, who coordinate care across various specialties, such as surgical oncology, radiation oncology, and pathology. Accordingly, medical oncologists are trained to manage all aspects of cancer treatment and must stay abreast of high-level guidelines across cancer types (Popescu et al., 2014). For simplicity, the term "oncologist" in this study refers specifically to medical oncologists.

Each NCCN Guidelines reflects a consensus of expert opinion and is overseen by a committee composed of dozens of specialists from NCCN member institutions. These guidelines are updated at least annually—and often more frequently—to incorporate the latest evidence and newly approved therapies. Substantive updates require approval by a quorum of committee members. In practice, the vast majority of recommendations are supported by strong consensus among members, meaning $> 85\%$ approved the recommendation (NCCN, 2024).

To quantify knowledge growth, changes in NCCN Guideline page length and citation counts are used as proxies for the expanding volume of clinically relevant information. Page lengths and citations were measured for the initial set of NCCN Guidelines for cancer screening, treatment, and risk reduction by type and by year. Changes in page lengths serve as a proxy for the growth in clinically relevant knowledge for each cancer type. As a robustness check, changes in the number of citations of any type and to identifiable scholarly works over time are also used as alternative proxies for knowledge expansion. Identifiable scholarly works were identified in using OpenAlex, a comprehensive database of research literature

(Priem et al., 2022). These citation-based metrics reflect the evolving volume of scientific literature that underpin the guidelines.

Importantly, the volume of relevant knowledge measured in this way is distinct from the advancement of treatment technologies. For example, many pages in the guidelines are devoted to managing side effects of existing therapies. In theory, if future treatments are developed that cure all cases with minimal side effects, clinical care could become simpler, and guidelines could be shorter. In practice, however, this scenario is rare. Despite therapeutic breakthroughs, some patients still present with cancers that are incurable or progress after first-line treatments. In such cases, the goal often shifts from curing the disease to slowing its progression. Clinicians continue exploring available therapies until no effective or unused options remain. Consequently, lines of therapy accumulate over time, contributing to longer and more complex guidelines.

Guideline expansion has also been driven by the medical system's increasing capacity to subtype cancers and tailor target treatments accordingly. For instance, lymphomas—cancers of the lymphatic system—were once braodly classified as either "Hodgkin's" (named after the English physician who first described it) or the general category "non-Hodgkin's lymphoma" (Armitage, 2010). As scientific understanding has advanced, "non-Hodgkin's" lymphoma has come to encompass a growing number of biologically distinct subtypes, each requiring unique treatment approaches (Jaffe, 2009).

Appendix Figures A.1 and A.2 provide examples of how the NCCN Guidelines changed between 2009 and 2014. Appendix Figures A.1a and A.1b display the table of contents for non-Hodgkin lymphoma (NHL). Over this period, the table of contents become longer and incorporated new NHL subtypes. Figures A.2a and A.2b show a diagnostic workup flowchart for lobular carcinoma in situ, a breast cancer subtype. Between 2009 and 2014, the flowchart increased in complexity: it added a new diagnostic test, introduced more decision splits, and expanded the number of linked flowcharts from three to five.

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 information. Reassuringly, the formatting of the NCCN Guidelines remained consistent and standardized across cancer types throughout the core study period (2008–2014), supporting the interpretation that increases in page length reflect added content rather than formatting changes (e.g., font size or layout). In fact, changes in page length may slightly understate knowledge growth. Figures A.1 and A.2 show examples where a single page contains more detailed content over time.

Second, the measure assumes that any relevant information not captured in the NCCN Guidelines is either proportional to the information that is 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 consistently represent a comparable share of information across treatments. 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 primarily accumulates rather than replaces prior content. For example, if a 20-page guideline grows to 30 pages, this is interpreted as 10 pages of new knowledge. If, however, 10 pages were removed and replaced by 20 pages of new content, the net measure would understate total knowledge growth. This assumption is evaluated empirically in Section 4.

Despite its limitations, this proxy offers several advantages over alternative knowledge growth metrics. Other measures—such as the number of new journal articles, patents, or drug approvals by cancer type—may capture innovation but often include information irrelevant to everyday clinical practice. In contrast, the NCCN Guidelines are curated to reflect actionable knowledge. Moreover, not all new research replaces older information. For example, many citations from the 2002 NCCN Guidelines remain present in the 2020 edition, highlighting the enduring relevance of some medical evidence.

3.2 Medicare Claims Data

I also use a 21-year panel of Medicare claims data to observe oncologists' clinical activity. Specifically, I use a 20% sample of Medicare Part B claims from 1999 to 2019. Medicare provides federally sponsored health insurance coverage to disabled adults and most U.S. adults aged 65 and older. This age group has the highest incidence of cancer in the United States, making it an ideal population for studying cancer care (Disease Control and Prevention, 2019).

Using 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 further limit the sample to physicians who treated at least 25 unique cancer patients per year in the claims data. This threshold ensures patient privacy and that specialization measures are based on a sufficient patient volume. The influence of patient volume on the measurement of specialization and on the overall results is discussed in detail in Section 6.4 and Appendix C. For the primary regression analysis, I focus on a balanced panel of oncologists from 2008 to 2014. The panel begins in 2008, which is the first year with data on many physician characteristics, as outlined in Section 3.3. It ends in 2014, the last full year before the transition from ICD-9 to ICD-10 diagnosis codes in 2015. This coding change introduces significant discontinuities in the measurement of physician specialization, as described in Section 4.3, leading me to end the panel just prior to the coding change.

Next, I measure the types of cancer each oncologist manages. The analysis is limited to outpatient evaluation and management claim lines billed by oncologists. These visits involve high-level medical decision-making and long-term care planning, in contrast to services such as chemotherapy administration prescribed by another physician or acute inpatient management. Each evaluation and management visit is assigned to a cancer type based on the patient's primary cancer diagnosis[3]. An advantage of this context is that cancer diagnoses

[3]Cancer patients are defined as those with at least $250 (in 1999 inflation-adjusted dollars) of Medicare

are standardized and determined independently of an oncologist's specialization, eliminating concerns that the oncologist's focus might influence the type of cancer diagnosed.

I also measure the time each oncologist devotes to managing different cancer types. Each evaluation and management code is mapped to work Relative Value Units (wRVUs), which Medicare uses to estimate the physician effort required. For evaluation and management services, wRVUs reflect expected physician time to provide care.

The resulting measure captures horizontal specialization across disease categories—for example, an oncologist focusing on lung cancer versus one treating a broad range of cancer types—by share of time managing each category. This measure does not directly capture vertical specialization, such as depth of knowledge related to disease stage, biomarkers, or clinical history. However, in practice, horizontal and vertical specialization are often linked: a narrower horizontal focus typically allows for deeper vertical expertise. For example, an oncologist specializing in lung cancer is more likely to have advanced expertise in managing complex metastatic lung cancer than a generalist managing multiple cancer types.

3.3 Medicare Data on Provider Practice and Specialty

Additional physician characteristics are obtained from the Medicare Data on Provider Practice and Specialty (MD-PPAS) dataset for the period 2008–2014. This dataset provides information on each physician's birth year, which I use to calculate their age at the beginning of the panel period in 2008.

MD-PPAS also reports the top tax identification (ID) number under which each oncologist bills. This tax ID reflects the organization or individual receiving Medicare payments. Multiple physicians sharing the same primary tax ID are interpreted as belonging to the

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. 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 (80%). 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 (1.9% of patients in 2009).

same organization or group practice (ResDAC, 2024). I use the number of physicians billing under a given tax ID as a proxy for the organization size. There are limitations to this approach. 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 physicians linked to a tax ID may understate the actual size of the organization. Additionally, tax ID names are used as a proxy for academic affiliation; the details of this construction are provided in Appendix B.2. Despite its imperfections, the tax ID serves as a useful indicator of whether a physician is part of a larger organization and the nature of that organization.

3.4 Measuring Relevant Knowledge Growth for Each Oncologist

I combine the Medicare claims data with NCCN Guidelines data to 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, as determined from that year's Medicare claims. For each cancer type, define 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 calculated as $\Delta k_i = \sum_{C_i} \Delta k_c$. As a robustness check (Section 6.2, I exclude cancer types for which a physician treated only one or two patients annually, thereby requiring a minimum level of engagement for inclusion.

For example, suppose oncologist i treated only lung and breast cancer in 2008. If the lung cancer guidelines expanded by 50 pages ($\Delta k_{lung} = 50$) and the breast cancer guidelines by 100 pages ($\Delta k_{breast} = 100$) over the next 5 years, then oncologist i is exposed to a total of 150 pages of knowledge growth between 2009 and 2014.

To link cancer types in the claims data to guideline categories, I construct custom groupings of diagnosis codes aligned with the NCCN Guidelines. These groupings depart from

standard systems such as the Clinical Classification Software to better reflect the taxonomy used by the NCCN Guidelines. They also combine certain NCCN Guidelines to maintain consistency over time. For example, the NCCN Guidelines for basal and squamous cell cancer (both forms of skin cancer) were presented in a single document until 2013, after which they were issued separately. I treat these jointly under the broader category of "skin cancer guidelines." Additional details are provided in Appendix B.1.

The groupings are organized by the anatomical origin of the cancer (e.g., lung). Historically, cancer classification has been based on the location of the primary tumor, regardless of metastatic spread. For example, lung cancer that metastasizes to the brain is still categorized as lung cancer, not brain cancer. This classification strategy is embedded in the structure of NCCN Guidelines, patterns of physician specialization, and treatment protocols during the study period. Although recent years have seen the emergence of biomarker-driven, tissue-agnostic cancer therapies, the first such FDA approval occurred in 2017 (Olivier et al., 2021), which falls outside the primary panel period analyzed here.

3.5 Measuring Physician Specialization

The main outcome of interest is oncologist specialization, measured using the Herfindahl—Hirschman Index (HHI) based on the distribution of time an oncologist spends managing different cancer types. The HHI is calculated as follows: $HHI_i = 10,000 \sum_C \left(\frac{T_{ic}}{T_i}\right)^2$, where T_{ic} represents the time oncologist i spends managing cancer type c and T_i is their total time spent managing all cancers.

The HHI_i ranges from 0 to 10,000. A value of 10,000 indicates complete specialization—— that is, the oncologist treats only one cancer type. If time is split equally across two cancer types, the HHI_i is 5,000. As the number of cancer types increases and time is more evenly distributed, the HHI_i approaches 0.

This measure has several advantages. First, it is continuous and grounded in actual clinical activity, unlike job titles, which may not reflect granular differences in practice. Second,

it captures subtle 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. As such, the HHI can reflect within-physician changes over time and reveal variation across physicians with the same formal titles or specialty designations.

4 Descriptive Patterns

This section outlines the characteristics of oncologists within the sample and examines trends in the expansion of clinical knowledge and specialization over time. Both the volume of clinical knowledge and the degree of specialization in oncology have increased considerably.

4.1 Knowledge Growth Over Time

The volume of clinical knowledge, as measured by the length of treatment guidelines, has significantly expanded over time. This increase in guideline length is primarily driven by the growing body of scientific literature, which has added new content. This knowledge growth is largely additive, with limited turnover. Furthermore, the scientific literature underlying these guidelines is almost entirely distinct across cancer types.

An analysis of historical guidelines reveals a substantial increase in the length of cancer treatment guidelines since 2002. Figure 1a illustrates the lengths of these guidelines from 2000 to 2020 for selected cancer types. Table A.1 provides a comprehensive list of the page counts for various guidelines in 2002, 2009, 2014, and 2020. Over this period, the total length of guidelines has grown by more than five times, from 1,129 pages in 2002 to 5,825 pages in 2020.

One significant factor contributing to the increase in guideline length is the rapid explosion in the number of FDA-approved cancer treatments. Figure 1c depicts the cumulative number of anti-cancer drugs approved by the U.S. Food and Drug Administration (FDA)

over time[4]. From 2002 to 2020, the number of FDA-approved anti-cancer drugs grew from 90 in 2002 to 243 in 2020. Not only has the total number of drugs increased rapidly, but the rate at which new drugs are being introduced has also accelerated.

The growing body of scientific literature is another contributor to the expansion of guidelines. Figure 1b shows the number of citations in these guidelines each year. Table 1 illustrates the close relationship between guideline length and the number of sources cited within them, both cross-sectionally and over time. In 2009, each additional citation was associated with an additional one-third of a page in the guideline length ($R^2 = 88\%$). By 2014, this relationship shifted, with each additional citation contributing to an increase of one-quarter of a page ($R^2 = 95\%$), suggesting either that citations were less informative or that the content of the guidelines had become more dense. Over time, the growth in citation counts is strongly correlated with guideline length. Between 2009 and 2014, each additional citation corresponded to an increase of about one-fifth of a page in the guidelines ($R^2 = 94\%$). Guidelines during this period typically included around 15 citations per page. This implies that each citation adds approximately $\frac{1}{15}$th of a page through the citation itself and an additional $\frac{2}{15}$th of a page through expanded content within the guideline.

While knowledge growth is largely additive, there is some turnover in the references cited across guidelines. The 2009 guidelines cited 4,013 unique scholarly works, with 2,071 sources (65%) also appearing in the 2014 guidelines. However, 1,106 sources (35%) were dropped. Additionally, 5,195 unique references were added between 2009 and 2014, significantly outpacing the number of dropped sources. Figure 2 shows these changes. These patterns suggest that the expansion of clinical knowledge is primarily additive, with minimal difference between net and gross knowledge growth. Appendix Figure A.3 illustrates the yearly counts of papers added, remaining, and removed from the guidelines based on their publication year. The newly added references to the guidelines have a diversity of ages but predominantly reflect recent literature.

[4]Data on anti-cancer drugs is sourced from Pantziarka et al. (2021), utilizing Anatomical Therapeutic Chemical Level 5 classifications to define individual drugs.

The majority of references in the guidelines are to published research papers. Of the 33,621 citations in 2009 and 2014, 76% can be confidently linked to published scholarly works. The remaining 24% of citations include sources such as drug package labels, conference posters, professional society reports, government documents, and papers that could not be reliably linked.

The overlap of knowledge across different cancer types, as measured by the overlap of scientific literature cited in the guidelines, is minimal. In 2009 and 2014, 94% of cited scholarly works were referenced by only one cancer type's guidelines. This suggests that the underlying scientific literature for each cancer type's guideline is largely distinct.

A limitation of the data is revealed in Table A.1, which indicates that a fewer cancer types did not have guidelines in 2002, 2009, and 2014, despite the NCCN considering their 2002 guidelines a "complete library" of clinical guidelines. One such example is mesothelioma, a rare cancer that lacked guidelines until 2010. A 2010 news article on the release of mesothelioma guidelines noted that authors faced challenges due to limited evidence on how to treat the disease (Mulcahy, 2010). This highlights a key issue in using guideline length as a measure of knowledge: zero pages of guidelines often reflect the absence of treatment evidence for a specific cancer type. Fortunately, the empirical strategy used does not require assuming that zero-page guidelines indicate zero knowledge, only that more pages correlate with more knowledge. However, the introduction of formal guidelines may cause a discontinuous jump in guideline length, potentially overestimating the relative growth of knowledge for this type of cancer. The impact of this measurement challenge affects only a small proportion of patients in the sample (2.80 in 2009 have a cancer for which guidelines are first introduced in the study period).

4.2 Oncologist Characteristics

This section presents summary statistics for a balanced panel of oncologists observed from 2008 to 2014. Table 2 provides descriptive statistics for key variables of interest based on a

panel comprising 3,948 oncologists. In the baseline year of 2008, each oncologist treated an average of 61 unique cancer patients in the outpatient setting, with the dataset including only those who saw a minimum of 25 patients. These patients were diagnosed with an average of 14 different cancer types, though this number ranged from as few as one to as many as 25 types. The mean HHI—used to quantify each oncologist's degree of clinical specialization—was 2,913 in 2009, with some oncologists reaching the maximum possible value of 10,000, they managed only one type of cancer in the outpatient setting over the course of a year. At the start of the panel, the average oncologist was 49 years old.

Regarding practice size, oncologists worked in organizations with an average of 19 oncology colleagues. However, this ranged widely, from solo practices with only one oncologist to large organizations with as many as 219 practitioners.

In 2009, the cumulative length of the NCCN Guidelines totaled 1,832 pages. Over the subsequent 5 years, this volume doubled to 3,534 pages. When filtered to include only cancers treated by each oncologist in 2008, the number of relevant guideline pages also doubled—from an average of 1,086 pages per oncologist in 2019 to 2,118 pages by 2014.

4.3 Trends in Oncologist Specialization Over Time

This section analyzes the evolution of oncologist specialization from 1999 to 2019. Figure 3 shows the 10th, 25th, 50th, 75th, and 90th percentiles of the HHI of oncologists over time for the sample of 12,537 oncologists. Bootstrapped confidence intervals for these percentiles are derived from re-sampling oncologists. The figure also includes the HHI for the overall distribution of cancer patients in the dataset, with confidence intervals calculated by bootstrapping over the patient-level sample. Several noteworthy patterns emerge.

First, there is a marked, discontinuous decline in HHI values across the distribution between 2014 and 2016. This drop aligns temporally with the transition from ICD-9 to ICD-10 diagnostic codes, which occurred in mid-2015. The abrupt nature of this shift supports the decision to end the main balanced panel in 2014 to avoid potential confounding due to

changes in disease classification systems.

Second, contrary to expectations given the rapid expansion of clinical knowledge and treatment options, the typical oncologist has not become more specialized over time. The median (50th percentile) and lower percentiles of the HHI distribution have remained flat or declined over the study period. These oncologists, whose specialization levels are similar to the overall patient distribution, appear to function as generalists. Throughout the observed years, the 10th, 25th, and 50th percentiles remain within a narrow 1,000-point HHI range that includes the population-level HHI, indicating relatively stable patterns of generalist practice despite considerable growth in knowledge.

In contrast, oncologists at the upper end of the specialization have become significantly more specialized. From 1999 to 2014, the 90th percentile of oncologist HHI increased by 1,031 [800, 1,363] points. Between 2016 and 2019, this trend continued, with a further increase of 298 [-13, 664] HHI points.

These findings are robust to alternative measures of specialization. Figure A.4a shows specialization trends using a sample of 25 randomly selected patients per oncologist, which controls for variation in patient volume but increases measurement error. Figure A.4b presents HHI calculated using the share of unique patients by cancer type rather than time-based measures. Using patient counts reduces the influence of certain confounding factors, such as variation in visit frequency across physicians. However, it also introduces additional noise due to instances where physicians temporarily see one-time patients—for example, when covering for colleagues on leave.

Despite these limitations, the overall trends depicted in both figures remain consistent. The 90th percentile of physician specialization increased markedly between 1999 and 2014 and showed variable growth from 2016 to 2019. In contrast, the 50th percentile and lower percentiles either remained stable or declined across both periods.

These divergent trends suggest that "generalist" and "specialist" oncologists may respond differently to the expansion of medical knowledge. While specialists appear to have adapted

their practices in response to knowledge growth, generalists show little change in their level of specialization during the same period. This observation supports the hypothesis that generalists do not respond to knowledge growth in the same way as specialists.

To formally test this hypothesis, the main analysis stratifies the sample into generalist and specialist oncologists and examines whether their responses to increasing knowledge differ. It is important to note that although this classification distinguishes between "specialist" and "generalist" oncologists, all individuals in the sample are medical subspecialists. The analysis highlights variation within this already specialized group to examine the extremes of specialization.

4.4 Differences Between More and Less Specialized Oncologists

Table 3 summarizes the baseline characteristics of generalist and specialist oncologists in 2008, the first year of the balanced panel. Generalists are defined as those below the 75th percentile of the physician HHI distribution, while specialists are those at or above the 75th percentile.

Substantial differences between the two groups are evident in the baseline year. In 2008, the average HHI for specialists was 4,858—more than twice that of generalists, whose average HHI was 2,262. Specialists were also more likely to be affiliated with academic institutions, with 24% of them identified as academic physicians, compared to only 6% among generalists.

Furthermore, specialists were more frequently located in large markets: 60% practiced in the top 20% of markets by size, compared with 47% of generalists. They also tended to work in larger organizations, with an average of 21 other oncologists in their institutions, versus 17 for generalists. These patterns suggest that specialization is facilitated by access to large firms, large markets, and academic environments.

Interestingly, the two groups did not differ meaningfully in terms of age: the average age at baseline was 50 years for specialists and 49 years for generalists. Anecdotal evidence supports the notion that oncologists typically make specialization decisions early in their ca-

reers, often during fellowship training. This implies that physicians rarely transition between generalists and specialist roles later in their careers.

Despite similarities in age, the observable and likely unobservable differences between specialists and generalists indicate that they are distinct groups. These baseline differences in specialization justify the decision to estimate responses to knowledge growth separately for the two groups. Moreover, the identification strategy accounts for potential bias arising from differences in practice breadth and associated exposure to knowledge growth.

5 Identification

This section outlines the identification strategy used to assess the extent to which knowledge growth contributes to increased specialization. Specifically, changes in specialization are regressed on a measure of exposure to knowledge growth at the physician level. The measure of exposure to knowledge growth may be subject to omitted variables bias. In particular, exposure to knowledge shocks could be non-random and correlated with changes in specialization due to other factors. To address this potential bias, the exposure measure is recentered to account for expected, or non-random, exposure to knowledge growth. This adjustment follows the general approach outlined in Borusyak and Hull (2023).

5.1 Estimating Equation

The primary estimating equation is as follows. Let ΔHHI_i denote physician i's change in specialization, as measured by the HHI, between 2009 and 2014. Let Δk_i represent their exposure to knowledge growth. Exposure is calculated as: $\Delta k_i = \sum_{C_i} \Delta k_c$, where C_i is the set of cancers treated by physician i in 2008 and Δk_c is the change in page length for each cancer type between 2009 and 2014. Additionally, let X_i denote the vector of other physician characteristics, including age, market, patient volume, and organization size at the beginning of the panel.

The estimating equation is as follows:

$$\Delta HHI_i = \alpha + \beta \Delta k_i + \theta X_i + \epsilon_i,$$

where the equation is estimated using ordinary least squares with robust standard errors.

This equation incorporates several key features. First, the outcome variable uses within-physician variation in specialization over time, which necessitates a balanced panel. This within-physician design allows specialization to respond immediately to knowledge growth. However, if the specialization response is significantly lagged, the estimates may be attenuated.

The main coefficient of interest is β, which represents the effect of knowledge growth on specialization. β is estimated using variation in exposure to knowledge growth (Δk_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 treat leukemia, and some treat both. The second source of variation is that different cancer subfields experience varying levels of knowledge growth. For example, leukemia may experience more knowledge growth than breast cancer, meaning physicians who treat leukemia exclusively are more exposed to knowledge growth than those who treat breast cancer exclusively.

The intercept term α absorbs any global shocks to knowledge. Specifically, it includes the impact of knowledge growth that affects all cancer fields or the general practice of medicine. Consequently, the main coefficient of interest, β, captures the effect of changes in specialization within different cancer subfields, isolating it from global knowledge shocks.

Moreover, the intercept term α also accounts for changes in the composition of oncology patients. The HHI of the overall distribution of patients in the sample shows a modest decline from 2009 to 2014, as evidenced by the following change: -335 [-416, -252]. This level change is absorbed by α.

5.2 Recentering the Instrument to Correct for Omitted Variables Bias

The measure of exposure to knowledge growth may be susceptible to omitted variables bias. To correct for potential bias and estimate causal coefficients, the measure was recentered using the general approach outlined in Borusyak and Hull (2023).

Omitted variables bias may arise because physicians who treat more cancer types are, by construction, exposed to more knowledge growth. For instance, a physician who specializes only in breast cancer will experience less exposure to knowledge growth than one who treats both breast cancer and leukemia. Moreover, these two physicians may differ in other important characteristics. For example, the breast cancer oncologist may be more likely to practice in a large healthcare market, such as New York City, where the patient volume justifies a breast cancer-only practice. On the other hand, oncologists practicing in smaller markets, like Paducah, Kentucky, may have different specialization trends, which are unrelated to knowledge growth. This variation in market size may be correlated with both exposure to knowledge growth (Δk_i) and changes in specialization (ΔHHI_i), leading to omitted variables bias in the estimate of β.

Empirically, realized exposure (Δk_i) is directly calculated from the data. Conceptually, it can be decomposed into two components:

$$\Delta k_i = \underbrace{\mathbb{E}[\Delta k_i]}_{\text{Expected Exposure}} + \underbrace{\Delta \tilde{k}_i}_{\text{Remaining Exposure}} . \tag{1}$$

The first component, $\mathbb{E}[\Delta k_i]$, represents the expected exposure to knowledge growth and contains the omitted variables bias. The second component, $\Delta \tilde{k}_i$, is the remaining exposure to knowledge growth, which is exogenous and uncorrelated with other factors influencing specialization, apart from knowledge growth.

The realized exposure (Δk_i) can be replaced with the "recentered" or residual exposure

($\Delta\tilde{k}_i$) for an unbiased and efficient estimate of β (Borusyak & Hull, 2023), as shown below:

$$\Delta HHI_i = \alpha + \beta\Delta\tilde{k}_i + \theta X_i + \epsilon_i. \tag{2}$$

The formal identification assumption here is that recentered exposure to knowledge growth is uncorrelated with the error term, including omitted variables ($\Delta\tilde{k}_i \perp \epsilon_i$). The real-world interpretation of this assumption depends on how one formulates the expectation of exposure to knowledge growth.

To formulate the expectation, assume there exists an expected knowledge growth for each cancer type, denoted $\mathbb{E}[\Delta k_c]$. The expected exposure for a physician, $\mathbb{E}[\Delta k_i]$, can be calculated as follows:

$$\mathbb{E}[\Delta k_c] = \sum_{C_i} \mathbb{E}[\Delta k_i]. \tag{3}$$

The central challenge is to estimate the expected knowledge growth by cancer type, $\mathbb{E}[\Delta k_c]$. The next section will discuss progressively more sophisticated approaches to estimating $\mathbb{E}[\Delta k_c]$ and their implications for the identification assumptions. More sophisticated approaches reduce observable omitted variables bias more, until none is detectable. The final approach is used in the main analysis in Section 6.

The simplest assumption is that there is no expected knowledge growth in any cancer field ($\mathbb{E}[\Delta k_c] = 0$). Under this assumption, recentered exposure ($\Delta\tilde{k}_i$) equals realized exposure (Δk_i), which is implicit in the original measure of exposure to knowledge growth. However, assuming no expected knowledge growth after 2009 is arguably implausible, given the historical trends in knowledge growth prior to that year.

A somewhat more plausible assumption is that all cancer fields share the same non-zero level of expected knowledge growth. While physicians may expect knowledge to grow across all fields, they do not predict which ones will experience more growth. This assumption can be operationalized as $\mathbb{E}[\Delta k_c] = \overline{\Delta k_c}$, where $\overline{\Delta k_c}$ represents the mean realized page increase across cancer types. This estimate of $\mathbb{E}[\Delta k_c]$ can then be used to derive the recentered

exposure measure ($\Delta\tilde{k}_i$) using Equations 1 and 3.

This approach compares physicians who treat the same number of cancers and is equivalent to controlling for the number of cancers an oncologist treats (Borusyak & Hull, 2023). Variation in the recentered exposure arises because oncologists, even when treating the same number of cancers, vary in the composition of cancers they treat. Some cancers experience more innovation than others. Importantly, this approach assumes that cancer-level knowledge shocks are unpredictable. However, one concern is that cancer-level rates of knowledge growth may be predictable and thus correlated with physician specialization. For example, an oncologist may anticipate rapid advancements in breast cancer treatments and choose to specialize in that field rather than in a field with less change. Addressing this potential bias motivates further refinement of the estimate of expected knowledge growth.

Next, I allow expected knowledge growth ($\mathbb{E}[\Delta k_c]$) to vary based on cancer characteristics. Acemoglu and Linn (2004) suggests 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 enable researchers to assess more quickly and cheaply if a treatment is working than using a primary outcome—suh as mortality—leading to cheaper, faster research. I incorporate both of these factors into the estimate of expected knowledge growth, $\mathbb{E}[\Delta k_c]$.

The expected knowledge growth for each cancer type is estimated using a Poisson regression of page count increases on cancer prevalence and an indicator for high-quality surrogate endpoints. Let p_c denote the share of cancer patients in Medicare with each cancer type in 2008, and let e_c be an indicator variable indicating whether a cancer type has high-quality surrogate endpoints, as defined by Budish, Roin, and Williams (2015). The Poisson maximum likelihood regression model is as follows:

$$E[ln\Delta k_c] = \delta_0 + \delta_1 e_c + \delta_2 p_c + \delta_3 e_c p_c. \tag{4}$$

This regression controls for knowledge shocks common to all cancer types (captured by δ_0).

Table 4 shows the coefficient estimates for the regression. The independent variables substantially predict knowledge growth, explaining 77% of the variation in the growth of knowledge. The remaining variation is attributed to randomness and uncertainty in research and development. For instance, some clinical trials and drug development lines fail, while others succeed.

After estimating the expected knowledge growth for each cancer type ($E[\Delta k_c]$), I use this information to calculate each physician's recentered exposure to knowledge growth ($\Delta\tilde{k}_i$), removing variation attributable to both the physician's scope and cancer characteristics.

This approach relies on several key assumptions. First, it assumes that patient prevalence, secondary endpoints, and shared knowledge shocks are the main factors driving predictable variation in innovation across cancer types. This assumption is supported by the finding that these factors explain a substantial share (77%) of the variation in guideline page count increases. Additionally, the method assumes that the residual variation in knowledge growth is unrelated to physicians' specialization decisions, except through its direct effect on specialization. While this assumption is untestable, it is consistent with the inherent unpredictability of clinical research. If clinical research outcomes were entirely predictable, the success rate of investigational drugs would likely be higher than the estimated 14% (Schuhmacher et al., 2025).

5.3 Correlation of Exposure to Knowledge Growth with Omitted Variables

This section empirically evaluates the extent to which the recentering procedure mitigates bias in the estimation of exposure to knowledge growth arising from omitted variables, following the methodology proposed by Borusyak and Hull (2023). The original (unadjusted) measure of knowledge growth exhibits a strong correlation with several observable omitted variables, indicating that recentering is necessary to obtain unbiased estimates. Encourag-

ingly, as increasingly refined estimates of expected knowledge growth ($\mathbb{E}[\Delta k_c]$) are used for recentering, the correlation between exposure to knowledge growth and omitted variables progressively declines——eventually becoming statistically insignificant.

Column 1 in Table 5 shows the coefficient estimates from regressions of observed omitted variables on the unadjusted measure of knowledge growth. These omitted variables include key baseline physician characteristics: market size (measured as the population within the physician's hospital referral region [HRR]), firm size (approximated by the number of other oncologists sharing the same tax ID number), number of unique patients seen, and physician age. As expected, the unadjusted knowledge growth measure is significantly correlated with all included omitted variables. For instance, an increase of 1 million in HRR population is associated with a statistically significant decrease in exposure to knowledge growth by 15 pages ($SE = 2$). Consequently, regressions using the unadjusted measure may inadvertently compare physicians practicing in markedly different market environments, thereby introducing bias into estimate of β due to unrelated specialization dynamics. Collectively, the omitted variables account for 33.5% of the variation in the unadjusted exposure measure, underscoring the severity of omitted variable bias.

Columns 2–4 of Table 5 sequentially eliminate additional sources of variation from the exposure measure. Specifically, Column 2 removes variation attributable to physician scope, Column 3 further adjusts for cancer prevalence, and Column 4 incorporates adjustments for surrogate endpoint variation. With each stage of recentering, the explanatory power of the observed omitted variables declines. By Column 4, the final recentered measure exhibits no statistically significant correlation with patient volume, market size, organization size, or physician age. For example, a 1 million increase in HRR populaton is now associated with a negligible and statistically insignificant change of -0.3 pages ($SE = 0.5$) in exposure to knowledge growth. Overall, the proportion of variation in exposure explained by observed omitted variables drops from 33.5%to just 1%—a reduction of more than 99%.

Tables A.2 and A.3 present the same regressions disaggregated by specialist and generalist

oncologists, respectively. The results are consistent across both subgroups: the recentering procedure reduces the explanatory power of observed omitted variables by 96% for specialists and over 99% for generalists.

In conclusion, the final recentering specification effectively eliminates the correlation between exposure to knowledge growth and observed omitted variables. These findings suggest that the procedure may also attenuates correlations with unobserved omitted variables, thereby enhancing the credibility of subsequent analyses. The fully recentered measure of exposure to knowledge growth is employed in all analyses presented in Section 6.

6 Effect of Knowledge Growth on Specialization

Knowledge and specialization in oncology have evolved substantially over time. This study investigates whether these trends are interconnected. The findings indicate that oncologists who were already specialized ("ex-ante specialists") tend to deepen their specialization as relevant medical knowledge expands. This intensified focus primarily occurs along the intensive margin—that is, within the subset of cancer types they already manage. In contrast, oncologists who were initially generalists exhibit a more limited response to knowledge growth. Their changes are modest and occur mostly on the extensive margin, reflecting a slight reduction in the breadth of cancer types they treat. The evidence suggests that generalists selectively discontinue managing cancer types that constitute a minimal share of their caseload, resulting in negligible changes in their overall HHI scores—a metric used here to quantify specialization.

6.1 Effects on Overall Specialization

Figure 4a presents the univariate relationship between exposure to knowledge growth—using the measure defined in Section 5—and changes in specialization among ex-ante specialist oncologists. The results reveal a statistically significant increase in specialization in response to

knowledge expansion. Specifically, every 100-page increase in guideline content is associated with a 196-point (s.e. = 80) increase in HHI[5]. Relative to the baseline mean, this represents a 4% increase in HHI in response to a 13% increase in guideline content. When fixed effects for oncologist age, hospital referral region, patient volume, and organization size are incorporated, the estimated increase in HHI rises slightly to 246 points (s.e. = 91) per 100-page increase (Table 6).

In contrast to specialists, general oncologists do not consistently respond to the growth of medical knowledge by increasing their degree of specialization. Figure 4b presents the univariate relationship between exposure to knowledge growth and changes in specialization, focusing exclusively on oncologists classified ex ante as "generalists." The estimated effect on the HHI is more precisely measured in this group than among specialists but remains statistically indistinguishable from 0 at $\beta = 50$ (31)[6]. Table 7 presents regression estimates incorporating fixed effects. When all fixed effects are included, the coefficient increases modestly and attains marginal statistical significance at the 95% confidence level ($\beta = 66$ (33)). These findings indicate only a minimal specialization response among generalists to the expansion of clinical guidelines.

Figure A.5 and Table A.4 provide univariate and fixed-effect regression estimates for the full sample of oncologists. Across all model specifications, the estimated coefficients exhibit small positive effects, and the inclusion of fixed effects does not substantively alter these results.

6.2 Effects on Different Margins of Specialization

Among specialists, increases in specialization appear to be driven primarily by adjustments along the intensive margin. In contrast, the absence of a significant specialization response

[5]Mean reversion in HHI is observed: HHI tends to decrease for oncologists in the top 25% of the distribution. More precisely, the reduction in HHI is attenuated by 196 points for every additional 100 pages of guideline exposure.

[6]Mean reversion in HHI is also observed within this sample. Although the HHI increases for oncologists in the bottom 75% of the distribution, there is no differential increase associated with higher exposure to guideline growth.

among generalists masks a modest reduction in scope along the extensive margin following exposure to knowledge growth.

Table A.6 shows that, for generalists, increased guideline exposure is associated with a small reduction in the number of unique cancer types treated. Specifically, estimates suggest a decrease of approximately 0.2 (0.1) cancer types. Across all model specifications, the data rule out that exposure to 100 additional guideline pages reduces the number of cancer types treated by 0.5 or more (on a mean of 14) at the 95% confidence level. These results suggest that general oncologists may cease treating cancers that occupy only a small fraction of their practice, thereby producing negligible changes in their overall HHI.

By contrast, Table A.5 indicates that specialists do not reduce the number of cancer types they treat in response to knowledge growth. Thus, observed increases in HHI among specialists must stem from adjustments along the intensive margin. In practical terms, specialists appear to concentrate more heavily on certain cancer types within their domain of expertise without discontinuing care for others.

Table A.7 provides pooled regression estimates for all oncologists. The findings from this analysis are broadly consistent with those derived separately for generalists and specialists, indicating that the average effect reflects a weighted combination of both groups' responses.

The observed patterns among specialists may reflect a tendency to continue treating established patients while narrowing the scope of new patient intake to specific cancer types. To test this hypothesis, an alternative specialization measure was calculated using only new patient evaluation and management codes (Tables A.8 and A.9). However, restricting the analysis to new patients significantly reduced the sample size to approximately one-quarter of its original size—even after lowering the patient count threshold to 11)—resulting in substantially larger standard errors and rendering the estimates statistically inconclusive.

Another plausible explanation is that oncologists respond to knowledge growth selectively, focusing only on cancer types that constitute a significant portion of their clinical practice. For example, a physician who frequently treats breast cancer iay keep up to date with new

guidelines for that cancer type while disregarding guidelines for rarer cases in their caseload. In the primary analysis, a physician was considered "exposed" to knowledge growth fro a given cancer type if they had treated at least one patient with that cancer in the baseline year. To eliminate the influence of one-off cases, this threshold was raised to two and then three patients in Table A.8. The resulting estimates remain consistent, showing similar effect sizes: $\beta = 186$ (87) for the two-patient threshold $\beta = 311$ (110) for the three-patient threshold. These findings underscore the robustness of the observed relationship between knowledge growth and increased specialization among specialists, even under more stringent exposure definitions. Comparable analyses for generalist oncologists, presented in Table A.9, reveal no statistically significant response to knowledge growth. The estimated coefficients are $\beta = 7$ (32) for the two-patient threshold and $\beta = 9$ (35) for the three-patient threshold, further confirming that generalists exhibit minimal changes in specialization regardless of how exposure is defined.

6.3 Robustness to Alternative Measures of Knowledge Growth

To assess the robustness of the main regression results, I also employ an alternative measure of knowledge growth: the number of sources cited in each clinical guideline. This metric serves as a proxy for the volume of scientific literature underpinning the guidelines. While the magnitude of the results differs—which is expected, given that each cited source represents approximately one-fifth the informational volume of a guideline page—the substantive conclusions remain consistent. An increase in cited sources is associated with a statistically significant rise in specialization among oncologists (Table A.13). For specialists, an additional 100 citations in the guidelines is estimated to increase specialization by 65 (18) HHI points in the univariate regression. For generalists, the effect is smaller but remains statistically significant, with an estimated increase of 16 (7) HHI points per 100 additional citations (Table A.14). At the aggregate level, the overall effect of increased citations on specialization remains positive statistically robust (Table A.15).

As an additional test, I also employ an alternative recentered measure of knowledge growth that recenters based on the estimated 2008 incidence rather than prevalence of each cancer type in Medicare (Appendix Table A.16 for specialists and Appendix Table A.17 for generalists). The results are extremely similar to main specifications, in part because cancer prevalence and incidence are highly correlated in Medicare ($R^2 = 89\%$).

6.4 Robustness to Volume-Dependent Measurement Error

Lastly, I examine the potential impact of volume-dependent measurement error in the specialization metric ($\hat{HHI}_i$) on the main findings. A more detailed discussion is provided in Appendix C, while key insights are summarized here.

The analysis is based on a 20% national sample of Medicare Part B enrollees, rather than the full patient population. This sampling may introduce measurement error into both the outcome variable (specialization, $\hat{HHI}_i$) and the treatment variable (recentered exposure to knowledge growth, $\Delta\tilde{k}_i$), especially for physicians with smaller patient volumes. Crucially, the treatment variable ($\Delta\tilde{k}_i$) is uncorrelated with patient volume, as demonstrated in Table 5, suggesting that it is unaffected by volume-dependent measurement error. By contrast, the specialization variable ($\hat{HHI}_i$) does exhibit volume-dependent measurement error, as illustrated in Figure A.8. Nonetheless, the primary regression estimates remain virtually unchanged when controlling for patient volume via fixed effects (Tables 6 and 7), indicating that this type of measurement error does not introduce systematic bias.

As an additional check, I replicate the key descriptive statistics and regression results using an alternative method: calculating both $\hat{HHI}_i$ and $\Delta\tilde{k}_i$ based on a random sample of 25 patients per physician. This approach standardizes the level of measurement error across physicians by artificially inflating it for high-volume providers. With the added noise, the estimated effects of knowledge growth on specialization become too imprecise to draw inference (Tables A.10, A.11, and A.12). However, the time trends in specialization (Figure A.4a) still reflect the upward trend in top-level specialization observed in Figure 3.

7 Constraints to Specialization

Why do specialist and generalist oncologists respond differently to knowledge growth? Empirical patterns suggest that specialist and generalist oncologists face distinct constraints on their ability to specialize. In particular, extensive specialization appears to be facilitated by scale: extreme specialization is primarily observed in large markets and within large academic organizations. By contrast, scale does not appear to enable specialization for the majority of oncologists. The median oncologist remains similarly general, regardless of whether they practice in large or small markets. Furthermore, most generalists are employed in non-academic settings, where organizational size does not appear to support specialization. Neither median nor top specialization increases systematically with organization size. These observations imply that the majority of oncologists face a combination of low private returns to specialization or high coordination costs, which collectively limit the extent to which they can narrow their clinical focus.

7.1 Market Size and Coordination Costs

Highly specialized oncologists appear to be constrained by the size of the markets in which they practice. Figure 5c illustrates the 90th percentile of physician specialization (referred to as "top specialization") by market size for the years 1999 and 2014. Table 8 reports the corresponding regression results. In 1999, top specialization increased by 495 (79) HHI points for every doubling in market size, a result consistent with the classical notion that specialization is limited by the extent of the market, as first articulated by Smith (1776). By 2014, this relationship had intensified, with top specialization rising by over 1,060 (94) HHI points per doubling in market size.

This growing gradient aligns with the framework discussed in Section 2.4. As medical knowledge expands, the optimal level of specialization increases—but only in markets with enough potential patients to support such niche expertise. These results suggest that health-

care delivery in highly specialized fields is characterized by increasingly large economies of scale over time.

The implications for patient access are significant. Growth in specialization over time has been concentrated in large markets, enhancing local access to subspecialist care in those areas. In contrast, access in smaller markets has either remained static or declined on average. If subspecialized oncologists offer superior care and patient mobility remains limited—as suggested by Dingel et al. (2023)——this divergence may exacerbate geographic disparities in health outcomes.

In contrast to top specialization, median oncologist has remained largely invariant to market size. Figure 5 and Table 8 show that in 1999, median specialization was effectively flat across the distribution of market sizes ($\beta = -24$ (45)). That is, the typical oncologist in a large market was no more specialized than one in a small market, despite the former having access to a larger patient base capable of supporting narrower scopes of practice. This pattern suggests that market size is not a binding constraint on specialization for the typical oncologist in 1999, as median specialization did not increase despite the availability of larger patient pools in bigger markets. Instead, the evidence is more consistent with coordination costs acting as the primary barrier to specialization for general oncologists, consistent with the theoretical framework of Becker and Murphy (1992) discussed in Section 2.2.

Figure 5 and Table 8 indicate that by 2014, median specialization had increased slightly with market size ($\beta = 130$ (34)). However, this gradient remains modest compared with that observed for top specialization. Even in the presence of substantial knowledge growth between 1999 and 2014, the typical oncologist in large markets continued to maintain a broad clinical scope. Interpreted through the lens of Becker and Murphy (1992), this pattern suggests that coordination costs remained high relative to the returns to specialization, thereby constraining general oncologists from narrowing their clinical focus.

Figure A.6 presents the gradients of both median and top specialization with respect to market size for the years 2016 and 2019. However, the short time interval between these

years limits the ability to identify significant shifts in these patterns.

7.2 Firms and Firm Types

A key distinction between generalists and specialists lies in the types of organizations they work for. As shown in Table 3, specialists are disproportionately employed in academic institutions. Large academic organizations tend to support high levels of specialization, whereas non-academic organizations do not, irrespective of size. One interpretation of this pattern is that academic institutions are better at facilitating coordination among oncologists, thereby enabling specialization—an advantage not observed in non-academic settings, even among the largest firms.

Academic organizations appear to drive the highest levels of specialization. Figure 6 illustrates the 50th and 90th percentiles of specialization within academic organizations by firm size in 2014. The 90th percentile is calculated for organizations with at least five oncologists. Among academic institutions, both median and top specialization increase markedly with organization size. This pattern suggests that academic settings successfully mitigate some coordination frictions, thereby enabling oncologists to subspecialize. Once coordination frictions are sufficiently low, organizational scale becomes a key enabler of specialization: larger academic institutions support higher degrees of specialization across both the middle and upper portions of the distribution.

In contrast, non-academic organizations exhibit signs of persistently high internal coordination frictions. Figure 6 also presents the 50th and 90th percentiles of specialization by firm size for non-academic institutions in 2014. In these settings, both median and top specialization are flat across the distribution of firm size. The most specialized oncologists at large non-academic firms are no more specialized than those at small firms. Likewise, the typical oncologist in the largest non-academic institutions is no more specialized than a solo practitioner. These patterns suggest that non-academic organizations do not reduce coordination frictions and therefore do not support increased specialization—neither at the

median nor at the top of the distribution.

Appendix Figure A.7 replicates this analysis for 2008, the first year with available data on organizational size. The results are consistent, with no discernible changes between 2008 and 2014. Table 9 and Appendix Table A.19 provide the corresponding regression tables.

These findings may also reflect dynamics consistent with an under-referral equilibrium, as described by Garicano and Santos (2004). In this setting, providers may know that a different doctor may be a better fit for a patient, but sending the patient to someone else gives up profits. Academic institutions appear to mitigate the referral inefficiencies that may otherwise hinder specialization, while non-academic institutions do not.

An alternative explanation for these patterns is that academic oncologists face higher returns to specialization rather than lower coordination frictions. Tenure and promotion systems in academic settings often reward narrow clinical expertise, incentivizing physicians to subspecialize.

Conversely, non-academic oncologists may face lower private returns to specialization. These physicians are often face strong incentives to maximize patient volume, a goal that may be more easily achieved through a broad scope of practice.

In addition, price incentives for producing high-quality care remain weak. 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. When physicians are not financially rewarded for delivering higher-quality, more specialized care, under-specialization may result. This dynamic is likely more pronounced in non-academic settings, where providers may fail to internalize the full benefits of specialization, particularly those tied to quality.

Finally, selection effects may play a role. Physicians with a strong preference for deep expertise, academic prestige, or scholarly work may self-select into academic institutions. In contrast, those who prioritize income or prefer clinical variety may choose non-academic

practice settings.

8 Technology Use and Specialization

Oncologist specialization varies substantially by market size and organization type. This raises an important question: Are there observable returns to specialization in terms of the quality of patient care? Unfortunately, the dataset lacks sufficient clinical detail about each patient's cancer—such as stage, histology, or molecular biomarkers—which precludes a direct analysis of whether specialization improves patient outcomes as the omitted variables bias is likely too great to permit plausible comparison of outcomes.

Nevertheless, the data do allow for an examination of oncologists' prescribing behavior with respect to anti-cancer drugs, which are central to treatment and significantly influence clinical outcomes. Anti-cancer drugs are inherently challenging to administer due to their complex dosing protocols and significant toxicity profiles. Among physicians, medical oncologists are uniquely specialized in managing these high-risk and carefully dosed therapies. Consequently, prescribing patterns for anti-cancer agents represent a core aspect of oncologists' clinical practice and technological engagement.

The findings indicate that specialization enhances the quality of care. Compared with general oncologists, specialized oncologists prescribe a narrower range of anti-cancer drugs and demonstrate greater familiarity with each medication they use. They also use newer therapies, including those designated as "Breakthrough Therapy" drugs, underscoring their responsiveness to pharmaceutical innovation. Notably, these differences in prescribing behavior have widened over time, suggesting increasing realized returns to specialization in oncology.

8.1 Data and Measurement

This analysis estimates key characteristics of the typical set of anti-cancer drugs prescribed annually by individual oncologists. Anti-cancer agents are identified using Anatomical Therapeutic Chemical (ATC) classification codes assigned by the World Health Organization, as referenced in Pantziarka et al. (2021). In cases where a drug is associated with multiple ATC codes[7], — typically reflecting its diverse clinical applications—— only one ATC code is selected for inclusion in this analysis. Prescription data are drawn from a 20% sample of Medicare Part B and Part D claims, with National Provider Identifiers (NPIs) used to track prescribing physicians. The dataset begins in 2008, the first year in which NPIs were broadly implemented in Part D claims.

The criteria for oncologist inclusion in this analysis differ slightly from those applied in Section 4. In addition to requiring a minimum of 25 total patients, oncologists must have prescribed anti-cancer medications to at least 11 unique patients within the 20% Medicare sample across either Part B or Part D. This threshold ensures sufficient prescribing volume for meaningful analysis and maintains compliance with privacy standards. As a result, the number of eligible oncologists is smaller than in the baseline sample. To offset this reduction and enhance longitudinal coverage, the analysis utilizes a full (unbalanced) panel of oncologists rather than a balanced panel. This approach allows for examination of trends over an extended period (2008–2019) while maximizing the sample size in each year. Given that the sample composition changes annually, oncologist specialization is measured using the HHI for each individual year.

8.2 Scope of Technology Use

Specialization allows oncologists to narrow the range of technologies they employ in clinical practice. Figure 7 illustrates the association between oncologists' level of specialization and the number of distinct anti-cancer drugs they prescribe annually. In the initial year of the

[7]Some drugs are assigned multiple ATC codes when they are indicated for distinct medical uses.

panel, differences in prescribing patterns are evident but modest. Oncologists in the highest quartile of specialization (75th—99th percentile of the HHI) prescribe an average of 16.1 (0.2) unique anti-cancer drugs per year. In contrast, general oncologists (first–50th percentile of HHI) prescribe an average of 18.3 (0.1) drugs—approximately 13% more.

Over time, the disparity in prescribing behavior becomes more pronounced. By 2019, highly specialized oncologists reduce their average scope to 11.9 (0.3) unique drugs per year. Generalist oncologists, however, continue to prescribe a broader range, averaging 20.6 (0.2) unique drugs—73% more than their specialized counterparts. These findings reinforce the notion that specialization narrows the scope of therapeutic technologies, with growing impacts on prescribing behavior over time.

8.3 Experience

A narrower therapeutic scope allows specialized oncologists to accumulate greater experience with the specific drugs they prescribe. In this analysis, an oncologist's experience with a given drug is measured by the number of patients to whom the drug is prescribed within a calendar year. This metric captures only recent experience, reflecting evidence that physician-specific human capital deteriorates quickly over time (Facchini, 2022; Hockenberry & Helmchen, 2014). To facilitate comparison, experience is normalized across physicians within each drug-year, resulting in a relative experience measure with a mean of 1.

Figure 8 shows the evolution of average relative drug experience by level of oncologist specialization. In 2008, the most specialized oncologists exhibits slightly less experience—approximately 8% lower—than their generalist counterparts. Over time, however, specialists accumulate more experience relative to generalists. By 2010, the pattern reverses, and by 2019, the most specialized oncologists exhibit 24% higher average relative experience with the drugs they prescribe. Given the complexity and toxicity risks associated with many anti-cancer therapies, increased familiarity with specific drugs is likely to enhance clinical management and reduce the incidence of adverse effects.

8.4 Drug Age and Quality

Specialization also appears to enhance oncologists' ability to adopt new technologies, 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[8]. Drug quality is proxied by whether the drug's active molecule ever received a "Breakthrough Therapy" designation from the U.S. Food and Drug Administration as of 2025. This designation, which is based on preliminary clinical evidence of substantial improvement over existing therapies, is positively associated with measures of clinical benefit and thus serves as a credible indicator of drug quality (Feng and Maini, 2016; Sun et al., 2020). Drug-level measures are patient-weighted to reflect the prescribing volume of each oncologist.

Figure 9a plots the average age of prescribed drug bundles by oncologist specialization over time. In 2008, drug age is comparable between specialist and generalist oncologists. However, the average age of drugs prescribed increases more steeply for generalists over time. By 2019, the average drug bundle prescribed by generalist oncologists is 3.0 years older than that of specialists. These results suggest that specialization facilitates more rapid adoption of newly approved therapies.

Figure 9b shows that specialists also prescribe a higher share of breakthrough-designated drugs than generalists. While the magnitude of this gap varies across years, the direction is consistent: specialized oncologists are persistently more likely to incorporate these high-quality, innovative therapies into their practice.

In summary, specialist oncologists exhibit increasingly distinct prescribing patterns relative to generalists. They use a narrower set of drugs, gain more experience with those drugs, and more quickly adopt newer and higher-quality therapies. These patterns suggest that specialists may deliver higher-quality care, and that these quality differences are widening over

[8]This approach likely overestimates drug age compared with methods that account for new indications. However, the data do not allow identification of prescribing indication or off-label use.

time. The findings are also consistent with the notion that specialization enables clinicians to stay closer to the technological frontier in a rapidly evolving therapeutic landscape.

Do these differences in prescribing behavior translate into higher rates of guideline-concordant care or better patient outcomes? Unfortunately, the available data lack detailed clinical information—such as cancer stage or biomarker status—needed to directly assess this. However, prior research provides suggestive evidence (Thiels et al., 2020; Voss et al., 2017). For example, Voss et al. (2017) find that soft-tissue sarcoma patients treated with guideline-concordant care experience better survival and that treatment at academic hospital centers——where specialists are concentrated——is associated with higher rates of such care. Similarly, Thiels et al. (2020) report improved survival and greater adherence to guidelines among stomach cancer patients treated in integrated academic systems. Taken together, these studies support the interpretation that specialists, by virtue of their narrower focus and greater drug-specific experience, may deliver superior care and improve patient outcomes.

9 Conclusion

This paper examines how the expansion of general medical knowledge influences expert specialization and shapes subsequent patterns of technology adoption in medical oncology. It documents a rapid and substantial increase in the number of anti-cancer drugs and the body of clinically relevant information over time. Despite this surge in knowledge, most oncologists maintain broad, generalist practices and do not become more specialized. In contrast, a small subset of initially specialized oncologists respond to knowledge growth by deepening their specialization.

These differences in specialization have important implications for patient care. Compared with generalists, specialized oncologists prescribe newer anti-cancer therapies and demonstrate greater experience with the drugs they use. These prescribing differences, which

have become more pronounced over time, are consistent with specialization facilitating the delivery of higher-quality care.

Specialization trends also have geographic consequences. In large healthcare markets, top-level specialization has grown significantly while it has stagnated—or even declined—in smaller markets. This divergence suggests that the growth of medical knowledge enhances economies of scale in care delivery and contributes to increasing geographic disparities in access to specialized medical care.

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.

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)." Accessed on 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.

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.

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."

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

Duffin, Jacalyn. 2019. *Stanley's Dream: The Medical Expedition to Easter Island.* Vol. 247, McGill-Queen's Press-MQUP.

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.

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

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.

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.

Jones, Benjamin F. 2009. "The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder?" *Review of Economic Studies*, 35.

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 and Drug Drug Benefits*, 1(5): 40–44. Interview with Bill McGivney, PhD.

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

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

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

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

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–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.

Popescu, RA, Robert Schäfer, Raffaele Califano, Robert Eckert, Robert Coleman, J-Y Douillard, Andrés Cervantes, Paolo G Casali, Cristiana Sessa, Eric Van Cutsem, et al. 2014. "The current and future role of the medical oncologist in the professional care for cancer patients: a position paper by the European Society for Medical Oncology (ESMO)." *Annals of Oncology*, 25(1): 9–15.

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.

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.

Sun, Mengzhao, Evan Kirkby, Jennifer Liu, Sonali Maniar, Karen Lin, Frederick Wong, Rohit Sarin, Daniel T. Chang, Prajnan Das, Jyothi Jagannathan, Anastassia Kirilova, Masha Kocherginsky, James D. Murphy, Parag Parikh, Thomas J. George, and Eugene J. Koay. 2020. "Chemoradiation in older adults with pancreatic cancer: A population-based

evaluation of comparative effectiveness and safety outcomes in the modern era." *Cancer*, 126(23): 5112–5123.

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.

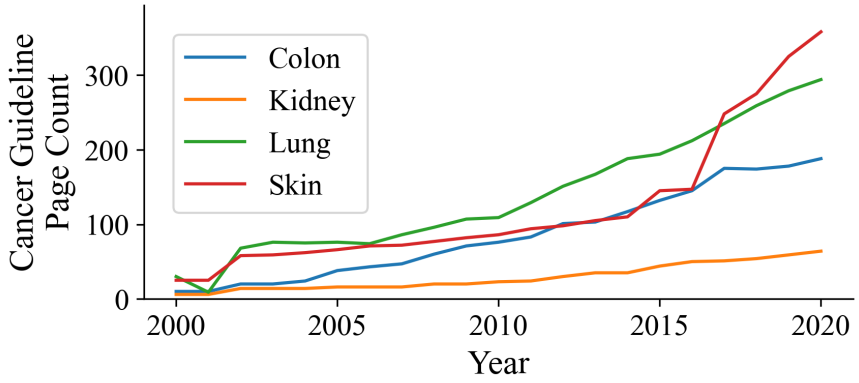
Voss, Rachel K, Yi-Ju Chiang, Keila E Torres, B Ashleigh Guadagnolo, Gary N Mann, Barry W Feig, Janice N Cormier, and Christina L Roland. 2017. "Adherence to National Comprehensive Cancer Network Guidelines is Associated with Improved Survival for Patients with Stage 2A and Stages 2B and 3 Extremity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

Wallis, Claudia. 2023. "A Wave of New Cancer Treatments Challenges Community Oncologists to Keep Up." *Scientific American*. Accessed: 2024-06-07.

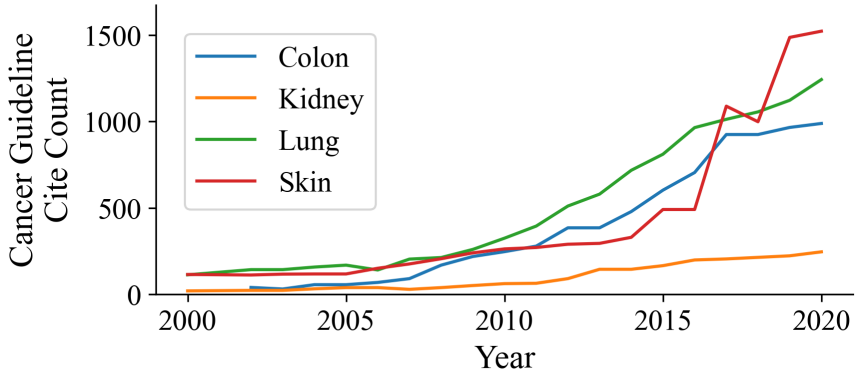
Welch, W Pete, and Andrew B Bindman. 2016. "Town and Gown Differences Among the 100 Largest Medical Groups in the United States." *Academic Medicine*, 91(7): 1007–1014.

Figure 1: Oncology has experienced large increases in available knowledge and technologies

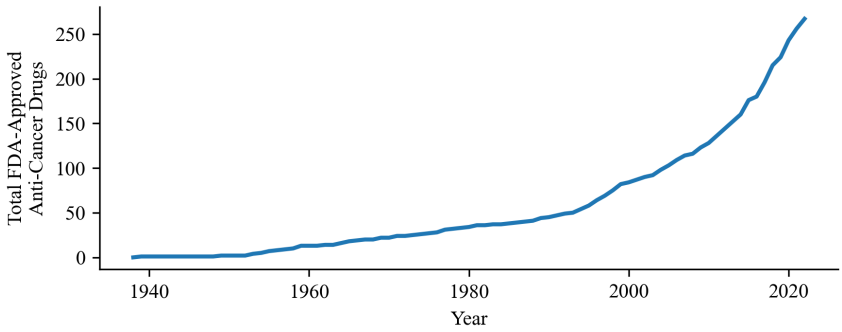
(a) Oncology clinical practice guidelines have become considerably longer



(b) Oncology clinical practice guidelines include many more citations



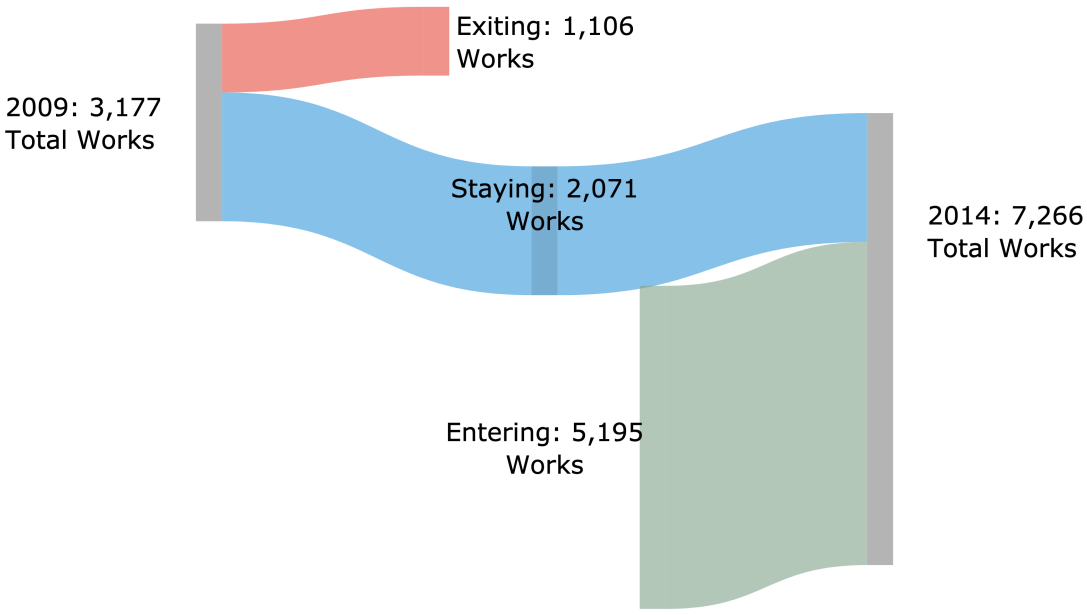
(c) The number of FDA-approved anti-cancer drugs has also increased



Notes: Panels (a), (b) and (c) show the growth in cancer guideline length, guideline citations, and approved pharmaceutical treatments 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 NCCN Guidelines. Panel (c) displays the number of unique anti-cancer drugs, where drugs are distinguished by unique Anatomical Therapeutic Chemical Level 5 Classification Codes. Note that the sample is limited to drugs with direct anti-cancer uses which are approved by the U.S. Food and Drug Administration. It does not include combinations of drugs, drugs used for cancer symptom relief or diagnosis, or unapproved drugs that are being investigated. The data in panel (c) is from Pantziarka et al. (2021).

Figure 2: Citation counts entering, exiting, and staying in 2009 and 2014 guidelines

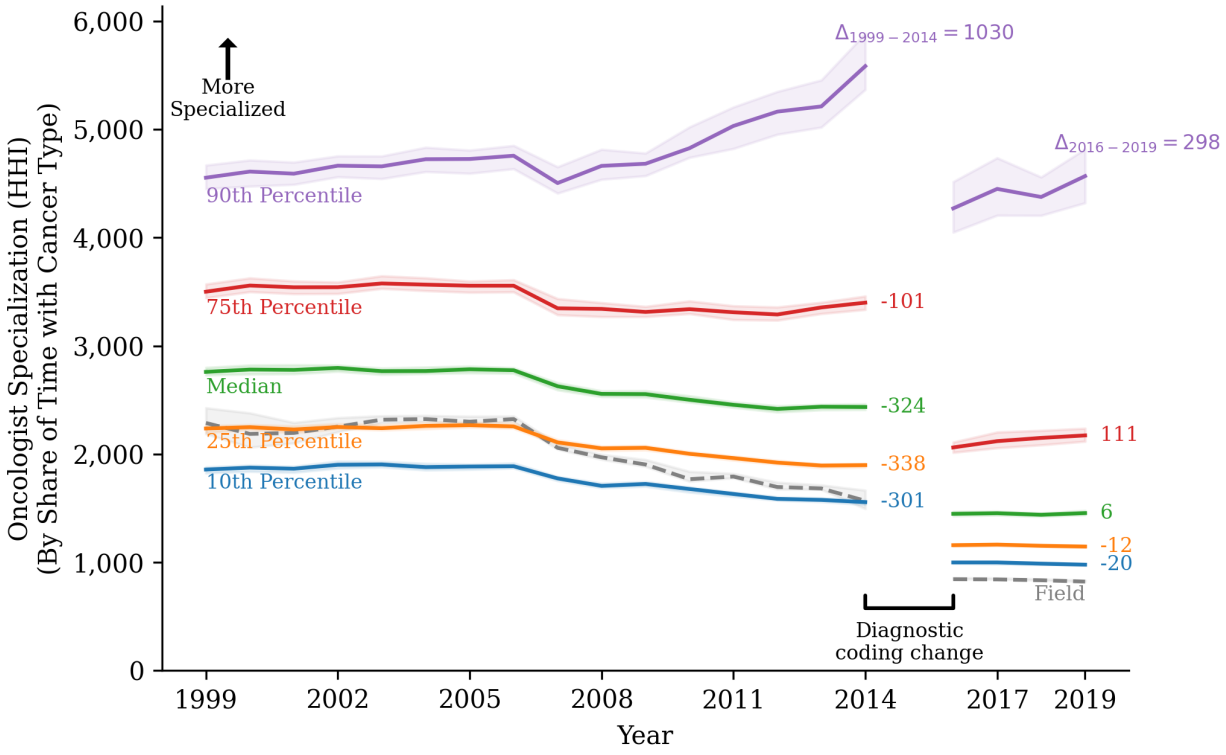
(a)



Notes: Figure A.3 shows changes in the references cited by NCCN Clinical Practice Guidelines in Oncology for treatment, screening, and risk reduction of cancer in each year between the 2009 and 2014 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 3: Extremes of specialization are increasing in oncology over time

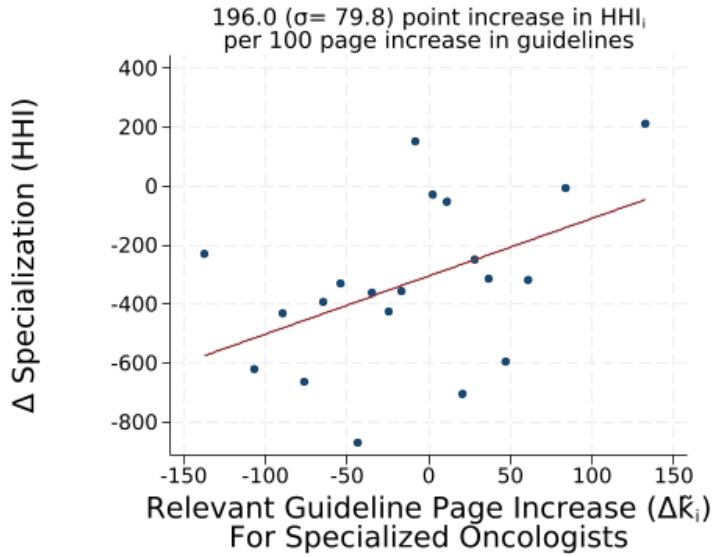
(a) Distribution of oncologist specialization over time



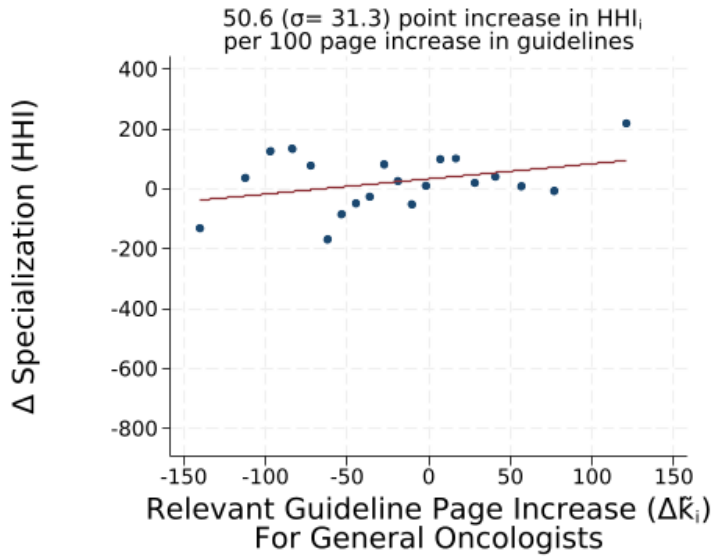
Notes: This figure shows the distribution of physician-level Herfindahl-Hirschman Index (HHI), a measure of specialization, from 1999-2019. All physicians in the sample are hematologist-oncologists, medical oncologists, or hematologists who saw at least 25 unique cancer patients in the year in a 20% sample of Medicare Part B outpatient evaluation and management claims. An HHI is calculated for each physician in each year using shares of work Relative Value Units, a proxy for time, spent evaluating and managing different types of cancer in the outpatient setting. The 10th, 25th, 50th, 75th, and 90th percentile are displayed and labeled. 95% confidence intervals for each percentile are calculated by bootstrapping physicians. The dashed grey line represents the HHI of patients in the sample. 95% confidence intervals are calculated by bootstrapping patients. The trend break in 2015 is attributed to the switch in diagnosis coding schemes, from ICD-9 to ICD-10, and 2015 is dropped from the graph for visual clarity.

Figure 4: Exposure to knowledge growth increases specialization only for oncologists who start specialized (2009-2014)

(a) Change in specialization (HHI) vs. exposure to knowledge growth for specialist oncologists (top 25% of initial specialization distribution)



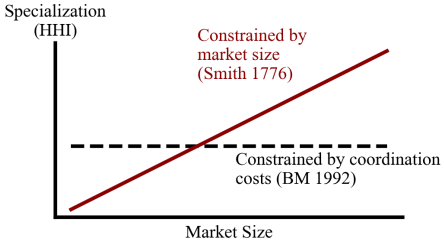
(b) Change in specialization (HHI) vs. exposure to knowledge growth for generalist oncologists (lower 75% of initial specialization distribution)



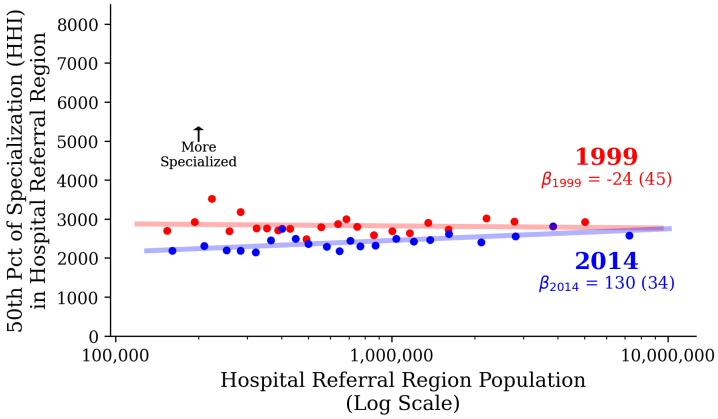
Notes: Panels (a) and (b) show binned scatterplots and regression estimates for the relationship between the recentered exposure to knowledge growth and change in specialization. Specialization is measured using a Herfindahl—-Hirschman Index (HHI) over types of cancer managed. Panel (a) shows this relationship for oncologists who start in the top 25% of the initial specialization distribution. Panel (b) shows this relationship for oncologists who start in the lower 75% of the initial specialization distribution. In the recentered measure, predicted exposure is subtracted from realized exposure to knowledge growth. Predicted exposure is estimated based on the cancer types a physician treats in 2008 and the predicted growth in knowledge over the next five years. The predicted knowledge growth by cancer type is estimated based on a Poisson regression of the log of number of Medicare patients with the cancer type in 2008, an indicator for well-established surrogate endpoints, and the interaction of the two. Regression standard errors are robust. 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.

Figure 5: Top specialization rises steeply with market size while median specialization is relatively flat, consistent with different contraints specialization

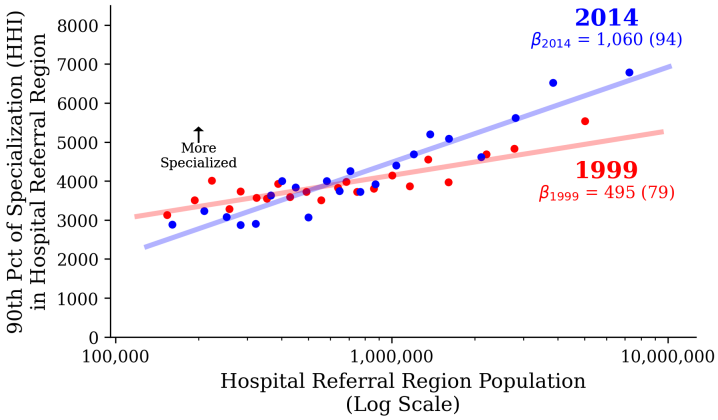
(a) Relationship between market size and specialization



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



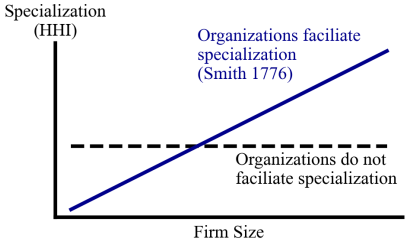
(c) 90th percentile of specialization by market size over time markets (1999 and 2014)



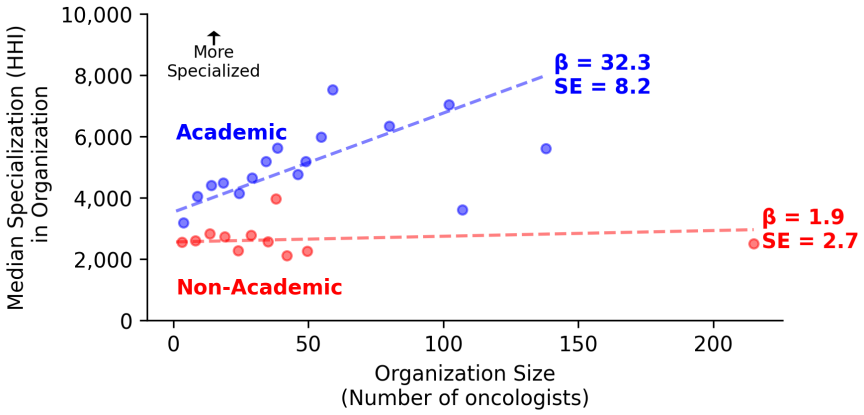
Notes: Panel (a) shows the predicted relationship between specialization and market size if specialization is constrained by market size (Smith, 1776) versus coordination costs (Becker and Murphy, 1992). Panels (b) and (c) show binned scatterplots of the 50th and 90th percentile of oncologist specialization 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 (HRR). In panel (b), the y-axis shows the 50th percentile of specialization in market in a year. In panel (c), the y-axis shows the 90th percentile of specialization in 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 that year. The line of best fit between log population and the relevant percentile of market size is plotted. 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: Academic organizations facilitate specialization while non-academic ones do not

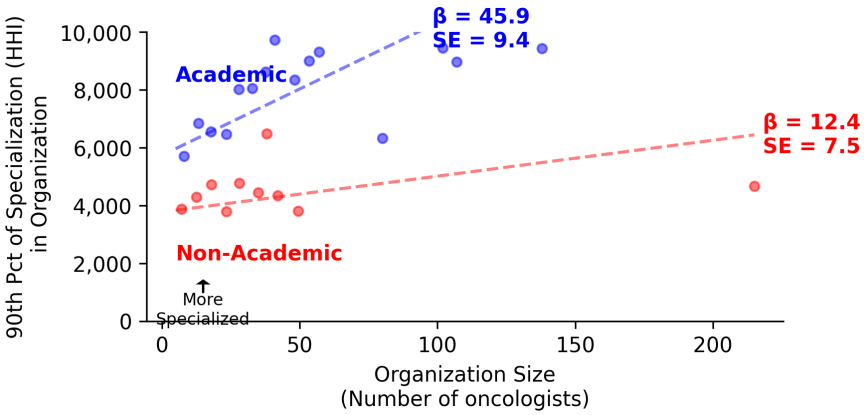
(a) Relationship between organization size and specialization



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



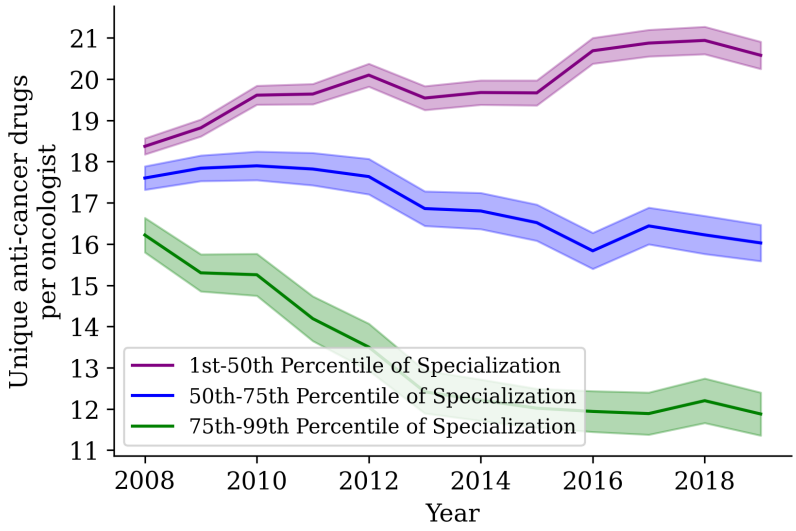
(c) 90th percentile of specialization by organization size and type (2014)



Notes: Panel (a) shows the predicted relationship between specialization and organization size if organizations do or do not facilitate specialization. Panels (b) and (c) show binned scatterplots of the 50th and 90th percentile of oncologist specialization in an organization by size and type in 2014. Organization size is on the x-axis and is estimated as the number of oncologists with at least 25 patients at the organization's tax ID in the 2014 Medicare Data on Provider Practice and Specialty (MD-PPAS), using oncologists' primary tax ID. In panel (b), the y-axis shows the 50th percentile of specialization in an organization. In panel (c), the y-axis shows the 90th percentile of specialization in an organization. In this panel, 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 an oncologist evaluated and managed in a year. The best fit line is plotted for the relationship between size and the relevant percentile of specialization in an organization. Points and lines are plotted separately for academic and non-academic organizations. Organization type is determined using tax ID names from the MD-PPAS. See Section 3.3 and Appendix B.2 for discussion of the limitations of variables from the MD-PPAS. Plotted points represent the mean value for organization size bins of width 5 and do not contain an equal number of organizations. 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 7: Over time, specialists prescribe fewer unique anti-cancer drugs than generalists in a year

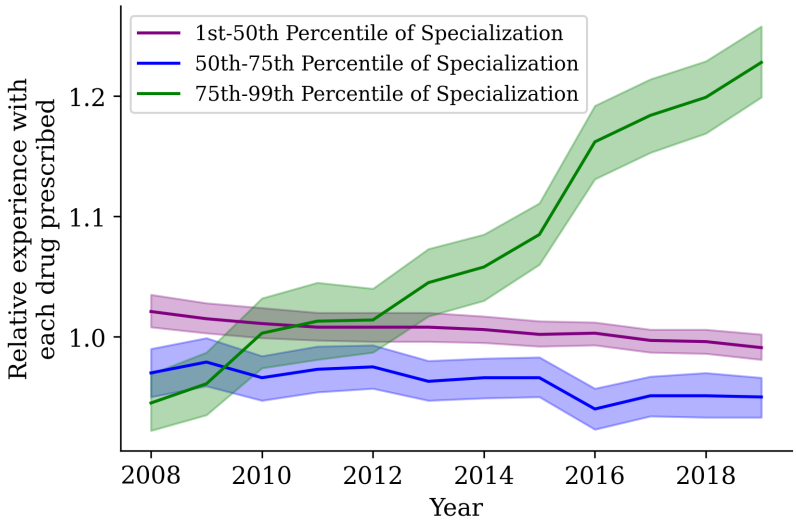
(a) Count of unique anti-cancer drugs prescribed by year and degree of specialization



Notes: This figure shows the average number of unique anti-cancer drugs that oncologists prescribe over time by level of specialization. In the early years of the panel, specialized and generalist oncologists prescribe a similar number of unique drugs. However, by the end of the panel, the more specialized an oncologist is, the fewer unique drugs they prescribe in a year. Unique anti-cancer drugs are defined as unique World Health Organization Anatomical Therapeutic Class codes that fall into the list of anti-cancer drugs from Pantziarka et al. (2021). Oncologists are grouped into specialization bins based on the percentile of their HHI, where a higher HHI indicates a more specialized oncologist. Within each specialization bin, the sample is restricted to oncologists who prescribed anti-cancer drugs to at least 11 patients across a 20% sample of Part B and Part D data. HHI is calculated using shares of evaluation and management work Relative Value Units over types of cancer, and the percentile is calculated among the population of oncologists with at least 25 unique cancer patients in a 20% sample of Medicare Part B outpatient evaluation and management claims.

Figure 8: Over time, specialists become relatively more experienced with the drugs they prescribe than generalists

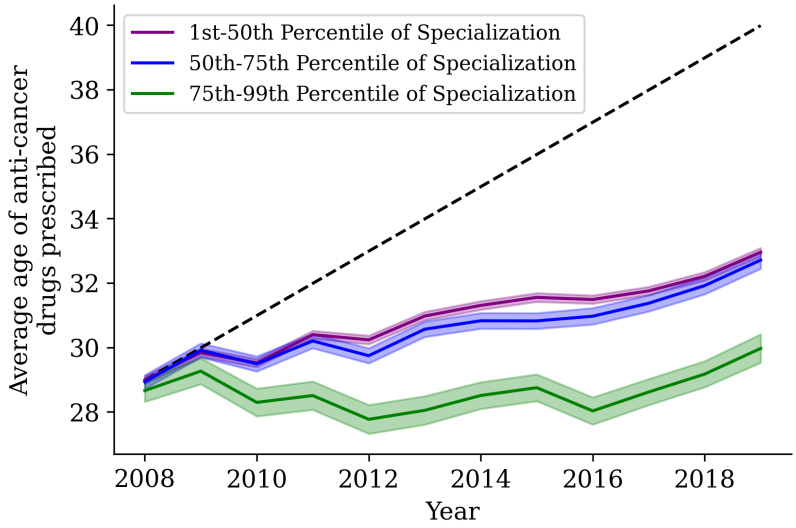
(a) Relative experience with anti-cancer drugs prescribed by year and degree of specialization



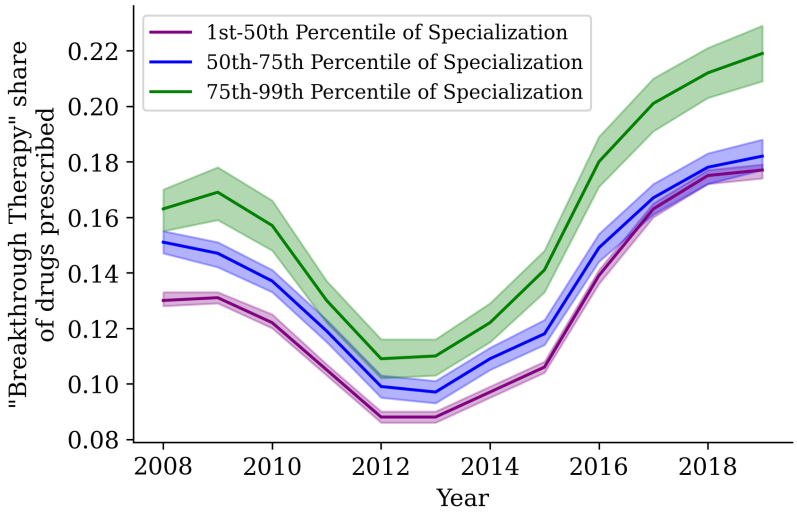
Notes: This figure shows oncologists' average relative experience with the anti-cancer drugs that they prescribe over time by level of specialization. Experience with a drug is calculated as the number of patients an oncologist prescribed a drug to in a year. Relative experience normalizes experience to have a mean of one within a drug-year. Each drug-specific relative experience is weighted equally within a physician. Physicians are weighted equally to estimate mean relative experience by degree of specialization. Weighting physician-level average relative experience equally by physician causes physician-drug-level relative experience to be weighted unequally, so physician-level average relative experience does not have a mean of one. Anti-cancer drugs are defined as unique World Health Organization Anatomical Therapeutic Class (ATC) codes that fall into the list of anti-cancer drugs from Pantziarka et al. (2021). Oncologists are grouped into quantile bins based on the percentile of their HHI, where a higher HHI indicates a more specialized oncologist. Within each quantile bin, the sample is restricted to oncologists who prescribed anti-cancer drugs to at least 11 patients across a 20% sample of Part B and Part D data. HHI is calculated using shares of evaluation and management work Relative Value Units over types of cancer, and the percentile is calculated among the population of oncologists with at least 25 unique cancer patients in a 20% sample of Medicare Part B outpatient evaluation and management claims.

Figure 9: Over time, specialists begin to prescribe more new and breakthrough drugs than generalists

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



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



Notes: Panel (a) shows the average age of anti-cancer drugs that oncologists prescribe over time by level of specialization. Panel (b) shows the average share of breakthrough anti-cancer drugs that oncologists prescribe over time by level of specialization. Anti-cancer drugs are defined as unique World Health Organization Anatomical Therapeutic Class (ATC) codes that fall into the list of anti-cancer drugs from Pantziarka et al. (2021). Drug age is measured as the time since FDA approval of the ATC code. A drug is classified as a breakthrough drug if the molecule, as of 2025, has ever received a "Breakthrough Therapy" designation from the U.S. Food and Drug Administration for any indication. Drugs are weighted by the number of patients the oncologist prescribed the drug to. Oncologists are grouped into quantile bins based on the percentile of their HHI, where a higher HHI indicates a more specialized oncologist. HHI is calculated using shares of evaluation and management work Relative Value Units over types of cancer, and the percentile is calculated among the population of oncologists with at least 25 unique cancer patients in a 20% sample of Medicare Part B outpatient evaluation and management claims. Within each quantile bin, the sample is restricted to oncologists who prescribed anti-cancer drugs to at least 11 patients across a 20% sample of Part B and Part D data.

Table 1: Guideline length is closely associated with the number of sources cited in guidelines

	(1) Page Count (2009)	(2) Page Count (2014)	(3) 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 1 shows the relationships between length of guidelines, measured in pages, and the number of sources cited in cross-section and over time. Column 1 contains a cross-sectional regression on guideline length on number of citations in 2009. Column 2 contains the same regression for 2014. Column 3 regresses changes in guideline length on changes in citation counts between 2009 and 2014. All page counts and citation counts are for the first set of guidelines in each year from NCCN Guidelines. In all regressions, citation counts are closely coupled to page lengths. Guidelines are grouped using the cancer grouping discussed in section 3.4 and appendix section B.1 to maintain a stable sample of guidelines over time that match with diagnosis code groupings.

Table 2: Summary statistics for balanced panel of oncologists (2008-2014)

	Mean	SD	Min	Max	N
N	3,948				
Unique Patients (2008)	61	29	25	291	3,948
Count of Cancer Types (2008)	14	4	1	25	3,948
HHI (2009)	2,920	1,517	1,028	10,000	3,948
HHI (2014)	2,855	1,660	897	10,000	3,948
Guideline Length (2009)	1,789				
Relevant Guideline Length (2009)	1,086	233	42	1,664	3,948
Change in Length of Guidelines (2009-2014)	1,745				
Change in Length of Relevant Guidelines (2009-2014)	1,032	227	24	1,606	3,948
Count Oncologists at Main Tax ID (2008)	18	40	1	219	3,948
Physician age (2008)	49	9	31	77	3,948

Notes: This table shows summary statistics for a balanced panel of oncologists from 2008 to 2014. 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. More information about each variable is calculated is detailed in section 3.

Table 3: Specialized oncologists are more concentrated in larger regions, larger organizations, and academics than general oncologists (2008)

	Specialists (Top 25% HHI)		Generalists (Lower 75% HHI)	
	Mean	Std	Mean	Std
N	985.0		2,953.0	
Unique Patients	53.5	(24.6)	64.1	(29.9)
Count of Cancer Types	11.1	(4.2)	14.7	(3.0)
HHI (2008)	4,865.8	(1,866.1)	2,269.0	(484.6)
HHI (2009)	4,440.9	(2,160.1)	2,408.3	(680.4)
HHI (2014)	4,113.4	(2,452.6)	2,433.1	(974.3)
Academic	0.24	(0.43)	0.06	(0.23)
Share in Largest 20% of Regions	0.60	(0.49)	0.47	(0.50)
Count Oncologists at Main Tax ID	21.0	(38.1)	17.0	(39.7)
Physician Age	50.4	(8.6)	49.1	(8.6)
Relevant Guideline Length (2008)	792.5	(263.1)	997.5	(156.4)
Change in Length of Relevant Guidelines (2009-2014)	869.5	(289.1)	1,087.2	(169.9)

Notes: This table compares key summary statistics for "specialist" and "general" oncologists in a balanced panel from 2008 to 2014. Specialist oncologists are defined as those in the top 25% of the specialization distribution as measured by a Herfindahl—-Hirschman Index (HHI) over types of cancer managed in 2008. Specialist oncologists are defined as those in the bottom 75% of the specialization distribution in 2008. All characteristics are measured in 2008 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. More information about how each variable is detailed in section 3.

Table 4: Guidelines grow faster for cancer types with more patients and surrogate endpoints

	(1) Page Count Increase	(2) Page Count Increase	(3) Page Count Increase
Log Patient Count in Medicare (2008)	0.312	0.332	0.268
	(0.0736)	(0.0632)	(0.0498)
Surrogate Endpoints		1.010	-3.149
		(0.322)	(1.413)
Surrogate x Log Pat Count (2008)			0.436
			(0.135)
Observations	35	35	35
R-squared	0.216	0.588	0.771
Estimation Method	PPML	PPML	PPML

Notes: This table shows the results of a regression of the increase in guideline pages from 2009 to 2014 on the log of the prevalence of the cancer type in Medicare in 2008, an indicator if the cancer type has surrogate endpoints, and the interaction of the two. The unit of observation is a cancer type e.g. breast cancer. The regression is estimated with Poisson maximum likelihood. Change in guideline page count is calculated using the change in page length for the first version of NCCN Guidelines in 2009 and 2014. The presence of surrogate endpoints is determined using Budish, Roin, and Williams (2015). 2008 prevalence in Medicare is calculated as the number of patients with the cancer type a 20% sample of 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.

Table 5: Physician characteristics no longer correlate with exposure to knowledge growth after recentering procedure

	(1) Unadjusted exposure to knowledge growth (Δk_i)	(2) Remove physician scope variation ($\Delta \tilde{k}_i$)	(3) Remove cancer prevalence variation ($\Delta \tilde{k}_i$)	(4) Remove endpoint type variation ($\Delta \tilde{k}_i$)
Market population (millions)	-15.19	-4.145	-1.467	0.316
	(2.001)	(0.772)	(0.710)	(0.541)
Count oncologists at main org	-0.382	-0.108	0.0303	0.0278
	(0.0900)	(0.0346)	(0.0332)	(0.0258)
Physician age	-1.281	0.210	0.207	-0.103
	(0.349)	(0.150)	(0.153)	(0.120)
Patient Count	4.304	0.241	-0.276	-0.0464
	(0.118)	(0.0413)	(0.0418)	(0.0350)
Observations	3,948	3,948	3,948	3,948
R-squared	0.335	0.024	0.010	0.001
Recentered		Yes	Yes	Yes

Notes: This table shows the results of a regression of exposure to knowledge growth on observable physician characteristics. The physician characteristics include the population of the physicians' Hospital Referral Region in millions, the number of other oncologists at the primary organization that they work at, and physicians' patient volume. All physician characteristics are from 2008, the baseline year for the panel. In Column 1, the outcome variable is raw exposure to knowledge growth, measured as the total increase in guideline page counts for cancer types from 2009 to 2014 for cancer types the physician managed in 2008. In Columns 2-4, the outcomes are increasingly recentered measures of exposure to knowledge growth. In recentering, predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. In Column 2, predicted knowledge growth is the same for each cancer and estimated as mean knowledge growth per cancer type. In Column 3, the predicted knowledge growth for each type of cancer is estimated using on a Poisson regression of the number of Medicare patients with the cancer types in 2008. Column 4 is similar, but predicted knowledge growth also adds indicator for if the cancer type has well-established surrogate endpoints and the interaction between the number of Medicare patients and surrogate endpoints. The full recentering procedure removes variation in exposure to knowledge growth that arises become some physicians have a broader scope and because some cancer types predictably have more innovation. 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. All regression standard errors are robust.

Table 6: Knowledge growth increases specialization for oncologists in the top 25% of the initial specialization distribution

	Dependent Variable: HHI					
	(1)	(2)	(3)	(4)	(5)	(6)
	Knowledge Growth Recentered ($\Delta\tilde{k}_i$)	Restrict Sample	Add Age FE	Add Market FE	Add Pat Count FE	Add Org Size FE
Increase in Guideline Length (100 Pages)	213.2	213.2	220.0	253.1	255.2	246.3
	(81.06)	(81.06)	(81.56)	(89.07)	(89.98)	(91.20)
Observations	924	924	924	924	924	924
R-squared	0.008	0.008	0.009	0.172	0.177	0.178
Age FE			Yes	Yes	Yes	Yes
Market FE				Yes	Yes	Yes
Org Size FE					Yes	Yes
Pat Count FE						Yes

Notes: This table shows regression estimates for the relationship between exposure to knowledge growth and change in specialization for oncologists who start in the top 25% of the initial specialization distribution. Specialization is measured using a Herfindahl—-Hirschman Index (HHI) over types of cancer managed. The measure of exposure to knowledge growth is recentered to control for variation in physician scope and field expected knowledge growth. Column 1 shows the coefficient estimate for the univariate regression. The regression standard errors are robust. Column 2 restricts the sample to physicians where we observe age, hospital referral region (HRR) and the number of other oncologists at their primary organization. Columns 3, 4, and 5 add fixed effects for the physicians age group, hospital referral region (HRR) and the number of other oncologists at their primary organization, respectively. In the recentered measure of exposure to knowledge growth, predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. Raw 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. Predicted exposure is estimated based on the cancer types a physician treats in 2008 and the predicted, rather than realized, growth in knowledge over the next five years. The predicted knowledge growth for each type of cancer is estimated based on a Poisson regression of the number of Medicare patients with the cancer types in 2008, an indicator for if the cancer type has well-established surrogate endpoints, and the interaction of the two. 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: Knowledge growth may marginally increase specialization for oncologists in the bottom 75% of the initial specialization distribution

	Dependent Variable: HHI					
	(1)	(2)	(3)	(4)	(5)	(6)
	Knowledge Growth Recentered ($\Delta\tilde{k}_i$)	Restrict Sample	Add Age FE	Add Market FE	Add Pat Count FE	Add Org Size FE
Increase in Guideline Length (100 Pages)	54.23	54.23	54.18	66.96	66.50	66.26
	(31.45)	(31.45)	(31.44)	(33.14)	(33.13)	(33.03)
Observations	2,937	2,937	2,937	2,937	2,937	2,937
R-squared	0.001	0.001	0.001	0.108	0.108	0.110
Age FE			Yes	Yes	Yes	Yes
Market FE				Yes	Yes	Yes
Org Size FE					Yes	Yes
Pat Count FE						Yes

Notes: This table shows regression estimates for the relationship between exposure to knowledge growth and change in specialization for oncologists who start in the bottom 75% of the initial specialization distribution. Specialization is measured using a Herfindahl—-Hirschman Index (HHI) over types of cancer managed. The measure of exposure to knowledge growth is recentered to control for variation in physician scope and field expected knowledge growth. Column 1 shows the coefficient estimate for the univariate regression. The regression standard errors are robust. Column 2 restricts the sample to physicians where we observe age, hospital referral region (HRR) and the number of other oncologists at their primary organization. Columns 3, 4, and 5 add fixed effects for the physicians age group, hospital referral region (HRR) and the number of other oncologists at their primary organization, respectively. In the recentered measure of exposure to knowledge growth, predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. Raw 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. Predicted exposure is estimated based on the cancer types a physician treats in 2008 and the predicted, rather than realized, growth in knowledge over the next five years. The predicted knowledge growth for each type of cancer is estimated based on a Poisson regression of the number of Medicare patients with the cancer types in 2008, an indicator for if the cancer type has well-established surrogate endpoints, and the interaction of the two. 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 8: Median and 90th percentile of specialization by market size in 1999 and 2014

	(1) Median HHI	(2) 90th Pct HHI
Current Population (Log)	14.30	584.5
	(30.08)	(56.41)
Current Population (Log) x 2014	152.2	575.8
	(39.55)	(88.94)
Observations	1,205	1,205
R-squared	0.144	0.279

Notes: Table 8 shows a regression of oncologist specialization on market size, year, and an interaction of market size and year. The baseline year is 1999 with regression coefficients for 2014. In Column 1, the outcome is the 50th percentile 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 (HRR). Market size is measured as population in the relevant year. Oncologist specialization is measured using a Herfindahl-Hirschman Index (HHI) over the types of cancer that the oncologist evaluated and managed in that year. 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 data.

Table 9: Median and 90th percentile of specialization by organization size in 2014

	(1) Median HHI Academic	(2) 90th Pct Academic	(3) Median HHI Non-academic	(4) 90th Pct HHI Non-academic
Organization Size	32.30	45.89	1.864	12.43
	(8.229)	(9.380)	(2.722)	(7.490)
Observations	159	119	1,466	345
R-squared	0.104	0.158	0.000	0.010
Minimum Org Size	1	5	1	5
Standard Errors	Robust	Robust	Robust	Robust

Notes: Table 9 shows regressions of oncologist specialization on organization size by organization type in 2014. In Column 1, the outcome is the 50th percentile 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 other oncologists in the sample. In Column 3, the outcome is the 50th percentile 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 other oncologists in the sample. Organization size is estimated as the number of oncologists with at least 25 patients 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. See Section 3.3 and Appendix B.2 for discussion of the limitations of variables from the MD-PPAS. Oncologist specialization is measured using a Herfindahl-Hirschman Index (HHI) over the types of cancer that the oncologist evaluated and managed in that year. 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 data.

10 Appendix

For Online Publication

Knowledge Growth and Specialization: Evidence from Oncologists

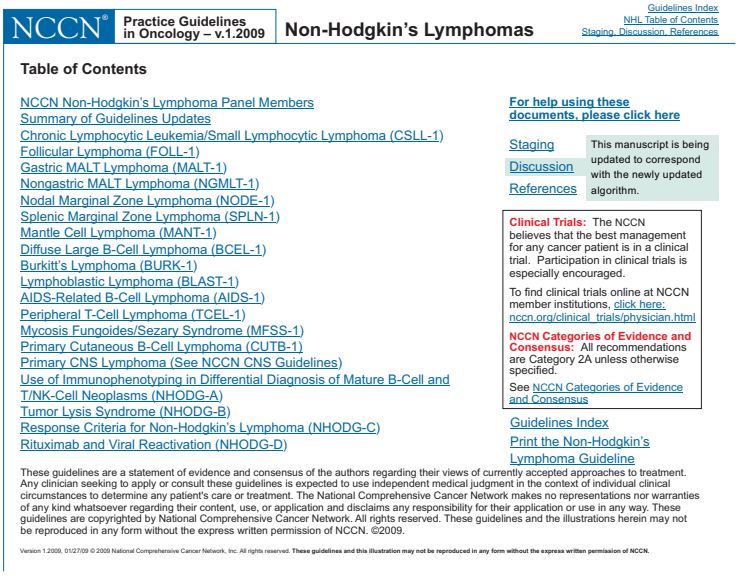
Maya Lozinski

August 2025

A Additional exhibits

Figure A.1: Changes in Table of Contents for Non-Hodgkins Lymphoma Over Time

(a) 2009 Table of Contents



NCCN® Practice Guidelines in Oncology – v.1.2009 **Non-Hodgkin's Lymphomas**

Guidelines Index
NHL Table of Contents
Staging, Discussion, References

Table of Contents

NCCN Non-Hodgkin's Lymphoma Panel Members
Summary of Guidelines Updates
Chronic Lymphocytic Leukemia/Small Lymphocytic Lymphoma (CSLL-1)
Follicular Lymphoma (FOLL-1)
Gastric MALT Lymphoma (MALT-1)
Nongastric MALT Lymphoma (NGMLT-1)
Nodal Marginal Zone Lymphoma (NODE-1)
Splenic Marginal Zone Lymphoma (SPLN-1)
Mantle Cell Lymphoma (MANT-1)
Diffuse Large B-Cell Lymphoma (BCEL-1)
Burkitt's Lymphoma (BURK-1)
Lymphoblastic Lymphoma (BLAST-1)
AIDS-Related B-Cell Lymphoma (AIDS-1)
Peripheral T-Cell Lymphoma (TCEL-1)
Mycosis Fungoides/Sezary Syndrome (MFSS-1)
Primary Cutaneous B-Cell Lymphoma (CUTB-1)
Primary CNS Lymphoma (See NCCN CNS Guidelines)
Use of Immunophenotyping in Differential Diagnosis of Mature B-Cell and T/NK-Cell Neoplasms (NHODG-A)
Tumor Lysis Syndrome (NHODG-B)
Response Criteria for Non-Hodgkin's Lymphoma (NHODG-C)
Rituximab and Viral Reactivation (NHODG-D)

For help using these documents, please click here

Staging
Discussion
References

This manuscript is being updated to correspond with the newly updated algorithm.

Clinical Trials: The NCCN believes that the best management for any cancer patient is in a clinical trial. Participation in clinical trials is especially encouraged.

To find clinical trials online at NCCN member institutions, click here: nccn.org/clinical_trials/physician.html

NCCN Categories of Evidence and Consensus: All recommendations are Category 2A unless otherwise specified.

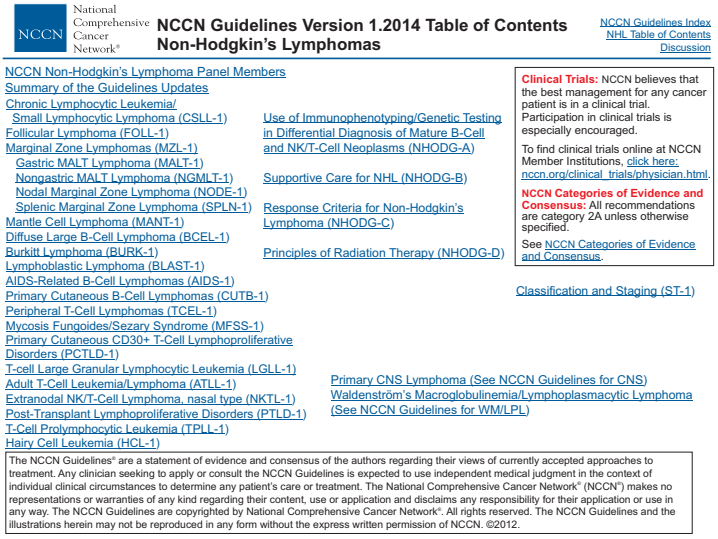
See NCCN Categories of Evidence and Consensus

Guidelines Index
Print the Non-Hodgkin's Lymphoma Guideline

These guidelines are a statement of evidence and consensus of the authors regarding their views of currently accepted approaches to treatment. Any clinician seeking to apply or consult these guidelines is expected to use independent medical judgment in the context of individual clinical circumstances to determine any patient's care or treatment. The National Comprehensive Cancer Network makes no representations nor warranties of any kind whatsoever regarding their content, use, or application and disclaims any responsibility for their application or use in any way. These guidelines are copyrighted by National Comprehensive Cancer Network. All rights reserved. These guidelines and the illustrations herein may not be reproduced in any form without the express written permission of NCCN. ©2009.

Version 1.2009, 01/27/09 © 2009 National Comprehensive Cancer Network, Inc. All rights reserved. **These guidelines and this illustration may not be reproduced in any form without the express written permission of NCCN.**

(b) 2014 Table of Contents



NCCN National Comprehensive Cancer Network®

NCCN Guidelines Version 1.2014 Table of Contents
Non-Hodgkin's Lymphomas

NCCN Guidelines Index
NHL Table of Contents
Discussion

NCCN Non-Hodgkin's Lymphoma Panel Members
Summary of the Guidelines Updates
Chronic Lymphocytic Leukemia/
Small Lymphocytic Lymphoma (CSLL-1)
Follicular Lymphoma (FOLL-1)
Marginal Zone Lymphomas (MZL-1)
Gastric MALT Lymphoma (MALT-1)
Nongastric MALT Lymphoma (NGMLT-1)
Nodal Marginal Zone Lymphoma (NODE-1)
Splenic Marginal Zone Lymphoma (SPLN-1)
Mantle Cell Lymphoma (MANT-1)
Diffuse Large B-Cell Lymphoma (BCEL-1)
Burkitt Lymphoma (BURK-1)
Lymphoblastic Lymphoma (BLAST-1)
AIDS-Related B-Cell Lymphomas (AIDS-1)
Primary Cutaneous B-Cell Lymphomas (CUTB-1)
Peripheral T-Cell Lymphomas (TCEL-1)
Mycosis Fungoides/Sezary Syndrome (MFSS-1)
Primary Cutaneous CD30+ T-Cell Lymphoproliferative Disorders (PCTLD-1)
T-cell Large Granular Lymphocytic Leukemia (LGLL-1)
Adult T-Cell Leukemia/Lymphoma (ATLL-1)
Extranodal NK/T-Cell Lymphoma, nasal type (NKTL-1)
Post-Transplant Lymphoproliferative Disorders (PTLD-1)
T-Cell Prolymphocytic Leukemia (TPLL-1)
Hairy Cell Leukemia (HCL-1)

Use of Immunophenotyping/Genetic Testing in Differential Diagnosis of Mature B-Cell and NK/T-Cell Neoplasms (NHODG-A)
Supportive Care for NHL (NHODG-B)
Response Criteria for Non-Hodgkin's Lymphoma (NHODG-C)
Principles of Radiation Therapy (NHODG-D)

Primary CNS Lymphoma (See NCCN Guidelines for CNS)
Waldenström's Macroglobulinemia/Lymphoplasmacytic Lymphoma (See NCCN Guidelines for WM/LPL)

Clinical Trials: NCCN believes that the best management for any cancer patient is in a clinical trial. Participation in clinical trials is especially encouraged.

To find clinical trials online at NCCN Member Institutions, click here: nccn.org/clinical_trials/physician.html.

NCCN Categories of Evidence and Consensus: All recommendations are category 2A unless otherwise specified.

See NCCN Categories of Evidence and Consensus.

Classification and Staging (ST-1)

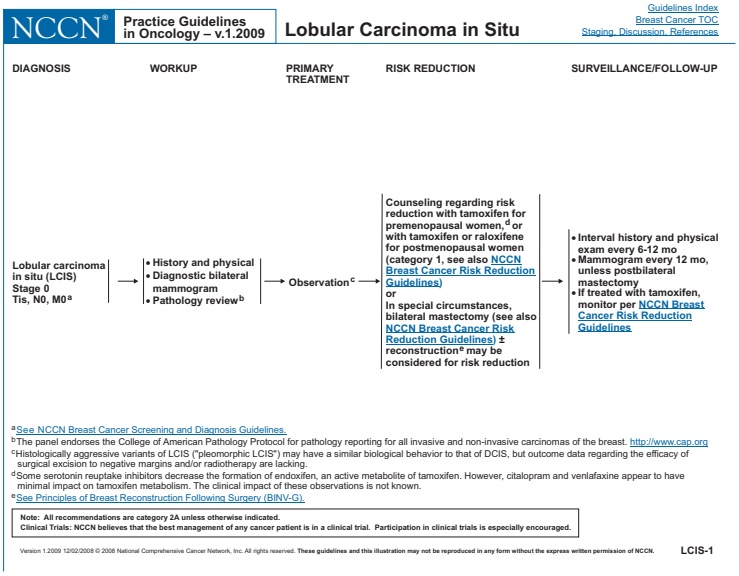
The NCCN Guidelines® are a statement of evidence and consensus of the authors regarding their views of currently accepted approaches to treatment. Any clinician seeking to apply or consult the NCCN Guidelines is expected to use independent medical judgment in the context of individual clinical circumstances to determine any patient's care or treatment. The National Comprehensive Cancer Network® (NCCN®) makes no representations or warranties of any kind regarding their content, use or application and disclaims any responsibility for their application or use in any way. The NCCN Guidelines are copyrighted by National Comprehensive Cancer Network®. All rights reserved. The NCCN Guidelines and the illustrations herein may not be reproduced in any form without the express written permission of NCCN. ©2012.

Version 1.2014, 12/20/13 © National Comprehensive Cancer Network, Inc. 2013, All rights reserved. **The NCCN Guidelines® and this illustration may not be reproduced in any form without the express written permission of NCCN®.**

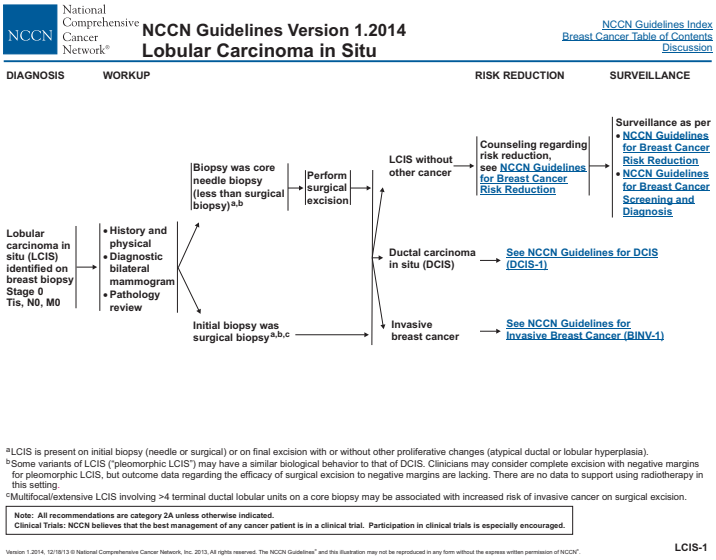
Notes: Figure A.1 shows example changes from NCCN Guidelines. Figures A.1a and A.1b show the table of contents for non-Hodgkins lymphoma (NHL) in 2009 and 2014, respectively. Figure A.1a reproduced with permission from the NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®) for Non-Hodgkin's 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: Changes in Flowchart for Lobular Carcinoma In-Situ Over Time

(a) 2009 Flowchart



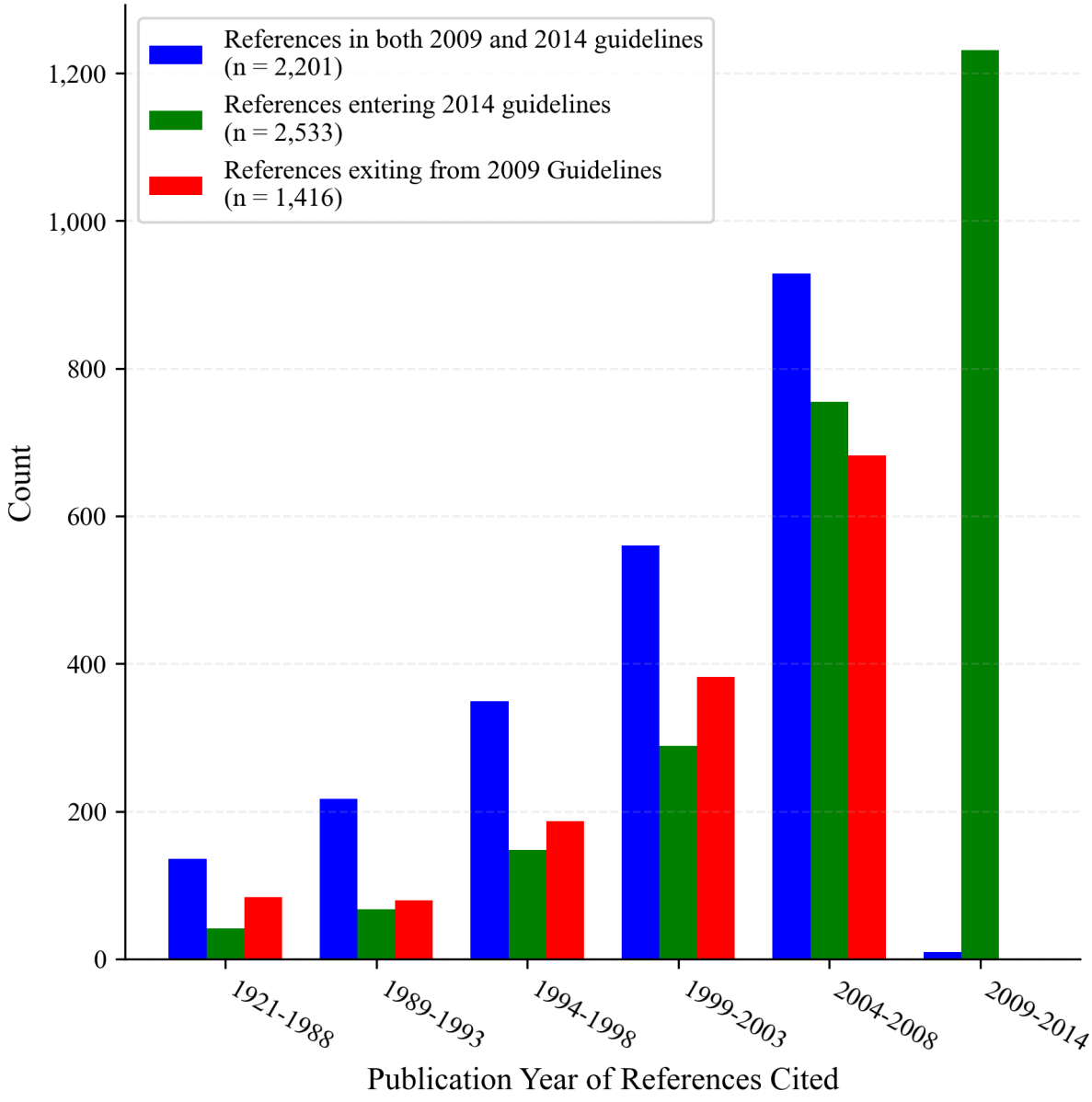
(b) 2014 Flowchart



Notes: Figure A.2 shows example changes from NCCN Guidelines. Figures A.2a and A.2b 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: Growth in oncology clinical practice guideline citations between 2009 and 2014 comes primarily from adding references to new and recent scholarly works

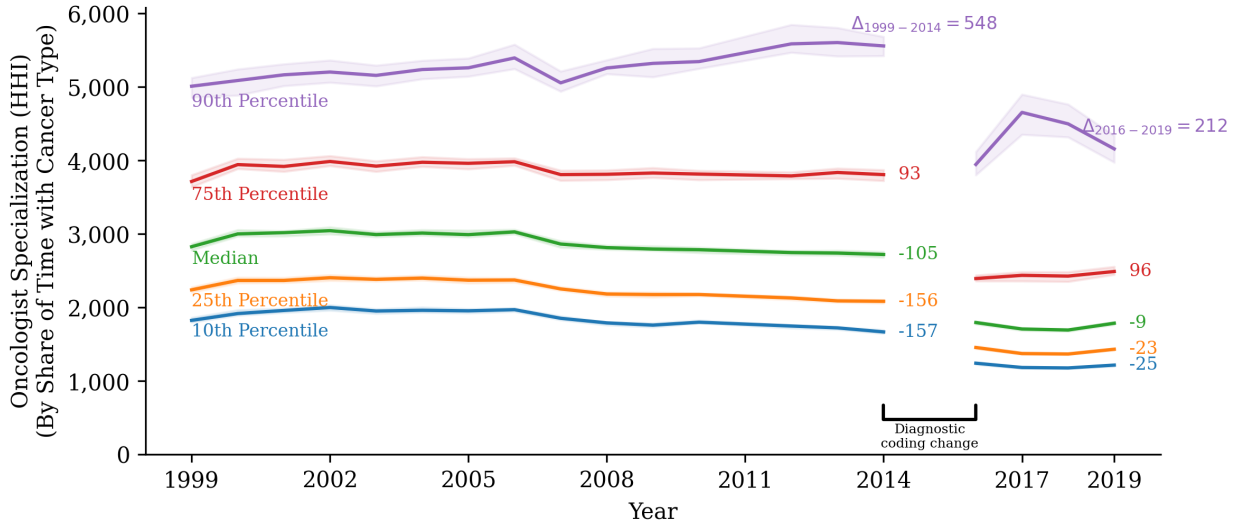
(a) Citations entering, exiting, and staying in 2009 and 2014 guidelines by publication year



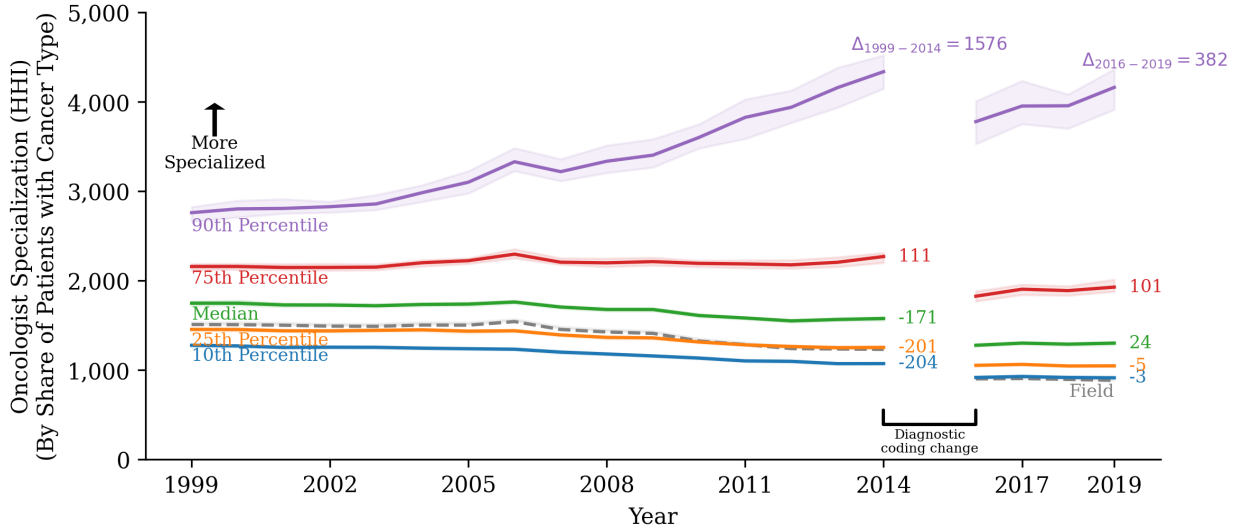
Notes: Figure A.3 describes changes in the publication year of references cited by NCCN Clinical Practice Guidelines in Oncology between the 2009 and 2014 guidelines for screening, risk reduction, and treatment. The x-axis shows the year bin in which the work was published. The y-axis shows the count of works. The count of works entering the 2014 guidelines are shown in green, the count of works in both 2009 and 2014 guidelines are shown in blue, and the count of works exiting the 2009 guidelines are shown in red. The sample includes all unique references in the first version of the 2009 and 2014 guidelines for which a year is identified.

Figure A.4: Specialized oncologists are growing more specialized over time

(a) Oncologist specialization over time, measured using only 25 patients per oncologist

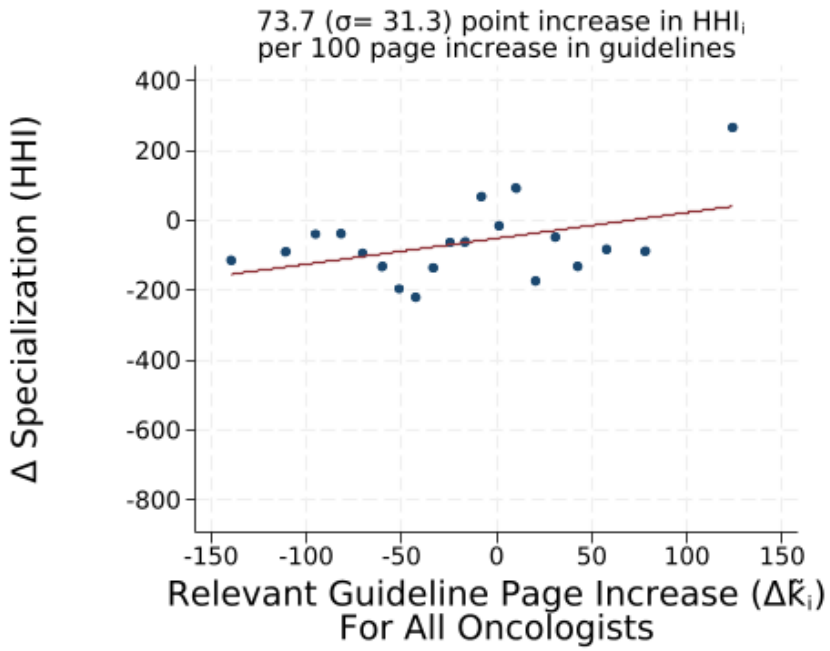


(b) Oncologist specialization over time, measured using patient counts



Notes: Panel (a) shows the distribution of physician-level Herfindahl-Hirschman Index (HHI), a measure of specialization, from 1999-2019, using only 25 random patients per physician per year to calculate HHI. Physician HHI is calculated within physician using shares of time spent evaluating and managing different types of cancer. Time is proxied using the work Relative Value Units for each evaluation and management HCPCS code. Panel (b) shows the distribution of HHI, or specialization, calculating HHI over shares of unique patients, rather than shares of time, seen for outpatient evaluation and management, by cancer type. All physicians in the sample are hematologist-oncologists, 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 10th, 25th, 50th, 75th, and 90th percentile are displayed and labeled. 95% confidence intervals for each percentile are calculated by bootstrapping physicians. The dashed grey line represents the HHI of patients in the sample. 95% confidence intervals are calculated by bootstrapping patients. The trend break in 2015 is attributed to the switch in diagnosis coding schemes, from ICD-9 to ICD-10, and 2015 is dropped from the graph for visual clarity.

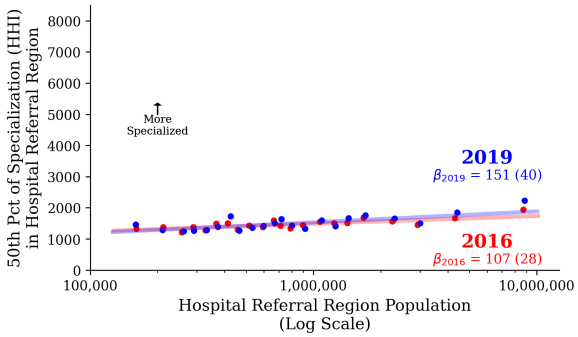
Figure A.5: Knowledge growth increases oncologist specialization in aggregate



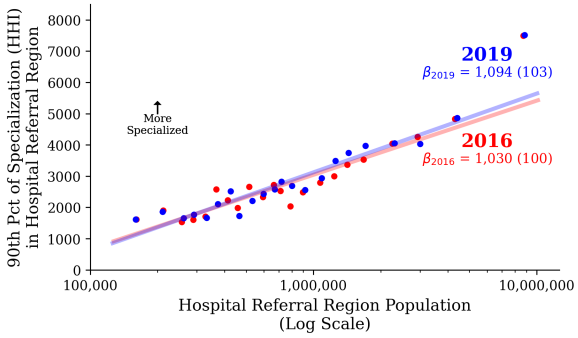
Notes: Figure A.5 shows a binned scatterplot and key regression estimates for the relationship between the recentered exposure to knowledge growth and change in specialization for all oncologists. In the recentered measure, the predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. Predicted exposure is estimated based on the cancer types a physician treats in 2008 and the predicted, rather than realized, growth in knowledge over the next five years. The predicted knowledge growth for each type of cancer is estimated based on a Poisson regression of the number of Medicare patients with the cancer types in 2008, an indicator for if the cancer type has well-established surrogate endpoints, and the interaction of the two. Regression standard errors are robust. All physicians in the sample are hematologist-oncologists, 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.

Figure A.6: No detectable change in specialization by market size between 2016 and 2019

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



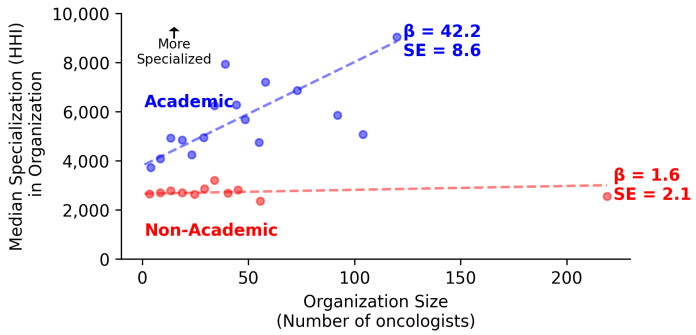
(b) 90th percentile of specialization by market size over time markets (2016 and 2019)



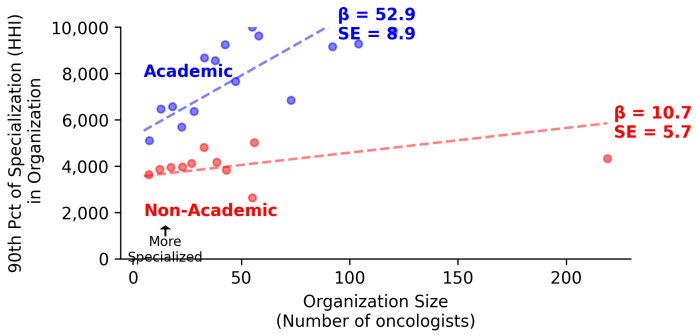
Notes: Panel (a) shows the predicted relationship between specialization and market size if specialization is constrained by market size (Smith, 1776) versus coordination costs (Becker and Murphy, 1992). Panels (a) and (b) show binned scatterplots of the 50th and 90th percentile of oncologist specialization in a market by market size in 2016 and 2019. The x-axis shows the market size, measured as population, in the relevant year. Markets are defined using Hospital Referral Regions (HRR). In panel (a), the y-axis shows the 50th percentile of specialization in market in a year. In panel (b), the y-axis shows the 90th percentile of specialization in 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 that year. The line of best fit between log population and the relevant percentile of market size is plotted. 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.7: Academic organizations facilitate specialization while non-academic ones do not (2008)

(a) Median specialization by organization size and type (2008)



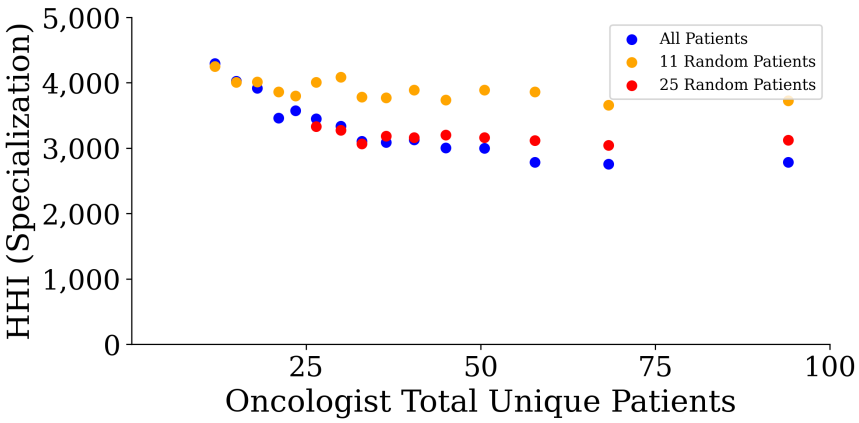
(b) 90th percentile of specialization by organization size and type (2008)



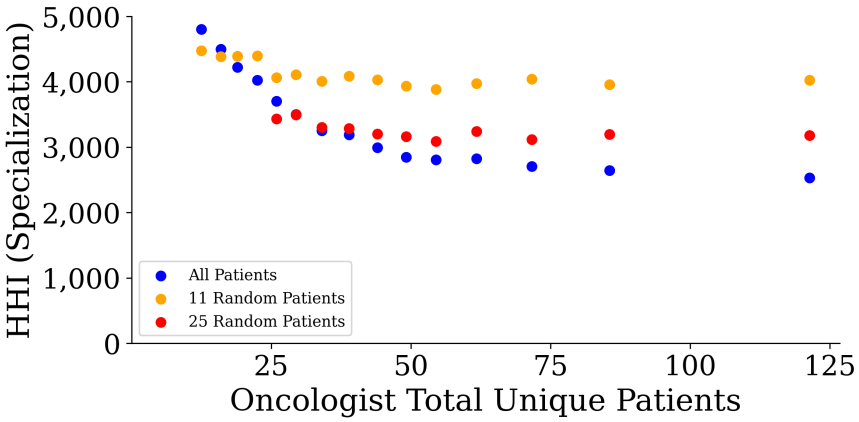
Notes: Panels (a) and (b) show binned scatterplots of the 50th and 90th percentile of oncologist specialization in an organization by size and type in 2008. Organization size is on the x-axis and is estimated as the number of oncologists with at least 25 patients at the organization's tax ID in the 2008 Medicare Data on Provider Practice and Specialty (MD-PPAS), using oncologists' primary tax ID. 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. In this panel, 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 an oncologist evaluated and managed in a year. The best fit line is plotted for the relationship between size and the relevant percentile of specialization in an organization. Points and lines are plotted separately for academic and non-academic organizations. Organization type is determined using tax ID names from the MD-PPAS. See Section 3.3 and Appendix B.2 for discussion of the limitations of variables from the MD-PPAS. Plotted points represent the mean value for organization size bins of width 5 and do not contain an equal number of organizations. 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: Physician level HHI declines with patient volume

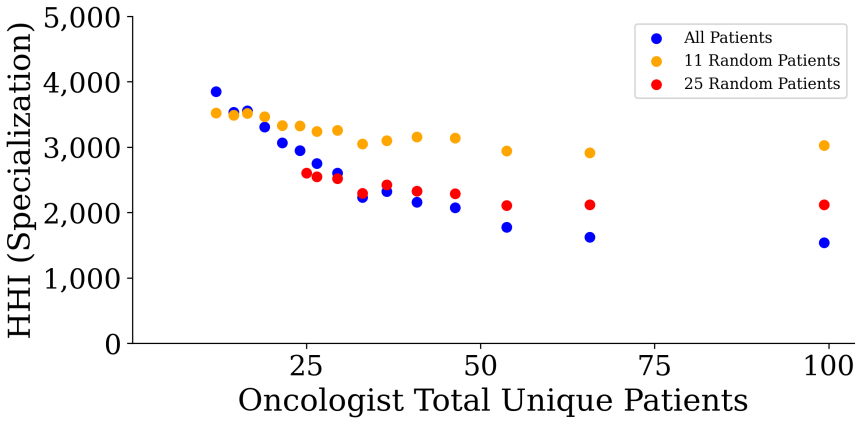
(a) HHI by patient volume (1999)



(b) HHI by patient volume (2009)



(c) HHI by patient volume (2019)



Notes: Figure A.8 displays binned scatter plots of physician-level Herfindahl-Hirschman Index (HHI) by patient volume. HHI is calculated using 11 random patients per physician in the yellow dots, 25 random patients per physician in the red dots, and all observed patients in the blue dots. Figures A.8a, A.8b, show the relationship for 1999, 2009 and 2019 respectively. Physician HHI is calculated within physician using shares of time spent evaluating and managing different types of cancer. Time is proxied using the work Relative Value Units for each evaluation and management HCPCS code. All physicians in the sample are hematologist-oncologists, oncologists, or hematologists who saw at least 11 unique cancer patients in the year in a 20% sample of Medicare Part B outpatient evaluation and management claims.

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

Year Site	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's 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	71	83	168	218
NHL	49	155	471	658
Neuroendocrine	46	64	98	141
Occult	30	43	55	65
Ovary and Adnexa	26	44	83	371
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	**1,129**	**1,832**	**3,589**	**5,825**

Notes: Figure A.1 displays the page count of the first set of treatment guidelines in 2002, 2009, 2014, and 2020 from NCCN Guidelines. The NCCN considered the 2002 guidelines to be a "complete library" of oncology guidelines (NCCN, 2023*b*). However, in the years since, they have added guidelines for additional types of cancer.

Table A.2: For specialized oncologists, recentering procedure removes most correlation between exposure to knowledge growth and physician characteristics

	(1) Unadjusted exposure to knowledge growth (Δk_i)	(2) Remove physician scope variation ($\Delta\tilde{k}_i$)	(3) Remove cancer prevalence variation ($\Delta\tilde{k}_i$)	(4) Remove endpoint type variation ($\Delta\tilde{k}_i$)
Market population (millions)	-29.35	-11.02	-5.809	-0.842
	(4.901)	(1.880)	(1.596)	(1.010)
Count oncologists at main org	-1.143	-0.262	0.121	0.109
	(0.288)	(0.110)	(0.0866)	(0.0494)
Physician age	0.305	0.836	0.171	-0.484
	(0.985)	(0.422)	(0.378)	(0.246)
Patient Count	5.168	0.720	-0.302	-0.129
	(0.352)	(0.120)	(0.115)	(0.0816)
Observations	988	988	988	988
R-squared	0.286	0.090	0.019	0.012
Recentered		Yes	Yes	Yes

Notes: This table shows the results of a regression of exposure to knowledge growth on observable physician characteristics for specialized oncologists. Specialized oncologists are defined as oncologists in the top 25% of the specialization distribution in 2009, measured using a Herfindahl—-Hirschman Index (HHI) over types of cancer managed. The physician characteristics include the population of the physicians' Hospital Referral Region in millions, the number of other oncologists at the primary organization that they work at, and physicians' patient volume in 2008. In Column 1, the outcome variable is raw exposure to knowledge growth, measured as the total increase in guideline page counts for cancer types from 2009 to 2014 for cancer types the physician managed in 2008. In Columns 2-4, the outcomes are increasingly recentered measures of exposure to knowledge growth. In recentering, predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. In Column 2, predicted knowledge growth is the same for each cancer and estimated as mean knowledge growth per cancer type, weighted by the number of physicians who treat that cancer. In Column 3, the predicted knowledge growth for each type of cancer is estimated using on a Poisson regression of the number of Medicare patients with the cancer types in 2008. Column 4 is similar, but predicted knowledge growth also adds indicator for if the cancer type has well-established surrogate endpoints and the interaction between the number of Medicare patients and surrogate endpoints. The full recentering procedure removes variation in exposure to knowledge growth that arises become some physicians have a broader scope and because some cancer types predictably have more innovation. All physicians in the sample are hematologist-oncologists, 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. All regression standard errors are robust.

Table A.3: For general oncologists, physician characteristics no longer correlate with exposure to knowledge growth after recentering procedure

	(1) Unadjusted exposure to knowledge growth (Δk_i)	(2) Remove physician scope variation ($\Delta \tilde{k}_i$)	(3) Remove cancer prevalence variation ($\Delta \tilde{k}_i$)	(4) Remove endpoint type variation ($\Delta \tilde{k}_i$)
Market population (millions)	-3.707	-0.368	-0.0355	0.474
	(1.371)	(0.677)	(0.758)	(0.640)
Count oncologists at main org	0.00466	-0.0177	0.00903	0.000144
	(0.0559)	(0.0282)	(0.0337)	(0.0302)
Physician age	-0.843	0.171	0.183	-0.0190
	(0.282)	(0.139)	(0.161)	(0.138)
Patient Count	3.635	0.0565	-0.238	0.000888
	(0.0999)	(0.0405)	(0.0438)	(0.0392)
Observations	2,960	2,960	2,960	2,960
R-squared	0.409	0.002	0.009	0.000
Recentered		Yes	Yes	Yes

Notes: This table shows the results of a regression of exposure to knowledge growth on observable physician characteristics for general oncologists. General oncologists are defined as oncologists in the bottom 75% of the specialization distribution in 2009, measured using a Herfindahl—-Hirschman Index (HHI) over types of cancer managed. The physician characteristics include the population of the physicians' Hospital Referral Region in millions, the number of other oncologists at the primary organization they work at, and physicians' patient volume in 2008. In Column 1, the outcome variable is raw exposure to knowledge growth, measured as the total increase in guideline page counts for cancer types from 2009 to 2014 for cancer types the physician managed in 2008. In Columns 2-4, the outcomes are increasingly recentered measures of exposure to knowledge growth. In recentering, predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. In Column 2, predicted knowledge growth is the same for each cancer and estimated as mean knowledge growth per cancer type, weighted by the number of physicians who treat that cancer. In Column 3, the predicted knowledge growth for each type of cancer is estimated using on a Poisson regression of the number of Medicare patients with the cancer types in 2008. Column 4 is similar, but predicted knowledge growth also adds indicator for if the cancer type has well-established surrogate endpoints and the interaction between the number of Medicare patients and surrogate endpoints. The full recentering procedure removes variation in exposure to knowledge growth that arises become some physicians have a broader scope and because some cancer types predictably have more innovation. All physicians in the sample are hematologist-oncologists, 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. All regression standard errors are robust.

Table A.4: Knowledge growth increases oncologist specialization in aggregate

	Dependent Variable: HHI					
	(1)	(2)	(3)	(4)	(5)	(6)
	Knowledge Growth Recentered ($\Delta\tilde{k}_i$)	Restrict Sample	Add Age FE	Add Market FE	Add Pat Count FE	Add Org Size FE
Increase in Guideline Length (100 Pages)	71.95	71.95	71.62	87.93	88.79	89.36
	(31.43)	(31.43)	(31.41)	(32.72)	(32.71)	(32.67)
Observations	3,931	3,931	3,931	3,931	3,931	3,931
R-squared	0.002	0.002	0.002	0.087	0.087	0.088
Age FE			Yes	Yes	Yes	Yes
Market FE				Yes	Yes	Yes
Org Size FE					Yes	Yes
Pat Count FE						Yes

Notes: This table shows regression estimates for the relationship between exposure to knowledge growth and change in specialization for all oncologists. Specialization is measured using a Herfindahl—-Hirschman Index (HHI) over types of cancer managed. The measure of exposure to knowledge growth is recentered to control for variation in physician scope and field expected knowledge growth. Column 1 shows the coefficient estimate for the univariate regression. The regression standard errors are robust. Column 2 restricts the sample to physicians where we observe age, hospital referral region (HRR) and the number of other oncologists at their primary organization. Columns 3, 4, and 5 add fixed effects for the physicians age group, hospital referral region (HRR) and the number of other oncologists at their primary organization, respectively. In the recentered measure of exposure to knowledge growth, predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. Raw 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. Predicted exposure is estimated based on the cancer types a physician treats in 2008 and the predicted, rather than realized, growth in knowledge over the next five years. The predicted knowledge growth for each type of cancer is estimated based on a Poisson regression of the number of Medicare patients with the cancer types in 2008, an indicator for if the cancer type has well-established surrogate endpoints, and the interaction of the two. All physicians in the sample are hematologist-oncologists, 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.5: For specialized oncologists, knowledge growth does not significantly reduce the number of types of cancers managed

	Dependent Variable: HHI					
	(1)	(2)	(3)	(4)	(5)	(6)
	Knowledge Growth Recentered ($\Delta\tilde{k}_i$)	Restrict Sample	Add Age FE	Add Market FE	Add Pat Count FE	Add Org Size FE
Increase in Guideline Length (100 Pages)	-0.135	-0.135	-0.172	-0.197	-0.222	-0.248
	(0.174)	(0.174)	(0.174)	(0.188)	(0.188)	(0.190)
Observations	924	924	924	924	924	924
R-squared	0.001	0.001	0.009	0.221	0.242	0.246
Age FE			Yes	Yes	Yes	Yes
Market FE				Yes	Yes	Yes
Pat Count FE					Yes	Yes
Org Size FE						Yes

Notes: This table shows regression estimates for the relationship between exposure to knowledge growth and the number of cancers managed for specialized oncologists. Specialized oncologists are defined as oncologists in the top 25% of the specialization distribution in 2009, measured using a Herfindahl—-Hirschman Index (HHI) over types of cancer managed. The measure of exposure to knowledge growth is recentered to control for variation in physician scope and field expected knowledge growth. Column 1 shows the coefficient estimate for the univariate regression. The regression standard errors are robust. Column 2 restricts the sample to physicians where we observe age, hospital referral region (HRR) and the number of other oncologists at the primary organization. Columns 3, 4, and 5 add fixed effects for the physicians age group, HRR and the number of other oncologists at their primary organization, respectively. In the recentered measure of exposure to knowledge growth, predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. Raw 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. Predicted exposure is estimated based on the cancer types a physician treats in 2008 and the predicted, rather than realized, growth in knowledge over the next five years. The predicted knowledge growth for each type of cancer is estimated based on a Poisson regression of the number of Medicare patients with the cancer types in 2008, an indicator for if the cancer type has well-established surrogate endpoints, and the interaction of the two. All physicians in the sample are hematologist-oncologists, 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.6: For general oncologists, knowledge growth reduces the number of types of cancers managed

	Dependent Variable: HHI					
	(1)	(2)	(3)	(4)	(5)	(6)
	Knowledge Growth Recentered ($\Delta\tilde{k}_i$)	Restrict Sample	Add Age FE	Add Market FE	Add Pat Count FE	Add Org Size FE
Increase in Guideline Length (100 Pages)	-0.182	-0.182	-0.180	-0.236	-0.244	-0.236
	(0.0909)	(0.0909)	(0.0908)	(0.0935)	(0.0937)	(0.0935)
Observations	2,937	2,937	2,937	2,937	2,937	2,937
R-squared	0.001	0.001	0.022	0.160	0.178	0.180
Age FE			Yes	Yes	Yes	Yes
Market FE				Yes	Yes	Yes
Pat Count FE					Yes	Yes
Org Size FE						Yes

Notes: This table shows regression estimates for the relationship between exposure to knowledge growth and the number of cancers managed for general oncologists. General oncologists are defined as oncologists in the bottom 75% of the specialization distribution in 2009, measured using a Herfindahl—-Hirschman Index (HHI) over types of cancer managed. The measure of exposure to knowledge growth is recentered to control for variation in physician scope and field expected knowledge growth. Column 1 shows the coefficient estimate for the univariate regression. The regression standard errors are robust. Column 2 restricts the sample to physicians where we observe age, hospital referral region (HRR) and the number of other oncologists at their primary organization. Columns 3, 4, and 5 add fixed effects for the physicians age group, hospital referral region (HRR) and the number of other oncologists at their primary organization, respectively. In the recentered measure of exposure to knowledge growth, predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. Raw 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. Predicted exposure is estimated based on the cancer types a physician treats in 2008 and the predicted, rather than realized, growth in knowledge over the next five years. The predicted knowledge growth for each type of cancer is estimated based on a Poisson regression of the number of Medicare patients with the cancer types in 2008, an indicator for if the cancer type has well-established surrogate endpoints, and the interaction of the two. All physicians in the sample are hematologist-oncologists, 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.7: For oncologists in aggregate, knowledge growth may reduce the number of types of cancers managed

	Dependent Variable: HHI					
	(1)	(2)	(3)	(4)	(5)	(6)
	Knowledge Growth Recentered ($\Delta\tilde{k}_i$)	Restrict Sample	Add Age FE	Add Market FE	Add Pat Count FE	Add Org Size FE
Increase in Guideline Length (100 Pages)	-0.145	-0.145	-0.156	-0.213	-0.233	-0.233
	(0.0801)	(0.0801)	(0.0798)	(0.0817)	(0.0812)	(0.0813)
Observations	3,931	3,931	3,931	3,931	3,931	3,931
R-squared	0.001	0.001	0.018	0.143	0.164	0.164
Age FE			Yes	Yes	Yes	Yes
Market FE				Yes	Yes	Yes
Pat Count FE					Yes	Yes
Org Size FE						Yes

Notes: This table shows regression estimates for the relationship between exposure to knowledge growth and the number of cancers managed for specialized oncologists. Specialized oncologists are defined as oncologists in the top 25% of the specialization distribution in 2009, measured using a Herfindahl—-Hirschman Index (HHI) over types of cancer managed. The measure of exposure to knowledge growth is recentered to control for variation in physician scope and field expected knowledge growth. Column 1 shows the coefficient estimate for the univariate regression. The regression standard errors are robust. Column 2 restricts the sample to physicians where we observe age, hospital referral region (HRR) and the number of other oncologists at the primary organization. Columns 3, 4, and 5 add fixed effects for the physicians age group, HRR and the number of other oncologists at their primary organization, respectively. In the recentered measure of exposure to knowledge growth, predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. Raw 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. Predicted exposure is estimated based on the cancer types a physician treats in 2008 and the predicted, rather than realized, growth in knowledge over the next five years. The predicted knowledge growth for each type of cancer is estimated based on a Poisson regression of the number of Medicare patients with the cancer types in 2008, an indicator for if the cancer type has well-established surrogate endpoints, and the interaction of the two. All physicians in the sample are hematologist-oncologists, 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: Knowledge growth increases specialization among specialized oncologists after raising patient count threshold for exposure to knowledge growth but not when restricting to new patient visits

	Dependent Variable: HHI			
	(1)	(2)	(3)	(4)
Increase in Guideline Length (100 Pages)	196.0	186.0	310.7	-20.22
	(79.81)	(86.76)	(109.5)	(162.5)
Observations	988	988	988	326
R-squared	0.007	0.004	0.008	0.000
Min Pat Count per Field	1	2	3	1
HHI Sample	All pat	All pat	All pat	All pat
Claims	Outpatient EM	Outpatient EM	Outpatient EM	New EM
Min Pat Count per Doc		25	25	11

Notes: This table shows regression estimates for the relationship between exposure to knowledge growth and change in specialization for oncologists who start in the top 25% of the initial specialization distribution. Specialization is measured using a Herfindahl—-Hirschman Index (HHI) over types of cancer managed. The measure of exposure to knowledge growth is recentered to control for variation in physician scope and field expected knowledge growth. Column 1 shows the baseline univariate regression. Column 2 and 3 raise the minimum patient count to be considered to be exposed to a cancer type in the baseline year. Column 4 restricts the sample to new evaluation and management claims (versus all evaluation and management claims) and lowers the threshold for a physician to be included in the sample to 11 patients per year. Column 5 restricts the sample to outpatient evaluation and management claims. In the recentered measure of exposure to knowledge growth, predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. Raw 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. Predicted exposure is estimated based on the cancer types a physician treats in 2008 and the predicted, rather than realized, growth in knowledge over the next five years. The predicted knowledge growth for each type of cancer is estimated based on a Poisson regression of the number of Medicare patients with the cancer types in 2008, an indicator for if the cancer type has well-established surrogate endpoints, and the interaction of the two. 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.

Table A.9: Knowledge growth does not increase specialization among general oncologists after raising patient count threshold for exposure to knowledge growth or when restricting to new patient visits

	Dependent Variable: HHI			
	(1)	(2)	(3)	(4)
Increase in Guideline Length (100 Pages)	50.57	7.200	8.637	68.64
	(31.32)	(32.01)	(34.67)	(67.85)
Observations	2,962	2,962	2,962	978
R-squared	0.001	0.000	0.000	0.001
Min Pat Count per Field	1	2	3	1
HHI Sample	All pat	All pat	All pat	All pat
Claims	Outpatient EM	Outpatient EM	Outpatient EM	New EM
Min Pat Count per Doc		25	25	11

Notes: This table shows regression estimates for the relationship between exposure to knowledge growth and change in specialization for oncologists who start in the lower 75% of the initial specialization distribution. Specialization is measured using a Herfindahl—-Hirschman Index (HHI) over types of cancer managed. The measure of exposure to knowledge growth is recentered to control for variation in physician scope and field expected knowledge growth. Column 1 shows the baseline univariate regression. Column 2 and 3 raise the minimum patient count to be considered to be exposed to a cancer type in the baseline year. Column 4 restricts the sample to new evaluation and management claims (versus all evaluation and management claims) and lowers the threshold for a physician to be included in the sample to 11 patients per year. Column 5 restricts the sample to outpatient evaluation and management claims. In the recentered measure of exposure to knowledge growth, predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. Raw 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. Predicted exposure is estimated based on the cancer types a physician treats in 2008 and the predicted, rather than realized, growth in knowledge over the next five years. The predicted knowledge growth for each type of cancer is estimated based on a Poisson regression of the number of Medicare patients with the cancer types in 2008, an indicator for if the cancer type has well-established surrogate endpoints, and the interaction of the two. 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.

Table A.10: Estimated effect of knowledge growth on specialization for specialists, estimated with 25 patients per oncologist

	Dependent Variable: HHI		
	(1)	(2)	(3)
Increase in Guideline Length (100 Pages)	203.3	-13.24	81.80
	(82.21)	(93.34)	(99.15)
Observations	988	988	988
R-squared	0.006	0.000	0.001
Knowledge Exposure Sample	25 pat	All pat	25 pat
HHI Sample	All pat	25 pat	25 pat

Notes: This table shows regression estimates for the relationship between exposure to knowledge growth and change in specialization for oncologists who start in the top 25% of the initial specialization distribution. Specialization is measured using a Herfindahl—-Hirschman Index (HHI) over types of cancer managed. The measure of exposure to knowledge growth is recentered to control for variation in physician scope and field expected knowledge growth. Column 1 shows the regression with exposure to knowledge growth calculated using only 25 random patients. Column 2 shows the regression with $H\hat{H}I_i$ calculated using only 25 random patients. Column 3 shows the regression with both values calculated using only 25 random patients In the recentered measure of exposure to knowledge growth, predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. Raw 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. Predicted exposure is estimated based on the cancer types a physician treats in 2008 and the predicted, rather than realized, growth in knowledge over the next five years. The predicted knowledge growth for each type of cancer is estimated based on a Poisson regression of the number of Medicare patients with the cancer types in 2008, an indicator for if the cancer type has well-established surrogate endpoints, and the interaction of the two. All physicians in the sample are hematologist-oncologists, 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.11: Estimated effect of knowledge growth on specialization for generalists, estimated with 25 patients per oncologist

	Dependent Variable: HHI		
	(1)	(2)	(3)
Increase in Guideline Length (100 Pages)	4.527	58.99	-1.202
	(31.74)	(42.84)	(43.66)
Observations	2,962	2,962	2,962
R-squared	0.000	0.001	0.000
Knowledge Exposure Sample	25 pat	All pat	25 pat
HHI Sample	All pat	25 pat	25 pat

Notes: This table shows regression estimates for the relationship between exposure to knowledge growth and change in specialization for oncologists who start in the lower 75% of the initial specialization distribution. Specialization is measured using a Herfindahl—-Hirschman Index (HHI) over types of cancer managed. The measure of exposure to knowledge growth is recentered to control for variation in physician scope and field expected knowledge growth. Column 1 shows the regression with exposure to knowledge growth calculated using only 25 random patients. Column 2 shows the regression with $H\hat{H}I_i$ calculated using only 25 random patients. Column 3 shows the regression with both values calculated using only 25 random patients In the recentered measure of exposure to knowledge growth, predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. Raw 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. Predicted exposure is estimated based on the cancer types a physician treats in 2008 and the predicted, rather than realized, growth in knowledge over the next five years. The predicted knowledge growth for each type of cancer is estimated based on a Poisson regression of the number of Medicare patients with the cancer types in 2008, an indicator for if the cancer type has well-established surrogate endpoints, and the interaction of the two. All physicians in the sample are hematologist-oncologists, 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.12: Estimated effect of knowledge growth on specialization in aggregate, estimated with 25 patients per oncologist

	Dependent Variable: HHI		
	(1)	(2)	(3)
Increase in Guideline Length (100 Pages)	33.89	35.44	13.77
	(31.40)	(39.95)	(40.62)
Observations	3,950	3,950	3,950
R-squared	0.000	0.000	0.000
Knowledge Exposure Sample	25 pat	All pat	25 pat
HHI Sample	All pat	25 pat	25 pat

Notes: This table shows regression estimates for the relationship between exposure to knowledge growth and change in specialization for all oncologists. Specialization is measured using a Herfindahl—-Hirschman Index (HHI) over types of cancer managed. The measure of exposure to knowledge growth is recentered to control for variation in physician scope and field expected knowledge growth. Column 1 shows the regression with exposure to knowledge growth calculated using only 25 random patients. Column 2 shows the regression with $\hat{HHI}_i$ calculated using only 25 random patients. Column 3 shows the regression with both values calculated using only 25 random patients In the recentered measure of exposure to knowledge growth, predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. Raw 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. Predicted exposure is estimated based on the cancer types a physician treats in 2008 and the predicted, rather than realized, growth in knowledge over the next five years. The predicted knowledge growth for each type of cancer is estimated based on a Poisson regression of the number of Medicare patients with the cancer types in 2008, an indicator for if the cancer type has well-established surrogate endpoints, and the interaction of the two. Unless otherwise specified, all physicians in the sample are hematologist-oncologists, 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.13: Knowledge growth, measured using guideline citations, increases specialization for oncologists in the top 25% of the initial specialization distribution

	Dependent Variable: HHI					
	(1)	(2)	(3)	(4)	(5)	(6)
	Knowledge Growth Recentered ($\Delta\tilde{k}_i$)	Restrict Sample	Add Age FE	Add Market FE	Add Pat Count FE	Add Org Size FE
Increase in Guideline Citations (100 Citations)	64.91	69.41	70.88	84.97	85.27	83.76
	(18.49)	(18.83)	(18.89)	(21.37)	(21.45)	(21.61)
Observations	988	924	924	924	924	924
R-squared	0.014	0.016	0.017	0.182	0.187	0.188
Age FE			Yes	Yes	Yes	Yes
Market FE				Yes	Yes	Yes
Org Size FE					Yes	Yes
Pat Count FE						Yes

Notes: This table shows regression estimates for the relationship between exposure to knowledge growth and change in specialization for oncologists who start in the top 25% of the initial specialization distribution. Specialization is measured using a Herfindahl—-Hirschman Index (HHI) over types of cancer managed. The measure of exposure to knowledge growth is recentered to control for variation in physician scope and field expected knowledge growth. Column 1 shows the coefficient estimate for the univariate regression. The regression standard errors are robust. Column 2 restricts the sample to physicians where we observe age, hospital referral region (HRR) and the number of other oncologists at their primary organization. Columns 3, 4, and 5 add fixed effects for the physicians age group, hospital referral region (HRR) and the number of other oncologists at their primary organization, respectively. In the recentered measure of exposure to knowledge growth, predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. Raw exposure to knowledge growth is calculated as the total increase in the count of guideline citations over 2009 to 2014 for the types of cancers a physician treats in 2008. Predicted exposure is estimated based on the cancer types a physician treats in 2008 and the predicted, rather than realized, growth in knowledge over the next five years. The predicted knowledge growth for each type of cancer is estimated based on a Poisson regression of the number of Medicare patients with the cancer types in 2008, an indicator for if the cancer type has well-established surrogate endpoints, and the interaction of the two. All physicians in the sample are hematologist-oncologists, 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.14: Knowledge growth, measured using guideline citations, increases specialization for oncologists in the bottom 75% of the initial specialization distribution

	Dependent Variable: HHI					
	(1)	(2)	(3)	(4)	(5)	(6)
	Knowledge Growth Recentered ($\Delta\tilde{k}_i$)	Restrict Sample	Add Age FE	Add Market FE	Add Pat Count FE	Add Org Size FE
Increase in Guideline Citations (100 Citations)	15.62	16.58	16.60	19.26	18.94	18.64
	(7.250)	(7.263)	(7.268)	(7.785)	(7.770)	(7.742)
Observations	2,962	2,937	2,937	2,937	2,937	2,937
R-squared	0.002	0.002	0.002	0.108	0.108	0.110
Age FE			Yes	Yes	Yes	Yes
Market FE				Yes	Yes	Yes
Org Size FE					Yes	Yes
Pat Count FE						Yes

Notes: This table shows regression estimates for the relationship between exposure to knowledge growth and change in specialization for oncologists who start in the bottom 75% of the initial specialization distribution. Specialization is measured using a Herfindahl—-Hirschman Index (HHI) over types of cancer managed. The measure of exposure to knowledge growth is recentered to control for variation in physician scope and field expected knowledge growth. Column 1 shows the coefficient estimate for the univariate regression. The regression standard errors are robust. Column 2 restricts the sample to physicians where we observe age, hospital referral region (HRR) and the number of other oncologists at their primary organization. Columns 3, 4, and 5 add fixed effects for the physicians age group, hospital referral region (HRR) and the number of other oncologists at their primary organization, respectively. In the recentered measure of exposure to knowledge growth, predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. Raw exposure to knowledge growth is calculated as the total increase in the count of guideline citations over 2009 to 2014 for the types of cancers a physician treats in 2008. Predicted exposure is estimated based on the cancer types a physician treats in 2008 and the predicted, rather than realized, growth in knowledge over the next five years. The predicted knowledge growth for each type of cancer is estimated based on a Poisson regression of the number of Medicare patients with the cancer types in 2008, an indicator for if the cancer type has well-established surrogate endpoints, and the interaction of the two. All physicians in the sample are hematologist-oncologists, 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.15: Knowledge growth, measured using guideline citations, increases oncologist specialization in aggregate

	Dependent Variable: HHI					
	(1)	(2)	(3)	(4)	(5)	(6)
	Knowledge Growth Recentered ($\Delta\tilde{k}_i$)	Restrict Sample	Add Age FE	Add Market FE	Add Pat Count FE	Add Org Size FE
Increase in Guideline Citations (100 Citations)	26.46	26.28	26.23	29.94	30.28	30.25
	(7.281)	(7.312)	(7.317)	(7.720)	(7.728)	(7.721)
Observations	3,950	3,931	3,931	3,931	3,931	3,931
R-squared	0.004	0.004	0.004	0.089	0.089	0.090
Age FE			Yes	Yes	Yes	Yes
Market FE				Yes	Yes	Yes
Org Size FE					Yes	Yes
Pat Count FE						Yes

Notes: This table shows regression estimates for the relationship between exposure to knowledge growth and change in specialization for all oncologists. Specialization is measured using a Herfindahl—-Hirschman Index (HHI) over types of cancer managed. The measure of exposure to knowledge growth is recentered to control for variation in physician scope and field expected knowledge growth. Column 1 shows the coefficient estimate for the univariate regression. The regression standard errors are robust. Column 2 restricts the sample to physicians where we observe age, hospital referral region (HRR) and the number of other oncologists at their primary organization. Columns 3, 4, and 5 add fixed effects for the physicians age group, hospital referral region (HRR) and the number of other oncologists at their primary organization, respectively. In the recentered measure of exposure to knowledge growth, predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. Raw exposure to knowledge growth is calculated as the total increase in the count of guideline citations over 2009 to 2014 for the types of cancers a physician treats in 2008. Predicted exposure is estimated based on the cancer types a physician treats in 2008 and the predicted, rather than realized, growth in knowledge over the next five years. The predicted knowledge growth for each type of cancer is estimated based on a Poisson regression of the number of Medicare patients with the cancer types in 2008, an indicator for if the cancer type has well-established surrogate endpoints, and the interaction of the two. All physicians in the sample are hematologist-oncologists, 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.16: Knowledge growth, recentered using cancer incidence instead of prevalence, increases specialization for oncologists in the top 25% of the initial specialization distribution

	Dependent Variable: HHI					
	(1)	(2)	(3)	(4)	(5)	(6)
	Knowledge Growth Recentered ($\Delta\tilde{k}_i$)	Restrict Sample	Add Age FE	Add Market FE	Add Pat Count FE	Add Org Size FE
Increase in Guideline Length (100 Pages)	243.7	243.7	251.1	283.1	285.7	276.5
	(82.54)	(82.54)	(82.91)	(90.74)	(91.71)	(93.01)
Observations	924	924	924	924	924	924
R-squared	0.010	0.010	0.012	0.174	0.179	0.180
Age FE			Yes	Yes	Yes	Yes
Market FE				Yes	Yes	Yes
Org Size FE					Yes	Yes
Pat Count FE						Yes

Notes: This table shows regression estimates for the relationship between exposure to knowledge growth and change in specialization for oncologists who start in the top 25% of the initial specialization distribution. Specialization is measured using a Herfindahl—-Hirschman Index (HHI) over types of cancer managed. The measure of exposure to knowledge growth is recentered to control for variation in physician scope and field expected knowledge growth. Column 1 shows the coefficient estimate for the univariate regression. The regression standard errors are robust. Column 2 restricts the sample to physicians where we observe age, hospital referral region (HRR) and the number of other oncologists at their primary organization. Columns 3, 4, and 5 add fixed effects for the physicians age group, hospital referral region (HRR) and the number of other oncologists at their primary organization, respectively. In the recentered measure of exposure to knowledge growth, predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. Raw exposure to knowledge growth is calculated as the total increase in the count of guideline citations over 2009 to 2014 for the types of cancers a physician treats in 2008. Predicted exposure is estimated based on the cancer types a physician treats in 2008 and the predicted, rather than realized, growth in knowledge over the next five years. The predicted knowledge growth for each type of cancer is estimated based on a Poisson regression of the incidence of cancer types in Medicare in 2008, an indicator for if the cancer type has well-established surrogate endpoints, and the interaction of the two. All physicians in the sample are hematologist-oncologists, 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.17: Knowledge growth, recentered using cancer incidence instead of prevalence, increases specialization for oncologists in the bottom 75% of the initial specialization distribution

	Dependent Variable: HHI					
	(1)	(2)	(3)	(4)	(5)	(6)
	Knowledge Growth Recentered ($\Delta\tilde{k}_i$)	Restrict Sample	Add Age FE	Add Market FE	Add Pat Count FE	Add Org Size FE
Increase in Guideline Length (100 Pages)	60.78	60.78	60.66	73.76	72.74	72.42
	(32.03)	(32.03)	(32.02)	(33.57)	(33.43)	(33.33)
Observations	2,937	2,937	2,937	2,937	2,937	2,937
R-squared	0.001	0.001	0.002	0.108	0.108	0.110
Age FE			Yes	Yes	Yes	Yes
Market FE				Yes	Yes	Yes
Org Size FE					Yes	Yes
Pat Count FE						Yes

Notes: This table shows regression estimates for the relationship between exposure to knowledge growth and change in specialization for oncologists who start in the bottom 75% of the initial specialization distribution. Specialization is measured using a Herfindahl—-Hirschman Index (HHI) over types of cancer managed. The measure of exposure to knowledge growth is recentered to control for variation in physician scope and field expected knowledge growth. Column 1 shows the coefficient estimate for the univariate regression. The regression standard errors are robust. Column 2 restricts the sample to physicians where we observe age, hospital referral region (HRR) and the number of other oncologists at their primary organization. Columns 3, 4, and 5 add fixed effects for the physicians age group, hospital referral region (HRR) and the number of other oncologists at their primary organization, respectively. In the recentered measure of exposure to knowledge growth, predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. Raw exposure to knowledge growth is calculated as the total increase in the count of guideline citations over 2009 to 2014 for the types of cancers a physician treats in 2008. Predicted exposure is estimated based on the cancer types a physician treats in 2008 and the predicted, rather than realized, growth in knowledge over the next five years. The predicted knowledge growth for each type of cancer is estimated based on a Poisson regression of the incidence of a cancer type in Medicare in 2008, an indicator for if the cancer type has well-established surrogate endpoints, and the interaction of the two. All physicians in the sample are hematologist-oncologists, 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.18: Knowledge growth, recentered using cancer incidence instead of prevalence, increases oncologist specialization in aggregate

	Dependent Variable: HHI					
	(1)	(2)	(3)	(4)	(5)	(6)
	Knowledge Growth Recentered ($\Delta\tilde{k}_i$)	Restrict Sample	Add Age FE	Add Market FE	Add Pat Count FE	Add Org Size FE
Increase in Guideline Length (100 Pages)	83.32	83.32	82.92	100.1	102.0	102.2
	(32.04)	(32.04)	(32.02)	(33.40)	(33.38)	(33.33)
Observations	3,931	3,931	3,931	3,931	3,931	3,931
R-squared	0.002	0.002	0.002	0.087	0.087	0.088
Age FE			Yes	Yes	Yes	Yes
Market FE				Yes	Yes	Yes
Org Size FE					Yes	Yes
Pat Count FE						Yes

Notes: This table shows regression estimates for the relationship between exposure to knowledge growth and change in specialization for all oncologists. Specialization is measured using a Herfindahl—-Hirschman Index (HHI) over types of cancer managed. The measure of exposure to knowledge growth is recentered to control for variation in physician scope and field expected knowledge growth. Column 1 shows the coefficient estimate for the univariate regression. The regression standard errors are robust. Column 2 restricts the sample to physicians where we observe age, hospital referral region (HRR) and the number of other oncologists at their primary organization. Columns 3, 4, and 5 add fixed effects for the physicians age group, hospital referral region (HRR) and the number of other oncologists at their primary organization, respectively. In the recentered measure of exposure to knowledge growth, predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. Raw exposure to knowledge growth is calculated as the total increase in the count of guideline citations over 2009 to 2014 for the types of cancers a physician treats in 2008. Predicted exposure is estimated based on the cancer types a physician treats in 2008 and the predicted, rather than realized, growth in knowledge over the next five years. The predicted knowledge growth for each type of cancer is estimated based on a Poisson regression of the incidence of a cancer type in Medicare in 2008, an indicator for if the cancer type has well-established surrogate endpoints, and the interaction of the two. All physicians in the sample are hematologist-oncologists, 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.19: Median and 90th percentile of specialization by organization size in 2008

	(1) Median HHI Academic	(2) 90th Pct Academic	(3) Median HHI Non-academic	(4) 90th Pct HHI Non-academic
Organization Size	42.18	52.89	1.574	10.66
	(8.602)	(8.927)	(2.096)	(5.719)
Observations	143	116	1,428	386
R-squared	0.129	0.171	0.000	0.012
Minimum Org Size	1	5	1	5
Standard Errors	Robust	Robust	Robust	Robust

Notes: Table A.19 shows regressions of oncologist specialization on organization size by organization type in 2008. In Column 1, the outcome is the 50th percentile 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 other oncologists in the sample. In Column 3, the outcome is the 50th percentile 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 other oncologists in the sample. Organization size is estimated as the number of oncologists with at least 25 patients at the organization's tax ID in the 2008 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. See Section 3.3 and Appendix B.2 for discussion of the limitations of variables from the MD-PPAS. Oncologist specialization is measured using a Herfindahl-Hirschman Index (HHI) over the types of cancer that the oncologist evaluated and managed in that year. 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 data.

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 and examples about the construction of the indicator for academic status for physicians.

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 created 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-Hodgkins'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 single category of Non-Hodgkins'Lymphoma in order to have a stable grouping over time.

I also group screening and diagnosis NCCN Guidelines with treatment NCCN Guidelines where relevant when counting guidelines lengths and citation counts. For most cancers, all types of recommendations are in a single guideline. However, breast and prostate cancer have separate NCCN Guidelines for screening and diagnosis than for treatment. I group all types of NCCN Guidelines for breast cancer and all types of NCCN Guidelines for prostate cancer together. This grouping ensure 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 switches 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 as NCCN Guidelines are also organized based on the anatomical origin of tumors.

In addition, I am reassured that anatomic origin is the appropriate high-level grouping because NCCN Guidelines are also organized anatomically. This observation 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 oncologists", or "thoracic oncologist".

B.2 Labeling Academic Physicians

This section discusses how I labeled 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. These limitations arise from the inherent challenges of binarizing the diversity of organizational arrangements in oncology and the limitations of the data. I detail the process below, which broadly emulates the approach from Welch and Bindman (2016).

I use the Medicare Data on Provider Practice and Specialty (MD-PPAS) from 2008 to proxy academic status. The data contains the name associated with the tax ID number from which the physician billed the most services in 2008. I assign academic status based on the name of this tax ID number.

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 it 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 matches the state and name of record for an academic institution's physician group. I also consider a tax ID to be academic if the institution named was affiliated 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 2008 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 Health System is labeled as academic since its name contains the phrase "University" and is currently affiliated with the University of Chicago. 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").

C Measurement Error in Physician Specialization

This section discusses measurement error in the main outcome variable, physician specialization. Estimated physician specialization has measurement error with a positive expected value. Fortunately, the recentering procedure used in the main estimates removes the correlation of this measurement error with the treatment variable (exposure to knowledge growth).

As an additional set of robustness checks, I replicate the key descriptive and regression results with a measure of specialization ($\hat{HHI}_i$) and recentered exposure to knowledge growth ($\Delta\tilde{k}_i$) calculated over 25 random patients per physician instead of all observed patients. This approach inflates measurement error for all physicians who see over 25 patients, but particularly so for high-volume physicians, equalizing measurement error across physicians. With this change, the time trends in specialization (Appendix Figure A.4a) are somewhat attenuated but show a similar rise in top specialization as Figure 3. With the additional measurement error, the estimates of the effect of knowledge growth on specialization (Appendix Tables A.10, A.11, and A.12) are noisy and unstable.

C.1 Measurement Error in Exposure to Knowledge Growth

Recall that I do not observe all patients nationally but rather a 20% sample of patients enrolled in Medicare Part B. Not observing all patients may lead to measurement error in exposure to knowledge growth, both unadjusted (Δk_i) and recentered ($\Delta\tilde{k}_i$). A key concern is that the degree of finite-sample measurement error is correlated with a physician's patient volume (i.e. how finite the sample is). Fortunately, patient volume is balanced across the recentered measure of exposure to knowledge growth, $\Delta\tilde{k}_i$ (recall Table 5), which implies that the degree of finite-sample measurement error is uncorrelated with the main treatment variable.

C.2 Measurement Error in Physician Specialization

Measurement error occurs because I observe only a subset of the patients a physician treats in a year. Recall that physician specialization is measured by a Herfindahl–Hirschman Index (HHI) over the shares of time that an oncologist spends managing each type of cancer: $HHI_i = 10,000 \sum_C \left(\frac{T_{ic}}{T_i}\right)^2$. The estimated value of HHI_i will be biased upwards on average in a finite subsample of patients. To illustrate, suppose physician i has $HHI_i = 2500$ when calculated over all patients. Suppose I observe only one patient; mechanically, I will estimate that $\hat{HHI}_i = 10000$. Suppose, I observe only two patients; mechanically, the minimum HHI I could estimate is $\hat{HHI}_i = 5000$. As the number of patients increases, the expected value of $\hat{HHI}_i$ approaches HHI_i from above. Recall that I observe only a 20% sample of Medicare patients treated by a physician rather than all patients. As such, I estimate HHI_i with finite sample measurement error. This measurement error will be larger for low-volume physicians and decline as patient volume increases.

Figure A.8 demonstrates that $\hat{HHI}_i$ declines with volume and that some of the decline is attributable to decreases in measurement error. I plot mean $\hat{HHI}_i$ by patient volume in 1999, 2009, and 2019. $\hat{HHI}_i$ declines with volume in all three years. How much of this decline comes from reductions in measurement error, and how much reflects real specialization among low-volume physicians? For example, the observed relationship could reflect real specialization differences if low-volume physicians tend to be academic sub-specialists who spend half their time treating one type of cancer and half their time researching that cancer.

To distinguish between measurement error and real differences in specialization, I also plot $\hat{HHI}_i$ estimated with 11 and 25 random patients per doctor.

Using a random subset of patients inflates the finite sample measurement error for physicians who see more than that number of patients a year so that $\hat{HHI}_i$ is measured with an equal degree of error regardless of observed physician volume.

There are several key findings.

First, small sample noise is not mean zero. The average HHI calculated with 11 patients

is consistently higher than the average HHI calculated with 25 patients. This HHI is again higher than average HHI calculated with all observed patients. Second, using at least 25 patients removes the majority of the mean difference in HHI between using 11 patients and all observed patients. Third, the relationship between volume and HHI changes over time. In 1999, low-volume and high-volume doctors have similar HHIs once I measure HHI with 25 patients for everyone. This pattern suggests that the decline in HHI with volume is largely a product of volume-dependent measurement error. However, in 2009 and 2019, HHI does decline with volume, even after equalizing volume-dependent measurement error. This pattern implies that specialists are genuinely lower-volume physicians in more recent years.

Next, I consider the impact of volume-dependent measurement error on my main identification strategy. A key concern for estimation is that patient volume, and therefore measurement error in specialization, is correlated with the treatment variable, creating omitted variables bias in the main regression estimates. Table 5 shows that the raw measure of exposure to knowledge growth is correlated with patient volume and, therefore, likely correlated with measurement error. Fortunately, the recentering procedure removes the correlation between exposure to knowledge growth and physician volume and, therefore, measurement error. If so, non-zero measurement error should not contribute to omitted variables bias in the main specifications. In addition, the main regression results are robust to adding fixed effects for patient volume (Tables 6 and 7), suggesting that patient volume, and therefore volume-dependent measurement error, is not a source of omitted variables bias.

C.3 Effect of equalizing volume-dependent measurement error in main results

Next, I investigate the impact of volume-dependent measurement error on the main findings. To do so, I replicate the key descriptive and regression results with a measure of specialization ($\hat{HHI}_i$) and recentered exposure to knowledge growth ($\Delta\tilde{k}_i$) calculated over 25 random patients per physician instead of all observed patients. This approach inflates measurement

error for high-volume physicians and thus equalizes measurement error across low- and high-volume physicians.

First, I consider the impact on the observed trends over time in specialization. Figure A.4a plots the distribution of HHI over time, similar to figure 3, but calculates HHI using only 25 random patients per physician rather than all patients. The overall patterns are essentially unchanged. The 10th through 75th percentiles of specializations are largely flat or declining. The 90th percentile of specialization increases substantially over time, consistent with rising top specialization.

Finally, I replicate the main regressions using only 25 patients per physician. Table A.10 shows the regression of exposure to knowledge growth on specialization for ex-ante specialized physicians. Column 1 shows the regression with exposure to knowledge growth calculated using only 25 random patients. Column 2 shows the regression with $H\hat{H}I_i$ calculated using only 25 random patients. Column 3 shows the regression with both exposure to knowledge growth and $H\hat{H}I_i$ calculated using only 25 random patients. The estimated effects are similar to the main results, with the same direction and positive magnitude. Table A.11 shows the same results for ex-ante generalist physicians. The estimates are all indistinguishable from zero, as in the main results. The similar estimates suggest that measurement error does not substantively impact or change the central findings.