Knowledge Growth and Specialization: Evidence from Oncologists *

Maya Lozinski
University of Chicago

March 2025

Abstract

How does growth in general knowledge impact worker specialization and subsequent technology adoption? This question is studied empirically in medical oncology, which has had explosive knowledge growth as measured by changes in historical cancer treatment guidelines. A small subset of initially specialized oncologists become even more specialized in response to knowledge growth. However, most oncologists remain general and do not become more specialized over time or in response to knowledge growth. These differences reflect differing constraints: specialized oncologists appear highly constrained by scale, with specialization rising steeply with organization size in academic settings and with market size. In contrast, generalists face coordination costs that are not alleviated by organization or market size. Specialization increasingly influences patient care. Over time, specialists prescribe newer anti-cancer drugs and have more experience with the drugs they prescribe than generalists. The results suggest that some knowledge users may respond to knowledge growth not by specializing but by falling behind the frontier.

Keywords: specialization, knowledge growth, healthcare, oncology, scale economies, technology adoption

JEL codes: I1, J2, L2, O3

*Thank you to Zarek Brot-Goldberg, Joshua Gottlieb, David Meltzer, Dan Black, Tamara Konetzka, Pauline Mourot, Keith Ericson and Jonathan Skinner for excellent feedback and comments. Thank you to Stephen Pardy for technical assistance digitizing guideline citations. Thanks to the NIH Medical Scientist Training Program (T32GM007281) and the NIA Program in Medicine, the Social Sciences, and Aging (T32AG051146) for support. Author's email address: mayalozinski@uchicago.edu.

1 Introduction

Many fields, such as computer science, molecular biology, and medicine, have a rapidly growing knowledge base and proliferation of subfields. Expert workers may need to work at the frontier of this knowledge base. However, as knowledge grows, staying up to date across many subfields can become costlier. Specialization, or narrowing one's scope, is one possible worker response to this growth, as this can reduce the cost of maintaining expertise. However, workers may also face constraints that prevent them from specializing. In this situation, expert workers can either bear the growing cost of staying at the frontier across multiple areas or fall behind it.

I study this question in the context of medical oncology, where I develop new measures of knowledge growth and evaluate how expert workers respond to growing knowledge. Medical oncologists are physicians who treat cancer patients with systemic therapies, like chemotherapy and immunotherapy (as opposed to surgery or radiation). They are knowledge workers with extremely high human capital; they must complete college, four years of medical school, three years of clinical training in general internal medicine, and at least two more years of training in oncology to practice in the United States. Their work largely involves cognitive and communication tasks. They are also primarily users of knowledge rather than producers of knowledge (Taylor, 2004; Jeffe and Andriole, 2018; Jain et al., 2019; Schlafly and Sebro, 2022).

Despite large upfront investments in human capital, oncologists must continue to invest to stay at the frontier. Cancer is a condition with a vast demand for treatment — the lifetime risk of developing cancer is 40% (American Cancer Society, 2024)[1] —, leading to rapid innovation and knowledge growth. In oncology, "incorporating this torrential evolution of knowledge is an impossible, Sisyphean task" (Wallis, 2023). As one oncologist told this author, "Some fields of cancer treatment are as complex now as the whole of oncology was 20

[1]Fortunately, treatment is often life-extending or curative; the lifetime risk of dying of cancer is much lower at 20%.

years ago." As such, this setting enables the study of the effects of rapid knowledge growth in a context with material welfare implications.

To estimate this relationship empirically, I create measures of knowledge growth based on changes in cancer treatment and screening guidelines. The NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines ®) are the leading comprehensive cancer treatment guidelines in the United States. Since 2001, they have constituted a "complete library" of guidelines regarding cancer management (NCCN, 2023*b*), Changes in the length of these guidelines serve as proxies of changes in the size of the clinically relevant knowledge base by cancer type. As a robustness check, the changes in the number of references in the guidelines are also used to measure knowledge growth. A key advantage of guideline-based measures is that they capture the volume of knowledge directly relevant to oncologists' everyday practice.

Examining guidelines confirms that oncology knowledge has grown dramatically. The total length of NCCN Guidelines for screening and treatment has grown from 1,129 pages in 2002 to 5,825 pages in 2020. This growth in guideline length appears closely tied to growth in the relevant scientific literature. This growth also occurred in a period with large increases in the number of available therapies: the total number of approved cancer drugs rose from 90 in 2002 to 243 in 2020.

Oncologist specialization has varied patterns over this time frame. Oncologist specialization is measured each year in a 21-year panel of Medicare claims data. Specialized oncologists have become even more specialized, with rising extremes of specialization over time. However, surprisingly, the typical oncologist has not become more specialized. Median specialization declines slightly despite substantial growth in medical knowledge over this period.

Next, I estimate oncologists' exposure to knowledge growth to test how it influences specialization. Oncologists' exposure to knowledge growth is proxied as the total increase in guideline length in the fields of cancer in which they work in a baseline year. If a doctor treats both kidney and lung cancer, and each set of guidelines grew 20 pages longer, they were exposed to 40 pages of knowledge growth. This measure combines two sources of variation:

variation in the sets of cancer subfields that physicians work in and variation in the rates of knowledge growth across those subfields.

This measure may suffer from omitted variables bias since cancer subfields are not chosen randomly. To purge potential bias, I recenter the measure using the general approach developed by Borusyak and Hull (2023). The recentering procedure removes variation in exposure to knowledge growth from the number of types of cancer a physician treats at baseline and from predictable knowledge growth, predicted by the cancer's prevalence and whether it has surrogate endpoints. The remaining variation is attributed to "unpredictable" knowledge shocks, e.g., the randomness in clinical trials' success or effectiveness of experimental molecular targets. Reassuringly, recentering removes the correlation between the measure and observable omitted variables such as market size and patient volume.

This measure is then used to estimate how exposure to knowledge growth influences oncologist specialization. Oncologists who start specialized become even more specialized in response to knowledge growth. For every 100-page increase in relevant guidelines, these oncologists increase their specialization by approximately 196 HHI points, a 4% increase in HHI for a 13% increase in relevant guidelines. In contrast, oncologists who start as general oncologists have a muted specialization response to knowledge growth. They see slightly fewer unique cancers and experience only marginally significant increases in their HHI.

Stylized facts suggest different constraints to specialization for specialists and generalists that may contribute to these different responses. Specialized oncologists appear to be highly constrained by scale. Specialization rises steeply with organization size in academics, where specialists are disproportionately concentrated. In addition, the top degree of specialization within a market rises with market size, with a gradient that has more than doubled over time. As a result, geographic inequality in top specialization and local access to highly specialized care have grown over time. This finding also suggests that market-level economies of scale have grown for highly specialized medical care over time.

In contrast, the typical oncologist faces coordination costs not alleviated by large orga-

nizations or markets. Most general oncologists work in non-academic organizations. Within these organizations, specialization does not increase with size, implying that organizational scale does not facilitate specialization outside of academics. Market size also does not strongly facilitate specialization for the typical oncologist. The median degree of specialization in a market is close to flat with respect to market size. These patterns suggest that general oncologists face some combination of low returns or high costs to specialization, regardless of market or organization size.

Given that specialization varies widely, does it impact patient care? Examining prescribing patterns suggests that specialization increasingly influences patient care. Over time, specialists use fewer unique cancer drugs than generalists, giving them relatively more experience with each drug they prescribe. In addition, generalists have, over time, fallen behind specialists in terms of adopting new cancer drugs, particularly breakthrough cancer drugs. These differences in practice patterns have grown substantially over time. They also imply that specialization has returns to quality and that specialists increasingly provide higher-quality cancer care than generalists.

Overall, the findings contribute to several discussions in economics and healthcare. First, the work contributes a measure of knowledge growth that captures changes in the currently relevant knowledge in cancer care. Recent medical literature has documented that NCCN Guidelines for four common cancers have grown longer and include more citations over time Kann et al. (2020). I expand on this by quantifying the growth across all types of cancer screening and treatment and documenting the degree of turnover in citations over time. In the economics literature, Glied and Lleras-Muney (2008) measures technological change in healthcare using stocks and flows of drug approvals. Drug approval-based approaches face some limitations that using guidelines overcomes. Specifically, guidelines include the full diversity of innovation and knowledge growth, which in oncology encompasses not only new drug approvals but also off-label drug use, new drug combinations, biomarkers, drug delivery strategies, and more. In addition, guidelines exclude information that is no longer pertinent,

such as information about old drugs that are no longer used. Patents have also been used to study innovation and knowledge growth by an extensive literature (Alcacer and Gittelman, 2006; Roach and Cohen, 2013). A key advantage of clinical guidelines over patent-based measures in this setting is that they enable the measurement of the volume of knowledge currently relevant to doctors' day-to-day work rather than the historical depth. For example, the chemotherapy drug sirolimus has a fascinating history that includes many patents and papers and is long enough to fill a book (Duffin, 2019). However, this history is largely irrelevant to oncologists' day-to-day clinical work and is not included in guideline-based measures of relevant knowledge.

Notably, this work addresses a similar question to Jones (2009) about how knowledge growth impacts specialization but finds differing results. Jones (2009) documents empirically that inventors face a growing historical "depth of knowledge" (as measured by patent citations) and that this leads them to specialize more. Similarly, a few highly specialized oncologists become even more specialized in response to knowledge growth in this setting. However, most oncologists are generalists and do not become more specialized over time or in response to knowledge growth. One potential reason for these contrasting results is differences between doctors and inventors. Inventors are creators of knowledge. In contrast, most oncologists, particularly general oncologists, are primarily users of knowledge. Users may bear the burden of knowledge more lightly than creators, as using knowledge may be less difficult than creating it. In addition, as I document, users of knowledge can fall behind the frontier when keeping up becomes costlier, an option that may not be available to inventors. In medicine and likely other fields, many more workers use knowledge than create it. As such, knowledge growth may have different consequences for this common type of knowledge worker than previous evidence suggests.

This work also contributes to an extensive literature on worker specialization. Similar to Baumgardner (1988) and Baicker and Chandra (2004), I find that physicians are more specialized in larger markets and larger firms even within a narrowly defined specialty, and

expand on this by showing these differences are growing over time. I also replicate the finding from Atalay, Sotelo, and Tannenbaum (2024) that the task content of work varies meaningfully by market size within narrowly defined occupations. I find evidence that physician specialization has returns to quality, a notable contrast to (Baicker and Chandra, 2004). The period examined may contribute to these differences, as realized returns to specialization are growing over time.

Additionally, this work contributes to an extensive empirical literature on the consequences of technological change in healthcare. Much of the literature focuses on single technologies with varied economic impacts on providers and patients (Skinner, 2011; Chandra, Malenka, and Skinner, 2014; Skinner and Staiger, 2015; Arrow, Bilir, and Sorensen, 2020). This work differs in considering the effect of changes in the overall corpus of technology. Evidence suggests that cumulative technological change in healthcare (as measured by the stock of drug approvals) increases educational inequality in health outcomes (Glied and Lleras-Muney, 2008). I build on this by examining how knowledge growth impacts healthcare workers and geographic inequality in care provision. Dingel et al. (2023) find large market-level economies of scale in healthcare. I build on this by providing evidence that economies of scale in healthcare appear to have grown over time, with knowledge growth as an important mechanism.

2 Conceptual Framework

This section presents a theoretical framework for understanding how oncologists respond to knowledge growth. Medical oncologists focus on treating cancer with drugs rather than radiation or surgery and can specialize both horizontally and vertically. While classic economic theory suggests that market size limits specialization, coordination costs may also significantly constrain the division of labor. Knowledge acquisition costs and the returns to specialization may increase when guidelines grow more complex, potentially driving further

specialization. These dynamics suggest that knowledge growth may exacerbate differences in specialization between large markets, which can support highly specialized oncologists, and small markets, which may only support generalists, increasing geographic inequality in specialized cancer care.

2.1 Medical Oncology

This research focuses on medical oncologists, who are differentiated from other physicians along two key dimensions: disease space and treatment modality. In disease space, medical oncologists differ from other doctors in their focus on cancer care. In treatment modality space, they are experts in using drugs, in contrast to radiation or surgery. Their residency training pathways are almost entirely different from those of doctors who train in radiation or surgery for cancer[2]. The patient problems are also often distinct in medical oncology relative to surgical or radiation oncology, with medical consensus in many, but not all, cases about which aspects of cancer care should be managed with surgery, drugs, or radiation.

2.2 Horizontal specialization

The costs and benefits of specialization can be understood through classic models of horizontal specialization and knowledge growth. Oncologists can horizontally differentiate and specialize, making them more productive at specific tasks. For example, one oncologist may focus on breast cancer while another focuses on lung cancer. This specialization reduces duplicative learning costs, as each physician only needs to stay current with developments in their specific domain. It also enables learning by doing. The breast cancer specialist will frequently prescribe breast cancer drugs, developing expertise in managing the side effects of these specific drugs, potentially leading to better patient outcomes.

In classic economic theory, specialization is highly productive and therefore limited largely

[2]For example, surgical oncologists first train in surgery for a particular region or structure in the body (e.g., the abdomen for general surgeons) and then in cancer surgery for that region or structure. Medical oncologists train in internal medicine and then in medical oncology.

by the extent of the market available to support it (Smith, 1776; Kim, 1989). As such, it should rise with market size.

However, Becker and Murphy (1992) observe this division of labor splits the broader task of cancer care for a population among more workers, leading to higher coordination costs. These coordination costs may take many forms, from literal communication costs to hold-up and incentive problems. For example, the marginal cancer patient may be highly profitable, incentivizing oncologists to keep all their patients for themselves rather than refer them to colleagues with more expertise. The optimal degree of specialization balances the benefits and costs of specialization. These coordination costs may constrain specialization, leading to a flat distribution of specialization by market size. Knowledge growth increases returns to specialization in a field, leading to more specialization in equilibrium.

2.3 Vertical specialization

Worker organization within medical oncology can also be understood through the lens of vertical knowledge hierarchies as characterized theoretically by Garicano (2000) and Garicano and Santos (2004). Formally, all oncologists are licensed to medically manage all types of cancer, placing them at the apex of a cancer care hierarchy that includes oncology nurses, technicians, care coordinators, and other non-physician health professionals. Informally, knowledge hierarchies still exist. In theory, general community oncologists can appropriately manage common cancer presentations, while complex cases should be referred to subspecialist care.

However, the referral market faces important frictions. Anecdotally, organizations are reluctant to refer patients out and give up those revenues due to the financial returns of providing cancer care. Due to anti-kickback laws, oncologists cannot be paid for their referrals and are, therefore, not compensated for the lost revenue from referring patients out. These incentives may lead to under-referrral in equilibrium, in line with the predictions from Garicano and Santos (2004) that in a spot market for referrals, particularly one with a price

of zero, the equilibrium involves under-referral relative to the social optimum. Garicano and Santos (2004) suggest this under-referral problem can be mitigated through ex-ante agreements about referrals and compensation, often facilitated through firms, partnerships, or other business relationships.

Guidelines can be seen as a technology that impacts the costs of knowledge acquisition. When introduced, guidelines reduce the costs of knowledge acquisition relative to no guidelines. However, as existing guidelines grow more complex, the cost of acquiring knowledge from them also increases. Guidelines serve as abbreviated references that reflect the extent of outside knowledge. The longer the guidelines are, the greater the outside knowledge required to treat cancer comprehensively. As knowledge acquisition costs rise, workers should become more specialized and create deeper vertical hierarchies (Garicano, 2000; Garicano and Santos, 2004).

2.4 Knowledge Growth and Market Size

A natural corollary to the predictions above is that knowledge growth may also increase differences in specialization by market size. A breast cancer specialist can only exist if there are enough breast cancer patients to support a specialist and thus may only exist in large healthcare markets. As breast cancer care grows more complex, these specialists may further focus on particular subtypes of breast cancer, leading to growing specialization but only in large markets. Small markets may continue to lack the population to support anything but a general oncologist, leading them to lag further behind the large markets in terms of specialization and any associated returns.

3 Data and Measurement

The empirical strategy combines multiple data sources to measure knowledge growth and specialization in medical oncology over time. Knowledge growth is proxied using changes

in the page length and reference count in cancer treatment guidelines over time by cancer type. Medicare claims and provider data are used to measure physician characteristics, including how physician specialization evolves over time. These data are then combined to get physician-level estimates of exposure to knowledge growth over a five-year panel.

3.1 Cancer guidelines data

To measure knowledge growth, I assemble a new dataset about past cancer guidelines from the National Comprehensive Cancer Network (NCCN). The NCCN is a prominent "not-for-profit alliance of 33 leading cancer centers devoted to patient care, research, and education" (NCCN, 2023*a*). In 1996, they started to release treatment guidelines for common types of cancer. In 2001, they released the first "complete library" of cancer treatment guidelines (NCCN, 2023*b*). They are widely used by oncologists; per one survey, "96% of oncologists think that the NCCN Guidelines are important to use when making decisions about patient care" (McGivney, 2008).

The NCCN Guidelines contain recommendations for all stages of the cancer treatment process, including "diagnosis, imaging, drug therapies, radiation, [and] surgery" (McGivney, 2008). All of this information is relevant to medical oncology, as medical oncologists serve as the central node in a network of other specialists who may be involved in a patient's cancer care, such as surgical oncologists, radiation oncologists, and pathologists. As such, medical oncologists are trained in comprehensive cancer care management and must stay current with the high-level recommendations for all types of cancer treatments (Popescu et al., 2014). For brevity, I use the term "oncologist" to refer only to medical oncologists.

The NCCN Guidelines represent an up-to-date expert consensus on approaches to cancer treatment. Each NCCN Guideline is managed by a committee of several dozen experts selected from NCCN member institutions. The committee updates the guidelines at least once a year (and often multiple times a year) to reflect the newest evidence and drug approvals. A quorum of committee members must approve substantive guideline changes of committee

members. The vast majority of recommendations in the guidelines have been ratified by at least 85% of the committee members, representing uniform expert consensus in favor of the recommendation (NCCN, 2024).

Changes in NCCN Guidelines are used as a proxy for changes in the volume of relevant clinical knowledge. Page lengths and citation counts are measured for the first set of NCCN Guidelines for screening and treatment by cancer type by year. Changes in page lengths proxy for changes in the volume of clinically relevant knowledge for each cancer type. As a robustness check, changes in the number of citations over time are also used as a proxy of knowledge growth. The citations-based measure reflects changes in the volume of the relevant scientific literature underlying the guidelines. In addition, citations in the NCCN Guidelines for 2009 and 2014 are linked to published scholarly works in OpenAlex, a database of research publications (Priem, Piwowar, and Orr, 2022).

Importantly, the measured volume of relevant knowledge is distinct from how advanced the current technology is. For example, guidelines have many pages of recommendations about managing side effects from existing drugs. In theory, if new treatments are created that have minimal side effects and cure all cases, treatment could become simpler, and guidelines would get shorter. However, in practice, this rarely happens. While there have been breakthroughs in treatment, some patients inevitably have cancers that cannot be cured and which advance after first-line therapy. In these cases, the goal of care is often to slow the cancer's progression rather than cure it. Doctors will continue to explore treatment options until all potential treatments stop working and there are no more unused treatments. As a result, lines of therapy tend to accumulate over time, leading to longer guidelines.

Guideline growth has also occurred because of the medical system's growing ability to subtype cancers and target treatments. For example, lymphomas, a cancer of the lymphatic system, were historically classified as either "Hodgkin's", after the English physician who first described the disease, or the catch-all "non-Hodgkins" lymphoma (Armitage, 2010). As science has advanced, "non-Hodgkins" lymphoma has been appreciated to represent a grow-

ing number of biologically distinct cancer subtypes, each with different optimal treatment strategies (Jaffe, 2009).

Appendix Figures A.1 and A.2 include examples of guideline changes from 2009 to 2014. Appendix Figures A.1a and A.1b show the table of contents for non-Hodgkins lymphoma (NHL). Between 2009 and 2014, the length of the table of contents grows longer and includes new subtypes of NHL. Appendix Figures A.2a and A.2b show a flowchart with the diagnostic workup for lobular carcinoma in-situ, a subtype of breast cancer. Between 2009 and 2014, the flowchart grows more complex. It includes an additional diagnostic test, has more decision splits, and links to five additional flowcharts, up from three.

Several key assumptions are required for a change in page length to be a valid proxy for the change in the volume of knowledge. First, it assumes that each page represents an equal volume of knowledge. Reassuringly, the formatting of the NCCN Guidelines appears to have remained unchanged and standardized across cancer types over the main time period for the research design (2008-2014), suggesting that longer page lengths result from more content, not changes in formatting (e.g., increases in font size). In fact, changes in page lengths are likely to understate, not overstate, knowledge growth slightly. Appendix Figures A.1 and A.2 illustrate cases where a single page contains more content over time.

The measure also assumes that any relevant knowledge not included in the NCCN Guidelines is either exactly proportional to the volume of information the guideline pages contain or is constant across cancer types. For example, the NCCN Guidelines may not include all relevant information about a chemotherapy drug, but they need to contain a consistent share of the information about each drug. Similarly, the NCCN Guidelines are based on preexisting knowledge about general internal medicine and principles of cancer care. The volume of general medical knowledge can change over time as long as it impacts all cancer types equally.

Lastly, growth in page count is a net measure of knowledge growth. This measure assumes knowledge growth is primarily additive. For example, if a 20-page guideline grows to 30

pages, this change is measured as 10 pages of new knowledge. If, in fact, 10 pages of old knowledge have been dropped and 20 pages of new knowledge have been added, the net measure will understate the amount of knowledge growth. I assess this assumption empirically in section 4.

This proxy measure provides several key advantages over other measures of knowledge growth. For example, alternative approaches include using new journal publications, patents, or drug approvals by cancer type. However, many of these new papers or patents may be irrelevant to the day-to-day practice of medicine. In addition, older papers or drugs may remain relevant. For example, many (but not all) citations from the 2002 NCCN guidelines are also included in 2020, suggesting some medical evidence has a long shelf life.

3.2 Medicare claims data

I also use a 21-year panel of Medicare claims data to observe oncologists. Specifically, I use a 20% sample of Medicare Part B Claims from 1999-2019. Medicare provides federally sponsored health insurance coverage for disabled adults and most U.S. adults over sixty-five. This age group bears the highest cancer incidence in the U.S., making it the ideal population to study cancer care (for Disease Control and Prevention, 2019).

Using this data, I construct an annual panel of oncologists. I restrict the sample to physicians whose most commonly listed specialty in a year is Hematology, Medical Oncology, or Hematology-Oncology. I further restrict the sample to physicians who evaluated and managed at least 25 unique cancer patients in the data in a year. This restriction ensures patient privacy and that specialization measures are based on a minimum patient volume. I consider how patient volume impacts the measurement of specialization and the overall results extensively in Section 6.4 and Appendix C. For the main regression analysis, I subset to a balanced panel of oncologists from 2008 to 2014. The panel starts in 2008 because it is the first year with data on many physician characteristics, as described in Section 3.3. The panel ends in 2014, as this is the last full year to use ICD-9 diagnosis codes in the

claims data before the switch to ICD-10 in 2015. This switch from ICD-9 to ICD-10 leads to significant discontinues in the measures of physician specialization, described in more detail in Section 4.3.

Next, I measure which types of cancer an oncologist manages. The sample is restricted to outpatient evaluation and management claim lines billed by oncologists. These types of visits require high-level medical management and long-term planning, in contrast to, for example, administering chemotherapy that someone else prescribed or managing acute medical issues for hospitalized cancer patients). Each evaluation and management visit is assigned to a cancer type based on the cancer patient's primary cancer type[3]. One advantage of the context is that cancer diagnoses are standardized and determined independently of the oncologist's specialization, so there is no concern that the oncologist's specialization determines the type of cancer diagnosed.

I also measure the time each oncologist spends managing each type of cancer, which is then used to construct measures of physician specialization. Each evaluation and management code is mapped to work Relative Value Units (wRVUs), a measure used by Medicare to estimate the amount of physician work the code requires. For evaluation and management codes, wRVUs reflect the expected time required.

The measure of specialization directly measures horizontal specialization in disease space, e.g., the decision to focus on lung cancer rather than see all types of cancer. The measure does not directly reveal vertical knowledge hierarchies, which remain difficult to measure without more detailed clinical information about each cancer, such as stage, biomarkers, and history. However, in practice, horizontal and vertical specialization often go hand-in-

[3]Cancer patients are defined as patients with at least $250 in 1999 inflation-adjusted spending of Medicare Part B spending for cancer diagnoses. A patient's primary cancer is defined as the most specific type of cancer for which they have the most diagnosis codes in a year. For example, if a patient has ten diagnosis codes for lung cancer and thirty for breast cancer, they are considered to have breast cancer. Most (67% of patients) have one type of cancer while only 33% have more than one type in 2009. Of those with multiple types, the primary cancer type represents most of the diagnosis codes (80%). If a patient has diagnosis codes for both "Unspecified" cancer and breast cancer, they are considered to have breast cancer, no matter how many codes they have for "Unspecified" cancer. Patients with only unspecified or non-specific cancer types are excluded from the sample (1.9% of patients in 2009).

hand, with a narrow horizontal scope enabling vertical depth. For example, a lung cancer specialist is more likely to have expertise in treating complex metastatic lung cancer than a general oncologist treating many types of cancer.

3.3 Medicare Data on Provider Practice and Specialty

Physician characteristics are measured using Medicare Data on Provider Practice and Specialty (MD-PPAS) from 2008 to 2014. This dataset allows me to observe the physician's birth year and determine their age at the start of the panel period in 2008.

MD-PPAS also contains the top tax identification (ID) number from which each oncologist bills. This tax ID number represents the organization or individual to which Medicare payments are made. Multiple physicians can share the same primary tax ID number, which implies they are in the same organization or group practice (ResDAC, 2024). Organization size can be proxied as the number of physician billing from a given tax ID number. Limitations of using tax IDs as an organizational identifier include that healthcare organizations may have multiple tax ID numbers, particularly if providers within the organization may all bill to personal tax ID numbers. In this case, the number of physicians using a tax ID will understate the number of people at an organization. Tax ID names are also used to proxy academic status. See Appendix B.2 for more details on how this proxy is constructed. Overall, tax IDs provide imperfect but useful information about whether a physician works in a big organization and the nature of that organization.

3.4 Measuring relevant knowledge growth for each oncologist

The Medicare claims and NCCN Guidelines data are combined to measure the main treatment variable: the growth in relevant knowledge for each oncologist. For each oncologist i, denote the set of cancer types they manage outpatient in 2008 as C_i. This set of cancers is determined from 2008 Medicare claims data. Each cancer, c, has an increase in relevant guideline page count Δk_c over the following five years (2009 to 2014). I assume that the full

guidelines and changes in guidelines for cancer c are relevant to physicians who treat cancer c in the baseline year. With this assumption, I calculate oncologist i's exposure to knowledge growth as $\Delta k_i = \sum_{C_i} \Delta k_c$. As a robustness check in section 6.2, I raise the threshold for a physician to be considered exposed to a field by ignoring cancers where they manage only one or two patients a year.

For example, suppose oncologist i treats only lung and breast cancer in 2008. Also, suppose that over the following five years, the lung cancer guidelines increased by 50 pages ($\Delta k_{lung} = 50$) and the breast cancer guidelines increased by 100 pages ($\Delta k_{breast} = 100$). Then, oncologist i is exposed to 150 pages of knowledge growth over the following five years (2009 to 2014).

To map from cancer guidelines to cancer types in claims data, I make a set of custom diagnosis code and cancer guideline groupings. The custom groupings aggregate diagnosis codes to align with how cancers are categorized in NCCN Guidelines, which is not the case in standard diagnosis code groupings such as the Clinical Classification Software. The custom groupings also aggregate over some NCCN Guidelines to ensure stable categories over time. For example, the NCCN Guidelines for basal and squamous cell cancer (both types of skin cancer) were contained in a single guideline document until 2013 and then divided into two separate documents starting in 2014. I group these and other related NCCN Guidelines into the overarching category of skin cancer guidelines. More details about these groupings can be found in appendix section B.1.

The custom groupings are based on the cancer's anatomical origin, e.g., the lung. Historically, cancer has been classified by the anatomical location of the original tumor, regardless of where it metastasizes. For example, lung cancer that spreads to the brain is considered metastatic lung cancer, not brain cancer. This classification strategy is reflected in the top-level organization of NCCN Guidelines, specialization, and cancer treatments during this period. In recent years, some drug approvals have been given based on tumor biomarkers without regard to the anatomical origin of the cancer. However, the first such approval was

given in 2017 (Olivier, Haslam, and Prasad, 2021), the tail end of my time frame and outside the key panel period studied.

3.5 Measuring physician specialization

The main outcome of interest is oncologist specialization. Specialization is measured using a Herfindahl–Hirschman Index (HHI) over the shares of time that an oncologist spends managing each type of cancer. HHI is calculated as follows, where T_{ic} is time that oncologist i spends managing cancer c and T_i is the total time they spend managing cancer: $HHI_i = 10,000 \sum_C \left(\frac{T_{ic}}{T_i}\right)^2$.

Observe that HHI_i has a maximum value of 10,000, achieved if an oncologist spends all their time managing one type of cancer. If an oncologist divides their time equally between two types of cancer, then HHI_i is 5,000. The lower limit of HHI_i is 0, which is approached as an oncologist spends an infinitesimally small amount of time on an infinite number of cancer types.

This measure has several key advantages. First, HHI is continuous and based on an oncologist's work in a year. This feature is a benefit over using job titles, a common measure of specialization that does not reflect this subtle variation. Second, it is sensitive to subtle intensive margin changes in specialization. The measure will capture if an oncologist shifts towards focusing on one particular type of cancer, even if they do not entirely stop seeing long-standing patients for other types of cancers. As such, it can capture differences in specialization for the same worker over time and across workers with identical job titles and specialty fields.

4 Descriptive Patterns

This section describes the characteristics of oncologists in the sample and the patterns of knowledge growth and specialization over time. The volume of clinical knowledge and degree

of top specialization have both grown in oncology over time.

4.1 Knowledge growth over time

The volume of clinical knowledge, as proxied by guideline page lengths, has grown substantially over time. This growth in page length is driven by growth in the relevant scientific literature, which is largely additive. In addition, based on the underlying scientific literature, knowledge is almost entirely distinct across guidelines.

Analysis of historical guidelines reveals that cancer treatment guidelines have multiplied in length since 2002. Figure 1a displays guideline lengths between 2000 and 2020 for selected cancer types. Appendix Table A.1 lists the page count for all types of guidelines in the data in 2002, 2009, 2014, and 2020. The total length of guidelines has increased over five-fold in this period, from 1,129 pages in 2002 to 5,825 pages in 2020.

One factor likely contributing to this growth in guideline length is the explosion in the number of drugs available to treat cancer. Figure 1c displays the number of U.S. Food and Drug Administration-approved anti-cancer drugs over time[4]. The number of FDA-approved anti-cancer drugs increased from 90 in 2002 to 243 in 2020. Not only has the total number of drugs been increasing rapidly over time, but the rate at which new drugs are introduced has been accelerating.

Another factor in guideline growth is the growing scientific literature. Figure 1b shows the number of citations in these guidelines each year. Table 1 shows the close relationship between guideline length and the number of sources cited in the guidelines both in cross-section and over time. In the cross-section, every additional citation in 2009 is associated with an additional $\frac{1}{3}$ of a page of guidelines ($R^2 = 88\%$). In 2014, each additional citation is associated with an additional $\frac{1}{4}$ of a page of guidelines ($R^2 = 95\%$), suggesting either that citations are less informative or page contents have grown more dense. Over time, the growth in citation counts is tightly correlated with growth in page length from 2009 to 2014

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

($R^2 = 94\%$). Over time, every additional citation is associated with an additional $\frac{1}{5}$ of a page of guidelines. Guidelines from this period fit about 15 citations on a single page. Together, this implies an additional citation increases a guideline by $\frac{1}{15}$th of a page directly through the text of the citation and $\frac{2}{15}$th of a page through additional contents in the guideline.

Knowledge growth appears largely additive, though there is some turnover based on citations. The 2009 guidelines reference 4,013 unique sources. Of these, 2,202 sources (55%) are also cited again in 2014 and 1,811 (45%) are dropped from the guidelines. An additional 7,286 unique references are added to guideline citations by 2014, greatly outpacing the number dropped. As such, knowledge growth appears largely additive, motivating the use of net, rather than gross, measures of knowledge growth.

The references added to guidelines are largely new or recent literature. Of works added to guidelines between 2009 and 2014, 49% are new (published in or after 2009), 30% are recent (published in the 5 years prior to 2009), and 22% are older (published more than 5 years prior to 2009). Of scholarly works dropped, 48% are recent and 52% are older. Figure 2 displays the counts of papers added, remaining, and removed from guidelines over this period by year of paper publication.

The vast majority of references are citations to published research papers. Of 33,621 digitized citations in 2009 and 2014, 76% can be linked with high confidence to published scholarly works. The remaining 24% of citations include drug package labels, conference posters, professional society issue briefs, government reports, and papers that cannot be linked with high confidence.

Knowledge appears to overlap minimally across cancers when measured by the overlap in scientific literature cited across guidelines. In 2009 and 2014, 94% of cited scholarly works are cited by only one set of guidelines. As such, the scientific literature underlying each guideline is largely distinct across cancer types.

One limitation of the data is highlighted by Appendix Table A.1, which shows that a handful of cancer types lack guidelines in 2002, 2009, and 2014, despite NCCN consider-

ing their 2002 guidelines to be a "complete library" of clinical guidelines. One example is mesothelioma, a rare type of cancer that did not have guidelines until 2010. A 2010 news article about the release of the first mesothelioma guidelines notes that the authors found the guidelines "challenging to prepare" partly because of limited available evidence about how to treat the condition (Mulcahy, 2010). This example illustrates a key aspect of the guideline measure: zero pages of guidelines largely reflect near-zero evidence about how to treat that specific type of cancer. However, patients with mesothelioma were being treated in some fashion before 2010. Fortunately, the empirical strategy does not require the assumption that zero pages of guidelines reflect zero knowledge, only that more pages reflect more knowledge. However, the introduction of new guidelines may create a discontinuous jump in the number of pages of guidelines and overstate the relative knowledge growth. This measurement challenge impacts a very small share of patients in the sample (2.80 in 2009).

4.2 Oncologist characteristics

Next, I describe summary statistics for the balanced panel of physicians from 2008 to 2014. Table 2 shows summary statistics for key variables of interest for a balanced panel of 3,948 oncologists. On average, they see 61 unique cancer patients in the outpatient setting in the data in the baseline year of 2008 (with an imposed minimum of 25). These patients span 14 distinct types of cancer on average, but this can be as few as one type and as many as 25 types. The mean oncologist HHI in 2009 is 2,913, with a maximum of 10,000. Their average age at the beginning of the panel is 49 years old. On average, 19 oncologists work at each oncologist's organization, but this can be as low as one, indicating a solo practice, and as high as 219.

In 2009, the total length of NCCN Guidelines is 1,832 pages and over the following five years, this doubles to 3,534 pages. Based on cancers treated in 2008, the number of relevant guideline pages also doubles from 1,086 pages per oncologist to 2,118 pages on average.

4.3 Trends in oncologist specialization over time

Next, I characterize the distribution of oncologist specialization over time. Figure 3 shows the 10th, 25th, 50th, 75th, and 90th percentile of oncologist HHI between 1999 and 2019, with confidence intervals estimated by bootstrapping over the sample of oncologists. Figure 3 also plots the HHI of the overall distribution of patients in the sample, with confidence intervals calculated by bootstrapping over patients. Several striking trends stand out.

First, there is a sharp, discontinuous decline in HHI at all points in the distribution between 2014 and 2016, corresponding to the switch from ICD-9 to ICD-10 midway through 2015. This clear trend break motivates the decision to end the main balanced panel in 2014, which prevents coding changes from contaminating the estimates of change in HHI.

Second, the typical oncologist has not become more specialized over time. The 50th percentile of specialization and below are all flat or declining over this period. These oncologists appear to be "general" oncologists who see a roughly similar distribution of cancer patients as the field overall. Within each year, the distribution's 10th, 25th, and 50th percentiles all stay within a 1000 HHI point range that also includes the HHI of the overall distribution of patients in the sample. This lack of growth in specialization for the median oncologist is surprising, given the several-fold growth in knowledge and treatment options for cancer over these 21 years.

In contrast, subspecialized oncologists have become dramatically more specialized over time. The 90th percentile of oncologist HHI increases by 1,031 [800, 1,363] HHI points from 1999 to 2014 and another 298 [-13, 664] points from 2016 to 2019.

These results persist when using alternative approaches to measuring specialization. Appendix Figure A.3a shows the distribution of specialization over time using 25 random patients per oncologist rather than all patients. Using 25 patients removes any variation in patient volume that might contribute to the observed trends. It also increases the measurement error with which specialization is measured. Appendix Figure A.3b shows the distribution of specialization using shares of unique patient counts rather than shares of

time by cancer type. Using patient counts reduces the contribution of some factors, such as variation in visit frequency across doctors. However, it increases noise from others, such as doctors seeing one-off patients while covering for a colleague on vacation. In both figures, the observed trends are similar. The 90th percentile of specialization rises significantly between 1999 and 2014, and variably over the shorter window of 2016 to 2019. The 50th percentile and below all flat or declining from 1999 to 2014 and 2016 to 2019.

These contrasting time trends suggest that "generalist" and "specialist" oncologists may respond to knowledge growth differently. Specialists may respond. Generalists likely do not, given that they have not become more specialized in a period with substantial knowledge growth. For the main estimation, the sample is split into generalist and specialist oncologists to test if they respond differently to knowledge growth. Of note, although I refer to specialist and generalist oncologists, all oncologists are considered medical subspecialists, highlighting the extremes of specialization studied here.

4.4 Differences between more and less specialized oncologists

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

Substantial differences are present between these two groups in the baseline year for the panel. In 2008, specialists have an average HHI of 4,858, double that of the generalists, who have an HHI of 2,262 on average. Specialists are far more likely to be academic physicians. 24% of specialists appear to be academic physicians, compared to only 6% of generalists. Specialists are also more concentrated in large markets: 60% of them are in the largest 20% of markets as compared to 47% of general oncologists. In addition, they work in larger organizations: On average, specialists work in places with 21 other oncologists, compared to 17 for generalists. These empirical differences are consistent with large firms, large markets,

and academic institutions all facilitating specialization. Notably, age does not differ meaningfully between specialists and generalists (50 versus 49 years old at the beginning of the panel). Anecdotally, this lack of age difference is consistent with how specialization decisions are currently made in medical oncology. Many oncologists decide whether to become generalists or specialists early in their careers while in fellowship training, rather than starting as generalists and then specializing or vice versa.

Despite the age similarities, specialists and generalists are otherwise observably distinct groups. The two groups likely also differ in unobserved ways that lead to their initial differences in specialization. These differences motivate the estimation of the response to knowledge growth separately for the two groups and the use of an identification strategy that removes variation in exposure to knowledge growth that stems from the breadth of practice.

5 Identification

The following section discusses the identification strategy used to test if knowledge growth increases specialization. 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 suffer from omitted variables bias. Specifically, exposure to knowledge shocks may be non-random and correlated with changes in specialization from other causes. This potential bias is removed by "recentering" the measure to account for expected, or non-random, exposure to knowledge growth using the general approach from Borusyak and Hull (2023).

5.1 Estimating Equation

First, I discuss the main estimating equation. Denote physician i's change in specialization, as measured by HHI, between 2009 and 2014 as ΔHHI_i. Denote their exposure to knowledge growth as Δk_i. Recall that exposure to knowledge growth is calculated as $\Delta k_i = \sum_{C_i} \Delta k_c$,

where C_i is the set of cancers that physician i treats in 2008 and Δk_c is the change in page length for each type of cancer between 2009 and 2014. Also, denote the vector of other physician characteristics as X_i. These include physician age, market, patient volume, and organization size at the beginning of the panel. The main estimating equation is as follows:

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

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

This estimating equation has several key features. First, the outcome uses within physician variation in specialization over time. This within-physician design requires a balanced panel. It also allows for specialization to respond immediately to knowledge growth. If the specialization response is highly lagged, that may attenuate the estimates.

The main coefficient of interest is β, the effect of knowledge growth on specialization. β is estimated using variation in exposure to knowledge growth (Δk_i). Variation in exposure to knowledge growth (Δk_i) comes from two sources. The first source is that physicians work in different sets of cancer subfields. For example, some treat breast cancer, some treat leukemia, and some treat both. The second source of variation is that different cancer subfields have different amounts of knowledge growth. For example, leukemia may have more knowledge growth than breast cancer, so physicians who treat only leukemia are more exposed to knowledge growth than those who treat only breast cancer.

The intercept term (α) absorbs any global shocks to knowledge. Specifically, it will include the impact of knowledge growth that impacts all fields of cancer or the general practice of medicine. As such, the main coefficient of interest (β) captures the effect of changes in different cancer subfields rather than global shocks.

The intercept term (α) also absorbs changes in the composition of oncology patients. The HHI of the overall distribution of patients in the sample falls modestly from 2009 to 2014, with a change of -335 [-416, -252]. This level change is absorbed by (α).

5.2 Recentering instrument to correct for omitted variables bias

The measure of exposure to knowledge growth may suffer from omitted variables bias. To correct for potential bias and estimate causal coefficients, I recenter the measure using the general approach from Borusyak and Hull (2023).

In this context, omitted variables bias may arise because physicians treating more cancer types are mechanically exposed to more knowledge growth. For example, someone who treats only breast cancer will be less exposed to knowledge growth than someone who treats both breast cancer and leukemia. These two doctors may also differ in other important ways. For example, the breast cancer oncologist may be more likely to work at a large healthcare market like New York City because it has the patient volume to support a breast cancer-only oncologist. Oncologists in large markets may have different specialization trends than those in small markets, like Paducah, Kentucky, for reasons unrelated to knowledge growth. This variation in market size may be correlated with both exposure to knowledge growth (Δk_i) and change in specialization (ΔHHI_i), leading to omitted variables bias in the estimate of β.

Empirically, realized exposure (Δk_i) is calculated exactly from the data. Conceptually, it can also 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, $\mathbb{E}[\Delta k_i]$, is 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. This component is exogenous and uncorrelated with other factors that impact specialization except for knowledge growth.

The realized exposure (Δk_i) can be substituted with the "recentered", or residual, expo-

sure ($\Delta\tilde{k}_i$) for an unbiased and efficient estimate of β (Borusyak and Hull, 2023), as 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 contents of the error term, including omitted variables ($\Delta\tilde{k}_i \perp \epsilon_i$). The real-world content of this mathematical assumption depends on exactly how one forms the expectation of exposure to knowledge growth.

The expectation can be formulated as Assume there exists an expected knowledge growth for each type of cancer, $\mathbb{E}[\Delta k_c]$. Then a physician's expected exposure to knowledge growth ($\mathbb{E}[\Delta k_i]$) can be calculated as below:

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

The central challenge is to formulate a plausible estimate of expected knowledge growth by cancer type, $\mathbb{E}[\Delta k_c]$. The next section discusses iteratively more sophisticated approaches to estimating $\mathbb{E}[\Delta k_c]$ and the implications for the substance of the identification assumptions. The final approach is ultimately used in the main analysis.

The simplest assumption is that all cancer fields have no expected knowledge growth ($\mathbb{E}[\Delta k_c] = 0$). In this case, recentered exposure ($\Delta\tilde{k}_i$) equals realized exposure (Δk_i). This assumption is implicit in the original measure of exposure to knowledge growth (Δk_i). However, the assumption of no expected knowledge growth after 2009 is arguably implausible given historical trends in knowledge growth prior to 2009.

A more plausible assumption is that all cancer fields have the same non-zero level of expected knowledge growth. Doctors may expect knowledge to grow in all fields, but they do not anticipate which ones will grow more than others. The assumption that all cancers have the same level of expected knowledge growth can be operationalized as $\mathbb{E}[\Delta k_c] = \overline{\Delta k_c}$ where $\overline{\Delta k_c}$ is the mean realized page increase across cancer types. This estimate of $\mathbb{E}[\Delta k_c]$

can then be used to back out the recentered measure of exposure to knowledge growth ($\Delta\tilde{k}_i$) using Equations 1 and 3.

This approach compares doctors who treat the same number of cancers and is equivalent to controlling for the number of cancers an oncologist treats (Borusyak and Hull, 2023). Variation in the recentered measure arises because, holding cancer set size constant, oncologists vary in the composition of cancers they treat, and some cancers have more innovation than others. Critically, this approach assumes that cancer-level knowledge shocks are unpredictable. One concern is that cancer-level rates of knowledge growth may be predictable and, therefore, correlated with physician sorting. For example, an oncologist may anticipate that treatments for breast cancer will advance rapidly and thus choose to specialize in breast cancer. Addressing this potential bias motivates further refinement of the estimate of expected knowledge growth.

Next, I additionally allow expected knowledge growth ($\mathbb{E}[\Delta k_c]$) to vary with cancer characteristics. Acemoglu and Linn (2004) find that health conditions with more potential patients had more innovation, suggesting more common cancers may have more knowledge growth. Budish, Roin, and Williams (2015) find that cancers with high-quality surrogate endpoints have more innovation. I incorporate both factors into an estimate of expected knowledge growth, $\mathbb{E}[\Delta k_c]$.

The expected knowledge growth for a 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 type of cancer in 2008. Also, define e_c as an indicator for if a cancer type has high-quality surrogate endpoints per Budish, Roin, and Williams (2015). I include both of these in a Poisson maximum likelihood regression to predict the expected page count increase for each type of cancer, 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 equation controls for knowledge shocks shared across all cancer types (loaded on δ_0).

Table 4 shows the coefficient estimates for the above regression. The right-hand side variables have substantial predictive power, explaining 77% of the variation in growth in knowledge. I attribute the remaining variation to randomness and uncertainty in research and development. For example, some clinical trials and lines of drug development fail, while others succeed.

As above, I then use the resulting estimate of expected knowledge growth by cancer type ($E[\Delta k_c]$) to back out each physician's recentered exposure to knowledge growth ($\Delta\tilde{k}_i$). This recentered measure of exposure to knowledge growth ($\Delta\tilde{k}_i$) removes variation from both physician scope and cancer characteristics.

This approach makes several key assumptions. First, it assumes that patient prevalence, secondary endpoints, and shared knowledge shocks are the central factors correlated with highly predictable variation in innovation across cancer types. This assumption is empirically supported by the observation that these factors alone explain 77% of the variation in guideline page count increases. This approach also assumes that the residual variation in knowledge growth is unrelated to oncologists' specialization decisions except through the direct effect on specialization. This assumption is untestable but 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

The next section empirically tests the extent to which the recentering procedures purge measurable omitted variables bias in estimated exposure to knowledge growth, following the strategy outlined in Borusyak and Hull (2023). The original measure strongly correlates with observed omitted variables, suggesting recentering is necessary for unbiased estimation. Fortunately, recentering with progressively more sophisticated estimates of expected knowledge

growth ($\mathbb{E}[\Delta k_c]$) reduces this correlation until it is undetectable.

Column 1 in Table 5 shows the coefficient estimates for a regression of observed omitted variables on the unadjusted measure of knowledge growth. These observed omitted variables consist of key physician characteristics, including market size (measured as the population of the physician's hospital referral region or HRR), firm size (proxied as the number of other oncologists at their tax ID numbers), count of unique patients seen, and physician age all in the baseline year. As hypothesized, the unadjusted measure of knowledge growth is significantly correlated with all omitted variables. For example, an increase in the HRR population of 1 million is associated with a decrease in exposure to knowledge growth of -15 (2) pages. A regression using this unadjusted measure of exposure to knowledge growth will compare physicians in larger and smaller markets. These physicians may have different specialization trends for reasons unrelated to knowledge growth, biasing the estimate of β. Overall, the included omitted variables explain 33.5% of the variation in exposure to knowledge growth. These results substantiate concerns about omitted variables bias in the unadjusted measure of knowledge growth.

Columns 2, 3, and 4 in Table 5 successively remove additional variation from the measure of knowledge growth. Column 2 removes physician scope variation, Column 3 removes cancer prevalence variation, and Column 4 removes surrogate endpoint variation. In each column, observed omitted variables explain less and less variation in exposure to knowledge growth. By Column 4, the recentering has purged all significant correlation between exposure to knowledge growth and the observed omitted variables of patient volume, market size, organization size, and physician age. For example, an increase in the HRR population of 1 million is now associated with a decrease in exposure to knowledge growth of -0.3 (0.5) pages. Overall, the observed omitted variables now explain only 0.1% of the total variation in the fully recentered measure as compared with 33.5% of the variation in the original measure, a decrease of over 99%.

Appendix tables A.2 and A.3 show the same regressions separately for specialist and

generalist oncologists. The results are similar, with recentering reducing the explanatory power of observed omitted variables by 96% and over 99%, respectively.

Overall, the final recentering procedure removes the correlation between exposure to knowledge growth and observed omitted variables. The result suggests that the recentering procedure may also largely eliminate the correlation between exposure to knowledge growth and any observed omitted variables. This recentered measure is used in all subsequent analyses in Section 6.

6 Effect of Knowledge Growth on Specialization

Both knowledge and specialization have changed over time. Are these changes related? Ex-ante specialized oncologists become more specialized in response to knowledge growth. These changes appear to be driven by the intensive margin, with oncologists focusing more within the set of cancers they already see. Ex-ante general oncologists display a more muted response to knowledge growth, with small decreases on the extensive margin in the breadth of cancers they manage. The findings suggest that general oncologists drop cancers that represent a small share of their time, with minimal effects on overall HHI.

6.1 Effects on overall specialization

Figure 4a shows the univariate relationship between the exposure to knowledge growth and change in specialization separately for specialist oncologists. The relationship between exposure to knowledge growth and increased specialization is large and significant for ex-ante specialized oncologists. For every 100-page increase in relevant guidelines, these oncologists increase their specialization by 196 (80) HHI points [5]. Relative to baseline means, this represents a 4% increase in HHI for a 13% increase in relevant guidelines. Table 6 shows the

[5]Of note, mean reversion in HHI is observed, so HHI for oncologists in the top 25% of the distribution falls on average. A more precise statement is that HHI falls by 196 points less for every 100 pages of exposure to guidelines.

coefficient estimates after adding fixed effects for age bin, hospital referral region, patient volume, and organization size. The estimates increase slightly to 246 (91) HHI points per 100 pages with all fixed effects.

In contrast to specialists, general oncologists do not consistently become specialized in response to knowledge growth. Figure 4b shows the univariate relationship between the exposure to knowledge growth and change in specialization separately for ex-ante "general" oncologists. The estimated effect on HHI is more precise than the effects for specialists and statistically indistinguishable from zero at $\beta = 50$ (31)[6]. Table 7 shows the coefficient estimates after adding fixed effects. With all fixed effects, the coefficient estimate increases slightly to become marginally significant at the 95% level ($\beta = 66$ (33)). These results are evidence of a minimal response to knowledge growth in terms of overall specialization.

Appendix Figure A.4 and Table A.4 show the univariate and fixed effect regression estimates for all oncologists. The estimated coefficients consistently show small positive effects and are substantively unchanged by adding fixed effects.

6.2 Effects on Different Margins of Specialization

For specialists, the increases in specialization are driven by intensive margin changes in specialization. For generalists, the lack of overall specialization response hides a small decrease in scope on the extensive margin in response to knowledge growth.

Appendix Table A.6 shows that knowledge growth slightly reduces the count of cancers treated on the extensive margin for generalists. The effects are small, with an estimated reduction of 0.2 (0.1) unique cancers managed. Across specifications, the estimates rule out that exposure to 100 additional guideline pages reduces the number of cancer types managed by 0.5 (on a mean of 14) at the 95% level. The results are consistent with general oncologists dropping cancers that represent a small share of their time, leading to minimal effects on HHI.

[6]Mean reversion in HHI is also observed in this sample. HHI rises for oncologists in the bottom 75% of the distribution but does not rise more for oncologists exposed to more guideline growth.

In contrast, Appendix Table A.5 shows that, for specialists, exposure to knowledge growth does not reduce the number of unique cancers treated. Changes on the intensive margin of specialization must drive the observed changes in HHI. In practical terms, specialized oncologists focus their practice more on some types of cancer but do not entirely stop seeing other types.

Appendix Table A.7 shows the combined regression with all oncologists. The findings are consistent with averaging the effects for specialists and generalists.

The observed patterns for specialists may reflect specialist oncologists continuing to see established patients but limiting their new patients to those with cancers in a narrower area of specialization. To test this, specialization is measured with only new patient evaluation and management codes (Appendix Table A.8 and Appendix Table A.9). Unfortunately, this restriction shrinks the sample to one-quarter of its original size (even after relaxing the patient count minimum to 11) and inflates the standard errors, making the estimates uninformative.

One additional possibility is that physicians respond to knowledge growth only in their main area of specialization. For example, they may keep up to date with breast cancer if they treat many breast cancer patients but will not do so for one-off patients with other cancers. Recall that in the main estimates, a physician is considered exposed to knowledge growth if they are observed to manage at least one patient with that cancer type in the sample in the baseline year. To remove one-off patients, this threshold for exposure to a cancer type is raised to two patients and then three patients in Table A.8. The effect sizes are quite similar when raising the patient count threshold (β = 186 (87) for two patients, β = 311 (110) for three patients). These results also demonstrate that the relationship between knowledge growth and specialization for specialists is robust to raising the patient count threshold for exposure to knowledge growth. These analyses are repeated for generalist oncologists in Table A.9. Table A.9 finds no response to knowledge growth, with estimated coefficients of β 4 = 7 (32) with a threshold of two patients, β= 9 (35) for three patients.

6.3 Robustness to alternative measures of knowledge growth

I also consider the main regression results' robustness to an alternative measure of knowledge growth: the number of sources cited in each guideline. This alternative measure reflects the quantity of academic literature informing the guidelines. The results differ in magnitude, which is expected given that a source cited represents roughly $\frac{1}{5}$ of the volume of knowledge of a guideline page. Substantively, the results are quite similar. More cited sources lead specialists to become significantly more specialized (Appendix Table A.13). An additional 100 citations in guidelines cause specialists to increase their specialization by an estimated 65 (18) HHI points in the univariate regression. More cited sources cause smaller but significant increases in generalists' degree of specialization (Appendix Table A.14) An additional 100 citations in guidelines cause generalists to increase their specialization by an estimated 16 (7) HHI points. In aggregate, more cited sources has a positive estimated effect on specialization (Appendix Table A.15)

6.4 Robustness to volume-dependent measurement error

Lastly, I consider how measurement error in the measure of specialization ($\hat{HHI}_i$) impacts the main findings. Appendix C addresses this issue in greater detail. The key findings are summarized here.

Recall that not all patients nationally are observed, but rather a 20% sample of patients enrolled in Medicare Part B. Not observing all patients may lead to measurement error in both the outcome variable of specialization ($\hat{HHI}_i$) and the treatment variable, recentered exposure to knowledge growth ($\Delta\tilde{k}_i$). In particular, these quantities may be measured with systematically more measurement error for physicians with lower patient volumes. Fortunately, the main treatment variable, recentered exposure to knowledge growth ($\Delta\tilde{k}_i$), is uncorrelated with patient volumes, implying it is uncorrelated with volume-dependent measurement error, as shown in Table 5. However, the main outcome variable, specialization, is measured with patient volume-dependent measurement error, shown in Appendix Fig-

ure A.10. However, the main regression results are essentially unchanged after 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 bias in estimation.

As an additional set of robustness checks, I replicate the key descriptive and regression results with a measure of specialization ($H\hat{H}I_i$) 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, particularly for high-volume physicians, and equalizes measurement error across physicians. With this change, the time trends in specialization (Appendix Figure A.3a) 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.

7 Constraints to Specialization

Why do specialist and generalist oncologists appear to respond differently to knowledge growth? Stylized facts suggest differing constraints to specialization. Top specialization appears limited by scale, with large markets and large academic organizations enabling extremes of specialization. In contrast, scale does not facilitate specialization for the typical oncologist. The median oncologist remains similarly general in large markets as in small markets. In addition, most generalists work in non-academic organizations, which also do not appear to facilitate specialization: neither median nor top specialization rises with organization size. These patterns suggest that most oncologists face some combination of low private returns to specialization and high coordination costs.

7.1 Market size and coordination costs

Highly specialized oncologists appear to face market size constraints. Figure 5c shows the 90th percentile of physician specialization, or "top specialization", by market size in 1999 and 2014, and Table 8 shows a regression of top specialization on market size. In 1999, top specialization increases with market size by 495 (79) HHI points for every doubling in market size. This result is highly consistent with top specialization being limited by the extent of the market, a classic pattern of specialization first observed by Smith (1776). By 2014, the gradient of top specialization with respect to market size has doubled to over 1,060 (94) HHI points per doubling in market size.

The growing gradient of top specialization with respect to market size is consistent with the discussion in section 2.4. Knowledge growth may increase the optimal degree of top specialization, but only in markets large enough to support these extremes. These findings are consistent with growing economies of scale in highly specialized healthcare.

These changes have implications for patient access to highly specialized healthcare. The observed growth in specialization over time occurs entirely in large markets. As a result, patients in large markets have increased local access to subspecialist care, while this access has stagnated or worsened in small ones. This growing gap could worsen geographic differences in health outcomes if subspecialized oncologists provide higher-quality care and patient travel for care remains costly, as observed in Dingel et al. (2023).

In contrast, the median oncologist in each market appears highly constrained by coordination costs with minimal variation in specialization by market size. Figure 5 and Table 8 show that median specialization is flat across the distribution of market size in 1999 ($\beta = -24$ (45)). The typical oncologist in a large market is no more specialized than in a small market. The typical oncologist in a market remains general, even in large markets with enough patients to support narrow specialization. This pattern suggests that market size is not a constraining factor for the typical oncologist in 1999, as specialization does not increase when this constraint is lifted. Instead, this pattern is consistent with coordination costs limiting

the specialization for the typical oncologist, in line with the theories of Becker and Murphy (1992) discussed in section 2.2.

Figure 5 and Table 8 also show that by 2014, median specialization rises slightly with market size ($\beta = 130$ (34)) but remains relatively flat compared to the gradient of top specialization. The typical oncologist in large markets maintains a broad clinical scope despite extensive growth in the volume of knowledge between 1999 and 2014. Through the lens of Becker and Murphy (1992), this pattern suggests that coordination costs remained high relative to the returns to specialization for general oncologists.

Appendix Figure A.5 shows the gradients of median and top specialization with respect to market size in 2016 and 2019. The time frame is too short to detect significant changes.

7.2 Firms and Firm Types

A key difference between generalists and specialists is the nature of the organizations they work in. As shown in Table 3, specialists are relatively concentrated in academic organizations. Large academic organizations facilitate high levels of specialization, while non-academic organizations do not, regardless of their size. One interpretation of these findings is that academic organizations improve coordination between oncologists, enabling specialization, while non-academic ones, even large ones, do not.

Academic organizations drive the extremes of specialization. Figure 6 displays the 50th and 90th percentiles of specialization within an organization by organization size in 2014 for academic places. The 90th percentile is calculated only for organizations with at least five oncologists. For academic organizations, the median and top levels of specialization rise steeply with the number of oncologists working there. One potential interpretation of this finding is that academic organizations successfully reduce coordination frictions to a level that allows oncologists to subspecialize. Without high coordination frictions, organization scale becomes an important limiting factor. The larger the organization, the higher the degree of specialization it supports at the top and middle of the distribution.

In contrast, non-academic organizations appear to have high internal coordination frictions. Figure 6 also displays the 50th and 90th percentiles of specialization by organization size in 2014 for non-academic places. Both median and top specialization are completely flat across the distribution of organization size. The top degree of specialization is not detectably higher at large organizations than at small ones. Similarly, the typical oncologist at the largest institutions is as general as a solo oncology practitioner. Non-academic organizations do not facilitate specialization at either the middle or the top of the distribution. This pattern suggests that non-academic firms do not reduce coordination frictions between oncologists.

Appendix Figure A.6 shows the same results for 2008, the first year for which organization size information is available. The findings are the same, with no detectable changes between 2008 and 2014. Appendix Tables A.16 and 9 display the associated regression tables.

These findings are also potentially consistent with an under-referral equilibrium, as described by Garicano and Santos (2004). Academic institutions appear to mitigate this referral problem, while non-academic ones do not.

An alternative interpretation of these results is that oncologists working at academic institutions face higher returns to specialization rather than lower coordination frictions. Oncologists note that tenure and promotion systems at academic organizations incentivize physicians to develop deep clinical expertise in a narrow domain.

In contrast, non-academic oncologists may face low returns specialization. Non-academic physicians often face strong incentives to see large numbers of patients, an objective that may be more easily achieved with a broad scope of practice.

Oncologists also face blunted price incentives to produce high-quality care. Traditional Medicare payment for oncology over this period did not incorporate any quality measures and, in more recent years, offered only limited, opt-in participation in payment models with quality measures (Chong et al., 2022). While Medicare is not the only payer, it does represent a large share of cancer patients. Providers may under-specialize if they face blunted

incentives for quality. Non-academic oncologists in particular may not internalize as much of the returns to specialization, particularly in the quality dimension.

Lastly, physicians may sort to organizations based on their preferences. Physicians who highly value the work style and status of expertise may sort to academic organizations. In contrast, those with a higher marginal utility of income or clinical variety may sort to non-academics.

8 Technology Use and Specialization

Oncologist specialization varies widely by market size and organization type. Are there detectable returns to specialization in terms of quality of care for patients? Unfortunately, the data does not contain sufficient detail about each patient's cancer (e.g., stage, biomarkers) to test if specialization improves patient outcomes.

However, the data does permit examination of oncologists' adoption and use of anti-cancer drugs, which influence patient outcomes. Anti-cancer drugs are complex to use, as they often have complicated dosing regimens and serious toxicity risks. Medical oncologists are the experts among doctors on the use of these complex and risky drugs. As such, prescription patterns for anti-cancer drugs reflect a central domain of oncologists' technology use.

The results are consistent with specialization improving quality. Specialized oncologists use fewer unique drugs and have more experience with each drug they prescribe than general oncologists. They also tend to prescribe newer drugs and more "Breakthrough Therapy" drugs to patients than generalists. Importantly, these differences have grown substantially, consistent with increasing realized returns to specialization over time.

8.1 Data and Measurement

I estimate various features of the typical bundle of anti-cancer drugs that an oncologist prescribes in a year. Anti-cancer drugs are defined here using World Health Organization

Anatomical Therapeutic Class (ATC) codes that fall into the list of anti-cancer drugs from Pantziarka et al. (2021). If a drug has multiple ATC codes[7], I consider only one code for this analysis. Oncologists' prescriptions are observed using a 20% sample of Medicare Part B and Part D claims. The data starts in 2008, the first year in which Medicare Part D claims use National Provider Identifiers (NPIs) to identify the majority of prescribers.

The sample selection changes slightly from that in section 4. In addition to the 25-patient minimum, the sample is additionally restricted to oncologists who prescribed anti-cancer drugs to at least 11 patients across a 20% sample of Part B and Part D data. This restriction ensures drug bundle features are calculated over a meaningful volume of patients and protects patient privacy. This criterion reduces the number of oncologists in my sample relative to the baseline sample. To counteract this, I examine the full, rather than balanced, panel of oncologists. Using the full panel of oncologists enables the examination of patterns over a longer time horizon (2008-2019) with a larger number of oncologists relative to using a balanced panel. Since the sample shifts slightly each year, oncologist specialization is determined based on their HHI in each year.

8.2 Scope of Technology Use

Specialization enables oncologists to limit the scope of technologies that they use. Figure 7 shows the relationship between oncologists' specialization and the number of unique anti-cancer drugs they prescribe annually. In the first year of the panel, the differences are significant but relatively small. The most specialized oncologists (75th—99th percentile of HHI) prescribe an average of 12.1 (0.2) unique drugs annually. The most general oncologists (1st—50th percentile of HHI) prescribe an average of 13.2 (0.1) drugs, or 9% more.

Over time, these differences grow substantially. In 2019, on average, the most specialized oncologists reduce their scope to 9.0 (0.2) unique drugs. In contrast, the most general oncologists prescribe an average of 12.5 (0.1) unique anti-cancer drugs, or 39% more than

[7]Drugs have multiple ATC codes when they have multiple distinct medical uses.

the specialists. Specialization narrows the scope of technologies that oncologists use.

As a robustness check, the analysis is replicated with a subset of 11 random patients per oncologist, drawn without replacement, rather than all patients in Appendix A.7. Using the same number of patients for every oncologist eliminates any mechanical relationship between the number of patients seen and the observed number of unique drugs prescribed. The same patterns hold, with more specialized oncologists reducing the number of drugs they prescribe relative to generalists over time.

8.3 Experience

This narrower scope enables specialized oncologists to have more experience with the drugs they prescribe. Oncologists' experience with a drug is measured by the number of patients they prescribe it to in a year. This measure incorporates only recent experience with a drug, a decision made because physician human capital appears to decay rapidly (Hockenberry and Helmchen, 2014; Facchini, 2022). Experience across drugs is then normalized to relative experience, with a mean value of one across physicians within each drug-year.

Figure 8 displays average relative experience by oncologist specialization over time. In 2008, the most specialized oncologists have slightly (6%) less, experience with the drugs they prescribe than general oncologists. This difference is driven by the most specialized oncologists having lower patient volumes; It disappears when patient volumes are equalized to 11 random patients per oncologist, as shown in Appendix Figure 8.

However, specialists gain experience relative to generalists over time. By 2013, this pattern has reversed, and by 2019 the most specialized oncologists have 28% more experience with the drugs they prescribe than general oncologists. As these drugs are typically complex and high-risk, more experience with a drug plausibly translates into better clinical management of drug treatment and associated side effects.

8.4 Drug Age and Quality

Specialization also helps oncologists to keep up with new technology, as measured by the age and share of "breakthrough" drugs in their drug bundles. Drug age is calculated as years since the first FDA approval for the ATC code for any indication, dosage, or delivery form[8]. A drug is classified as a breakthrough drug if the molecule, as of 2025, ever received a "Breakthrough Therapy" designation from the U.S. Food and Drug Administration. This designation is given when preliminary evidence suggests the drug may be a substantial improvement over current treatment options. It also correlates with measures of clinical benefit, providing a measure of drug quality (Feng and Maini, 2016; Sun et al., 2020). Features of oncologist drug bundles are calculated weighting by the number of patients prescribed to.

Figure 9a shows the average age of drug bundles by degree of specialization over time. In 2008, specialist and generalist oncologists prescribe drugs of a similar age. The average age of drugs prescribed rises over time for everyone, but more for generalist oncologists. By 2019, generalist oncologists prescribe a significantly older set of drugs than specialists, 2.7 years older on average. These results are consistent with specialization enabling oncologists to adopt new drugs more quickly.

Notably, generalist oncologists initially appear not to update their drug bundles. Their drug bundle ages one year every year. However, this result is driven by the use of steroids, an old medication class, which are given to many patients and hence weighted heavily in the drug bundle. Appendix figure A.9 drops steroids. This figure reveals that both generalist and specialist non-steroid drug bundles age less than one year every year, suggesting both groups do adopt new drugs. As before, specialists update their drug bundles faster over time than generalists, pulling ahead in using new technology. Generalist oncologists update their prescribing, but not fast enough to keep up with the frontier.

Figure 9b shows that specialized oncologists also prescribe a greater share of breakthrough

[8]This approach raises the average drug age relative to a strategy that accounts for new indications. However, the data does not permit observation of which indication a drug was given for or if it was given off-label without an FDA-approved indication.

drugs than generalists. The magnitude of the difference varies year to year, but specialized oncologists consistently prescribe significantly more of these novel high-quality drugs than generalists.

In summary, specialized oncologists have increasingly different practice patterns from generalist oncologists. Specialized oncologists use a narrower set of drugs, enabling them to have more experience with each drug and to prescribe newer sets of drugs. These results suggest that specialized oncologists provide higher quality care than generalists and that these quality differences are growing over time. They are also consistent with greater specialization enabling expert workers to stay at the frontier of a rapidly evolving set of technologies.

Do the observed differences in prescribing patterns lead to differences in guideline-concordant care or patient health outcomes? Unfortunately, I do not observe sufficient detail about each patient's cancer (e.g., stage, biomarkers) in the data to plausibly answer this question. However, several case studies from the medical literature find that academic centers provide more guideline-concordant care and are associated with improved patient survival (Voss et al., 2017; Thiels et al., 2020). For example, Voss et al. (2017) find that soft-tissue sarcoma patients treated with guideline-concordant care have better survival, and patients are more likely to receive guideline-concordant care if treated at an academic hospital. (Thiels et al., 2020) find similar results for stomach cancer. As specialists are concentrated in academic centers, the findings suggest that specialists provide higher-quality care and improve survival.

9 Conclusion

This paper examines how growth in general knowledge impacts expert specialization and documents subsequent technology adoption in medical oncology. I document large and rapid increases in anti-cancer drugs and the corpus of clinically relevant knowledge. Despite this knowledge growth, most oncologists remain general and do not become more specialized

over time. However, a small subset of initially specialized oncologists become even more specialized in response to knowledge growth.

These differences in specialization impact patient care. Specialized oncologists prescribe newer anti-cancer drugs and have more experience with each drug they prescribe than general oncologists. These differences in drug prescribing are consistent with specialization enabling higher quality care and have grown over time.

The patterns of specialization also affect the geography of healthcare delivery. Top specialization has increased substantially in the largest markets while stagnating or declining in the smallest ones. This growing dispersion provides evidence that knowledge growth increases economies of scale and geographic inequality in specialized medical care.

10 Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author used Claude from Anthropic and ChatGPT from OpenAI in order to revise sections of the paper. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

References

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

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.

Jeffe, Donna B, and Dorothy A Andriole. 2018. "Prevalence and Predictors of Us Medical Graduates' Federal F32, Mentored-K, and R01 Awards: A National Cohort Study." *Journal of Investigative Medicine*, 66(2): 340–350.

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

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

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.

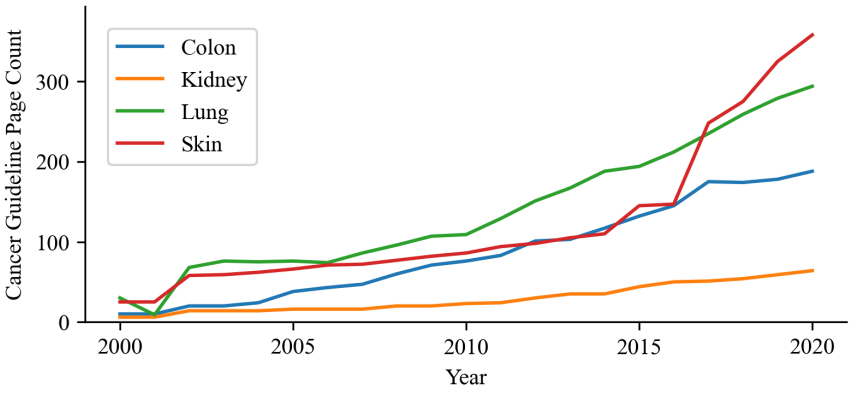
Voss, Rachel K, Yi-Ju Chiang, Keila E Torres, B Ashleigh Guadagnolo, Gary N Mann, Barry W Feig, Janice N Cormier, and Christina L Roland. 2017. "Adherence to National Comprehensive Cancer Network Guidelines is Associated with Improved Survival for Patients with Stage 2A and Stages 2B and 3 Extremity 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.

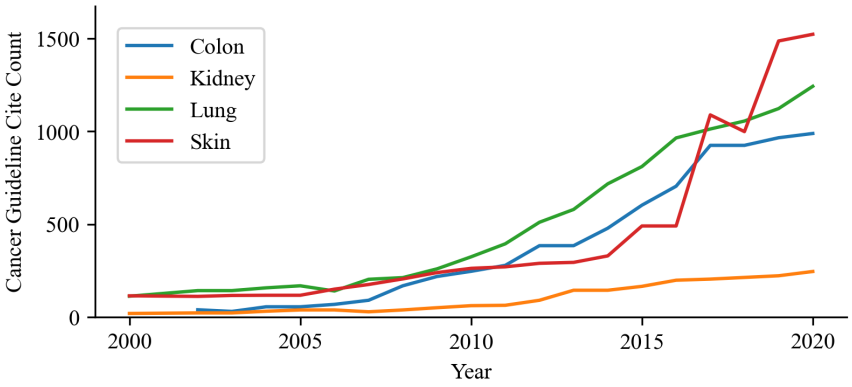
Welch, W Pete, and Andrew B Bindman. 2016. "Town and Gown Differences Among the 100 Largest Medical Groups in the United States." *Academic Medicine*, 91(7): 1007–1014.

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

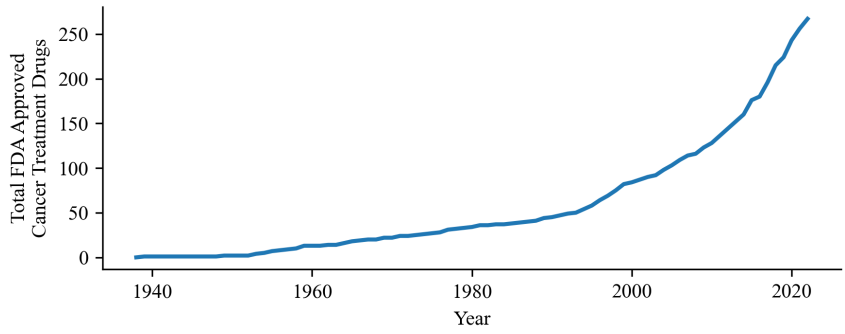
(a) Oncology clinical practice guidelines have become considerably longer



(b) Oncology clinical practice guidelines include many more citations



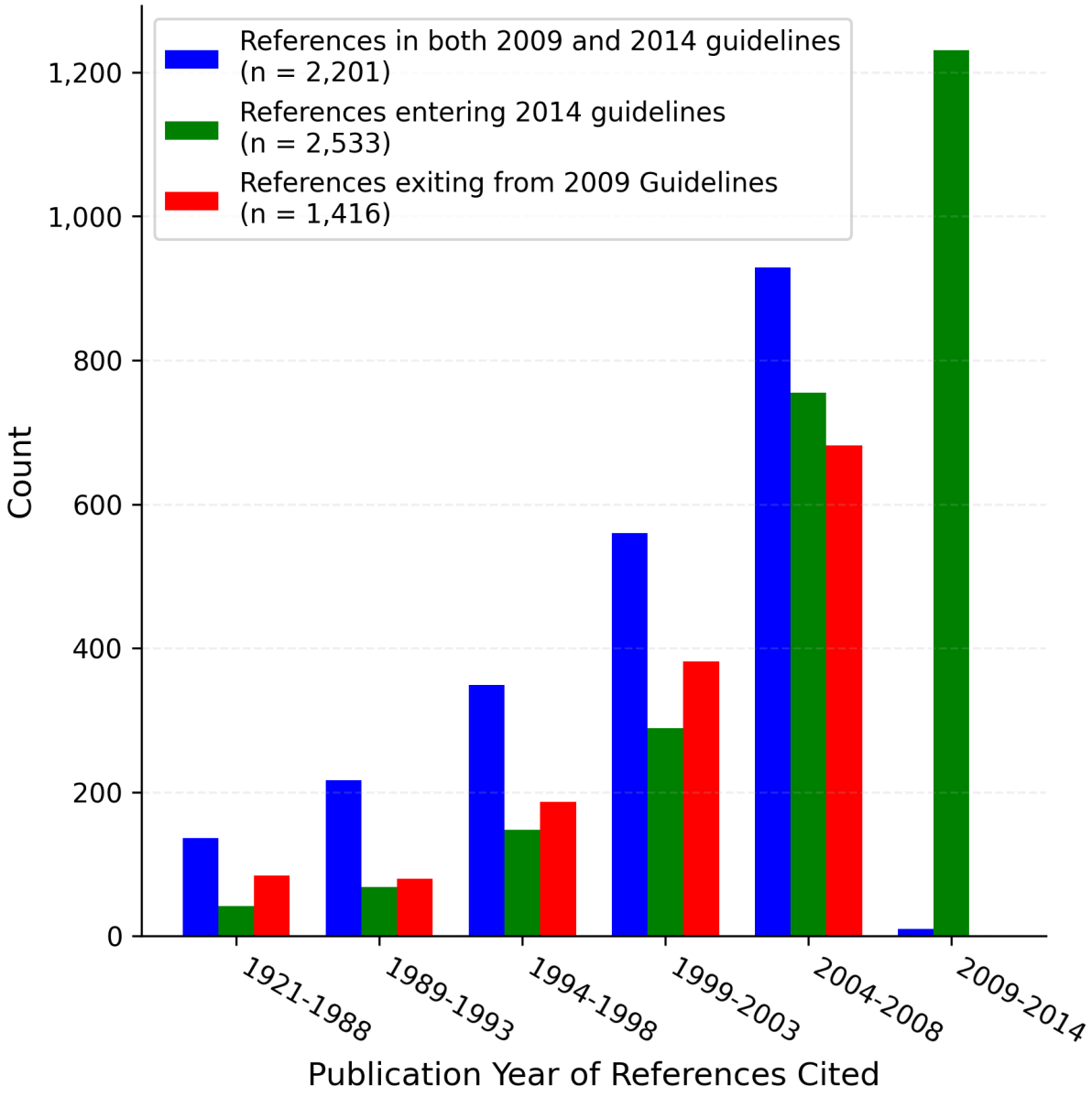
(c) The number of FDA-approved anti-cancer drugs has also increased



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 page count and citation count of the first set of treatment 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: Growth in oncology clinical practice guideline citations between 2009 and 2014 comes primarily from adding references to new and recent scholarly works

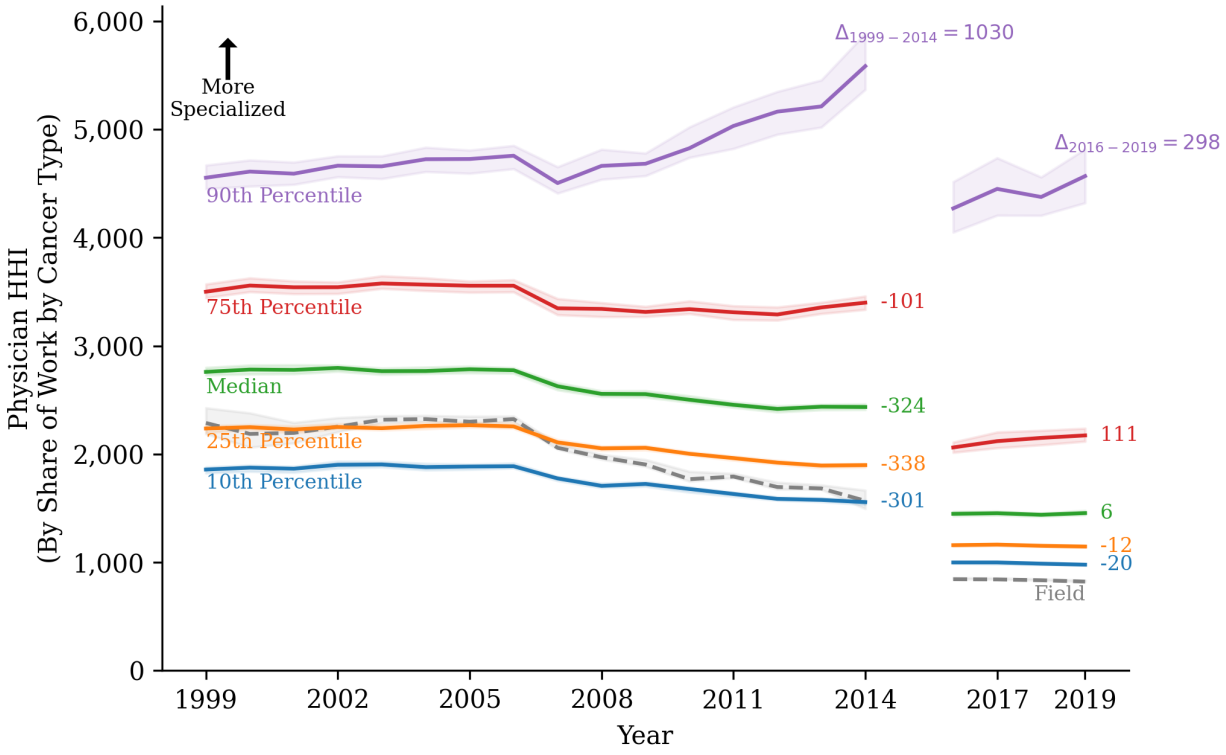
(a) Citations entering, exiting, and staying in 2009 and 2014 guidelines by publication year



Notes: Figure 2 describes changes in the publication year of references cited by NCCN Clinical Practice Guidelines in Oncology between the 2009 and 2014 guidelines. 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 3: Only specialized oncologists are growing more specialized over time

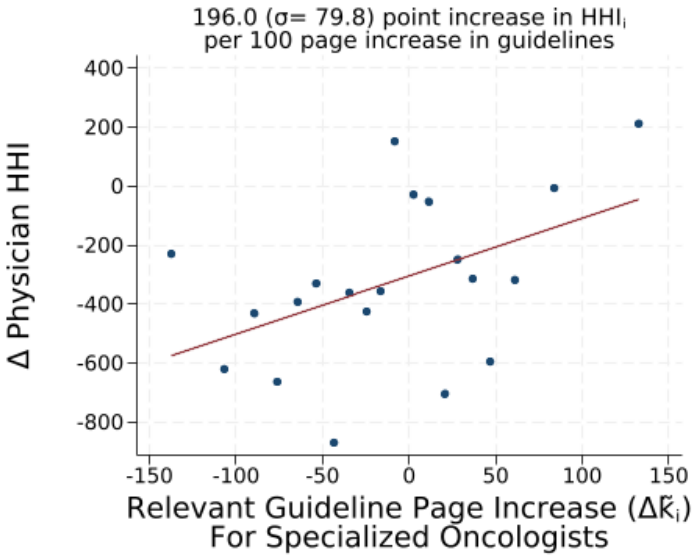
(a) Distribution of oncologist specialization over time



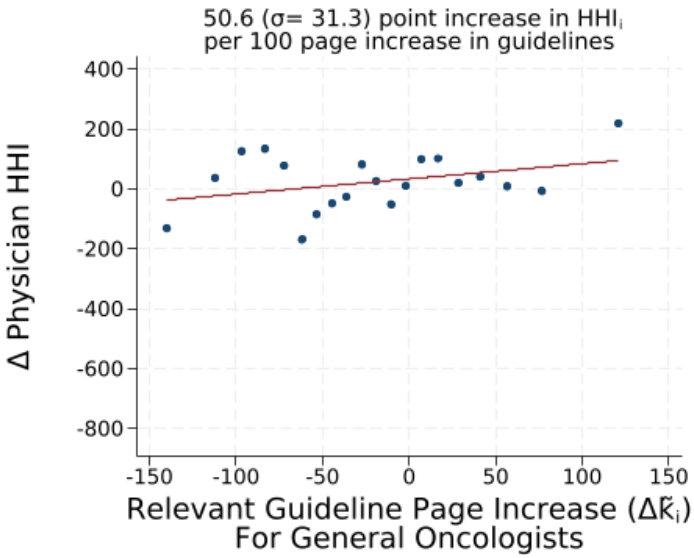
Notes: This figure shows the distribution of physician-level Herfindahl-Hirschman Index (HHI), a measure of specialization, from 1999-2019. All physicians in the sample are hematologist-oncologists, medical oncologists, or hematologists who saw at least 25 unique cancer patients in the year in a 20% 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)



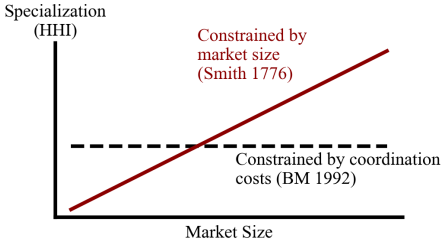
(b) Change in specialization (HHI) vs. exposure to knowledge growth for generalist oncologists (lower 75% of initial specialization distribution)



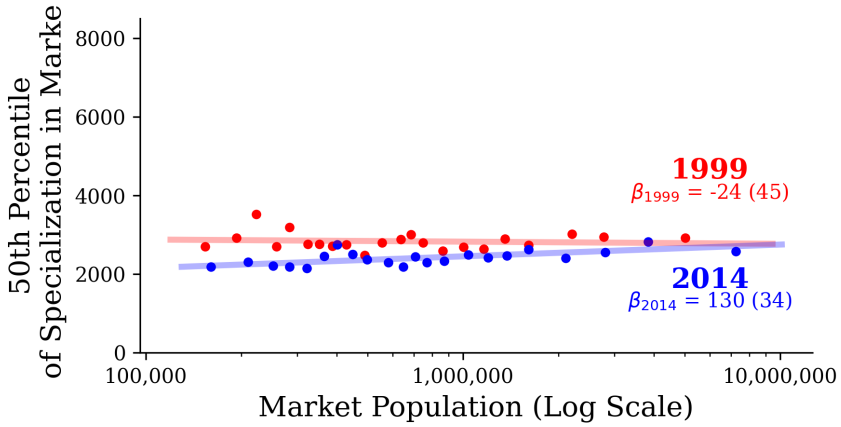
Notes: Panels (a) and (b) show binned scatterplots and regression estimates for the relationship between the recentered exposure to knowledge growth and change in specialization. Specialization is measured using a Herfindahl—-Hirschman Index (HHI) over types of cancer managed. Panel (a) shows this relationship 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

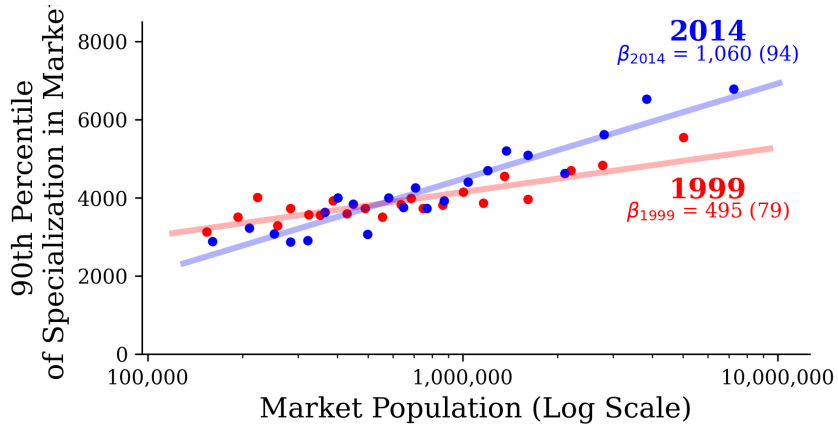
(a) Relationship between market size and specialization



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



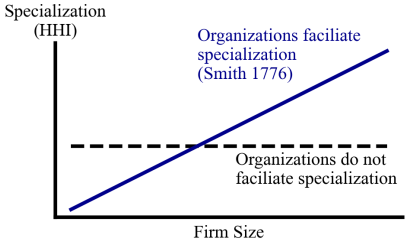
(c) 90th percentile of specialization by market size over time markets (1999 and 2014)



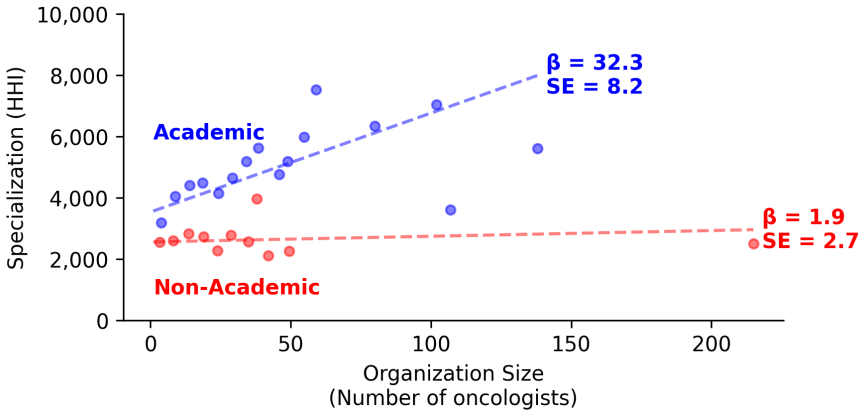
Notes: Panel (a) shows the predicted relationship between specialization and market size if specialization is constrained by market size (Smith, 1776) versus coordination costs (Becker and Murphy, 1992). Panels (b) and (c) show binned scatterplots of the 50th and 90th percentile of oncologist specialization 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

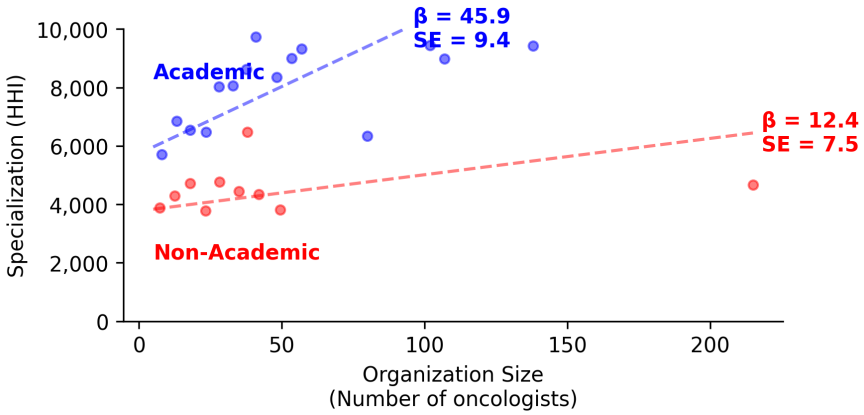
(a) Relationship between organization size and specialization



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



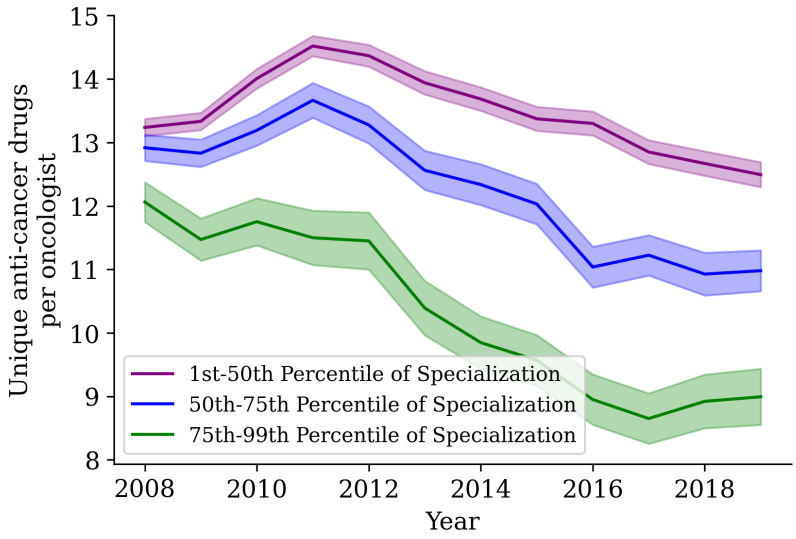
(c) 90th percentile of specialization by organization size and type (2014)



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

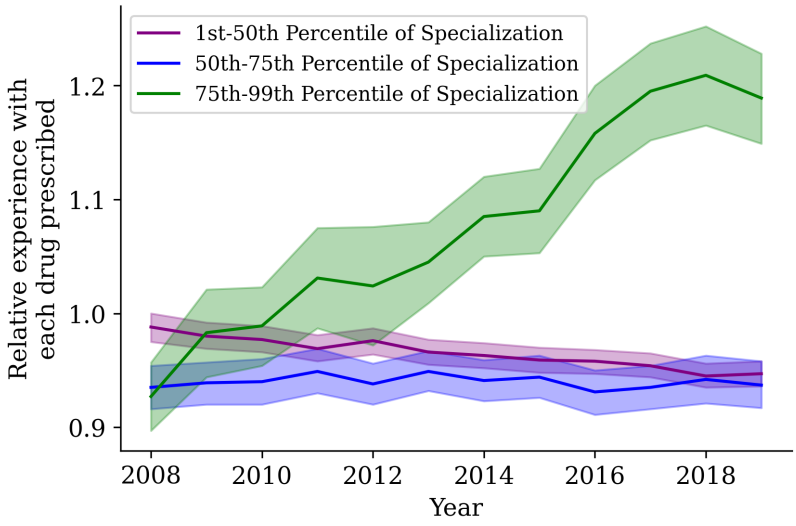
(a) Count of unique anti-cancer drugs prescribed by year and degree of specialization



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

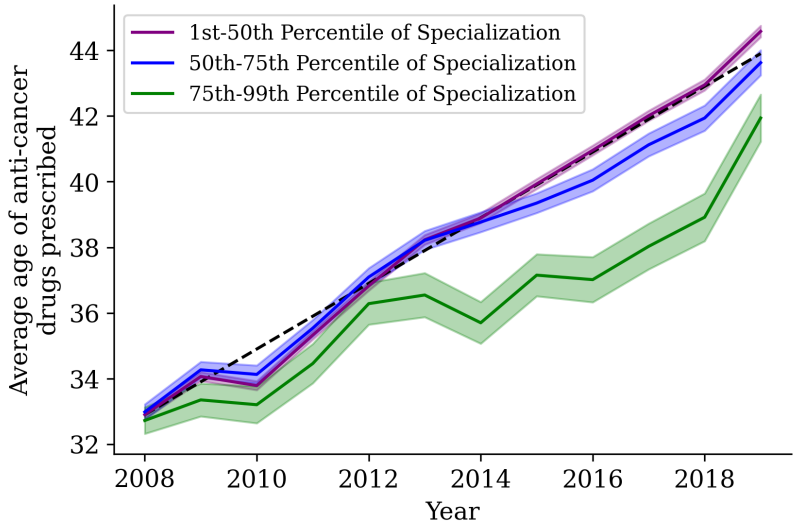
(a) Relative experience with anti-cancer drugs prescribed by year and degree of specialization



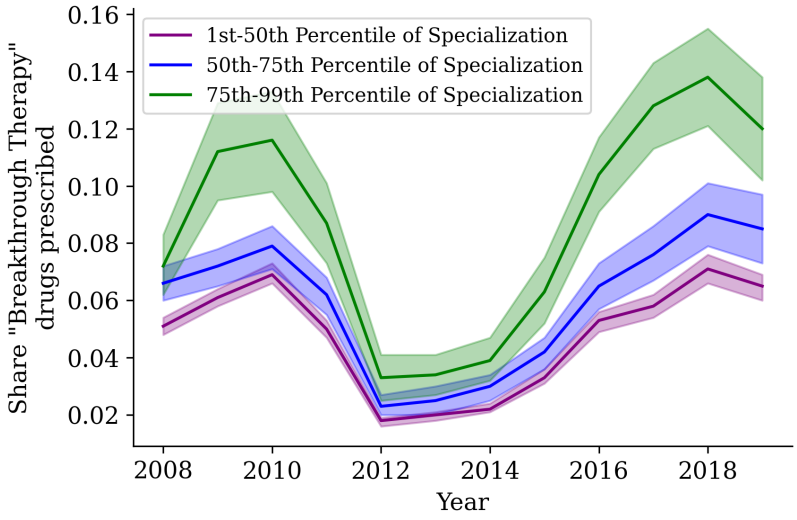
Notes: This figure shows oncologists' average relative experience with the anti-cancer drugs that they prescribe over time by level of specialization. Experience with a drug is calculated as the number of patients an oncologist prescribed a drug to in a year. Relative experience normalizes experience 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) Share of "Breakthrough Therapy" anti-cancer drugs prescribed by year and degree of specialization, excluding steroids



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

multicolumn7cDependent Variable: HHI	(1) Knowledge Growth Recentered ($\Delta\tilde{k}_i$)	(2) Restrict Sample	(3) Add Age FE	(4) Add Market FE	(5) Add Pat Count FE	(6) Add Org Size FE
Increase in Guideline Length (100 Pages)	196.0	213.2	220.0	253.1	255.2	246.3
	(79.81)	(81.06)	(81.56)	(89.07)	(89.98)	(91.20)
Observations	988	924	924	924	924	924
R-squared	0.007	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

multicolumn7cDependent Variable: HHI	(1) Knowledge Growth Recentered ($\Delta\tilde{k}_i$)	(2) Restrict Sample	(3) Add Age FE	(4) Add Market FE	(5) Add Pat Count FE	(6) Add Org Size FE
Increase in Guideline Length (100 Pages)	50.57	54.23	54.18	66.96	66.50	66.26
	(31.32)	(31.45)	(31.44)	(33.14)	(33.13)	(33.03)
Observations	2,962	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.

11 Appendix

For Online Publication

Knowledge Growth and Specialization: Evidence from Oncologists

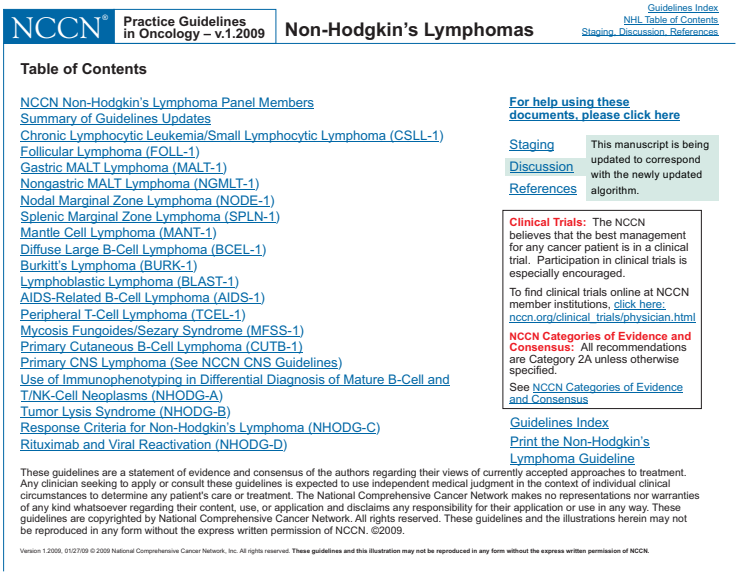
Maya Lozinski

March 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.

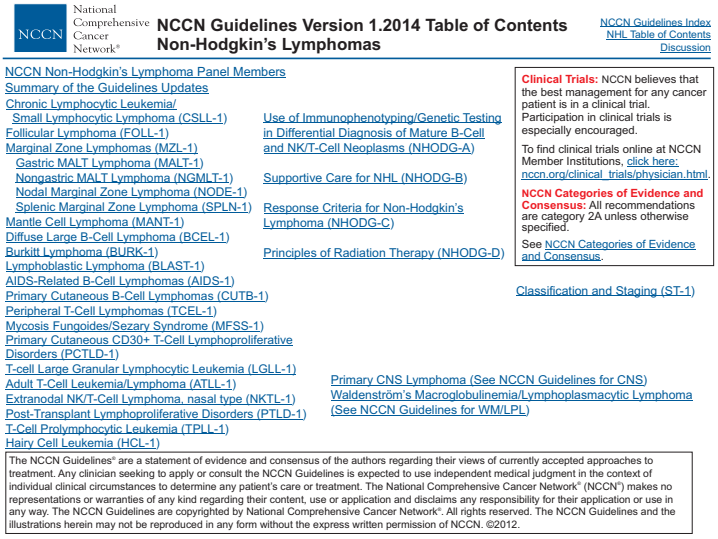
See NCCN Categories of Evidence and Consensus

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)

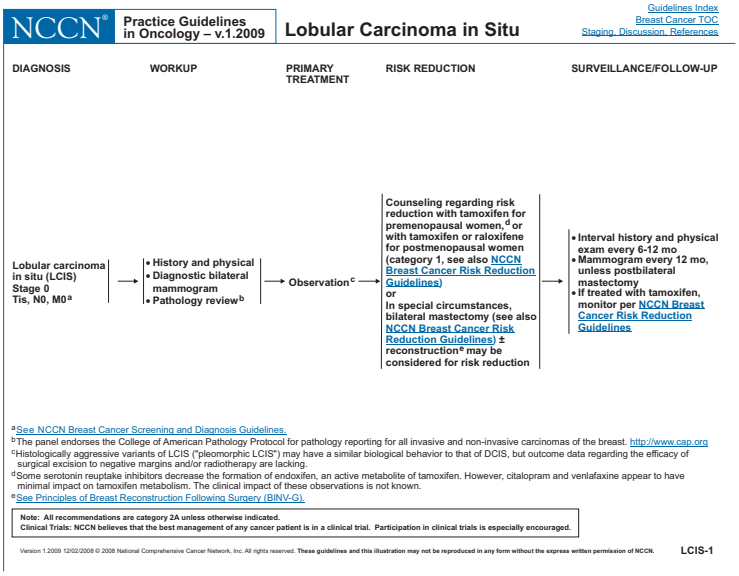
The NCCN Guidelines® are a statement of evidence and consensus of the authors regarding their views of currently accepted approaches to treatment. Any clinician seeking to apply or consult the NCCN Guidelines is expected to use independent medical judgment in the context of individual clinical circumstances 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®.**

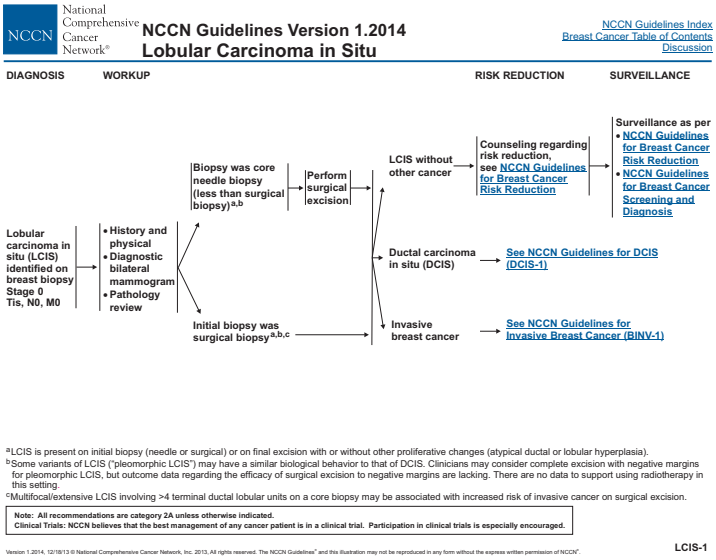
Notes: Figure A.1 shows example changes from NCCN Guidelines. Figures A.1a and A.1b show the table of contents for non-Hodgkins lymphoma (NHL) in 2009 and 2014, respectively. Figure A.1a reproduced with permission from the NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®) for Non-Hodgkin's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



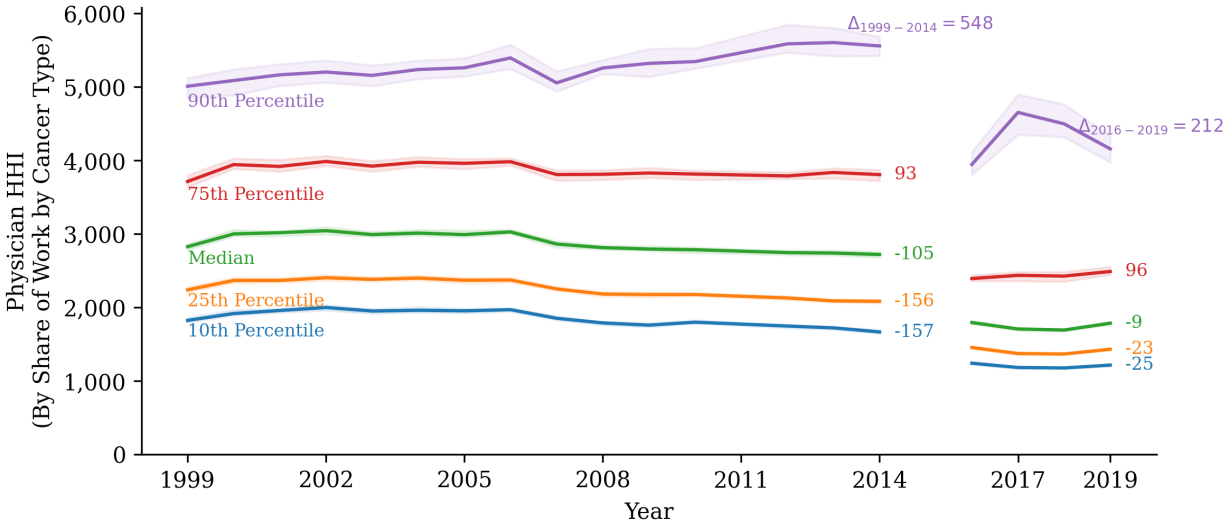
(b) 2014 Flowchart



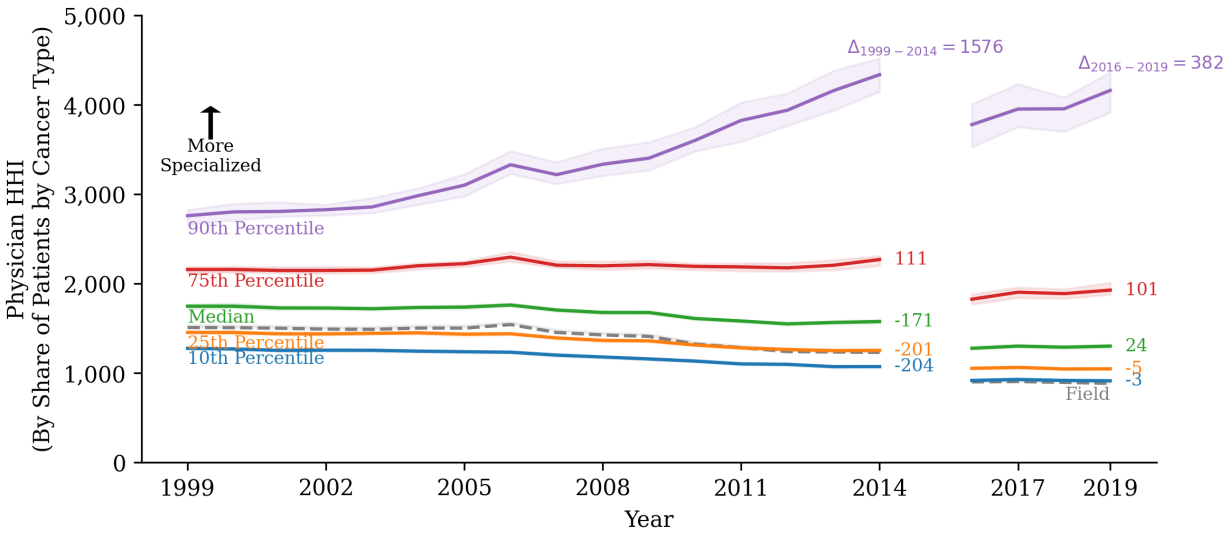
Notes: Figure A.2 shows example changes from NCCN Guidelines. Figures A.2a and A.2b show a flowchart with the diagnostic workup for lobular carcinoma in-situ, a subtype of breast cancer, in 2009 and 2014, respectively. Figure A.2a reproduced with permission from the NCCN Clinical Practice Guidelines 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: Specialized oncologists are growing more specialized over time

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

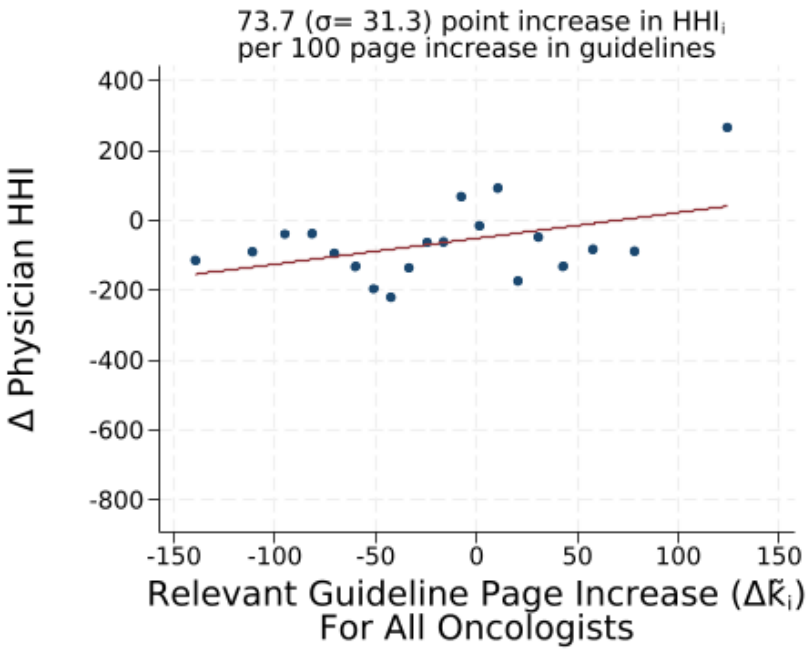


(b) Oncologist specialization over time, measured using patient counts



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.

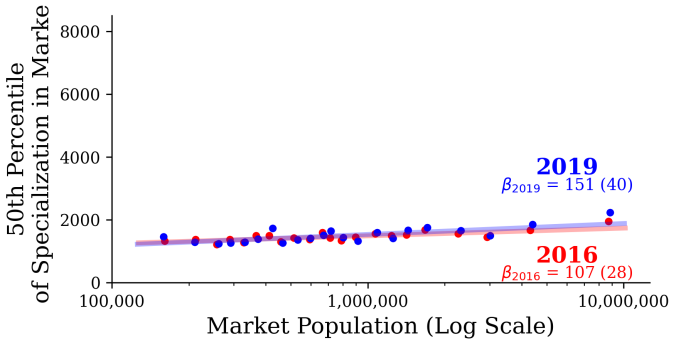
Figure A.4: Knowledge growth increases oncologist specialization in aggregate



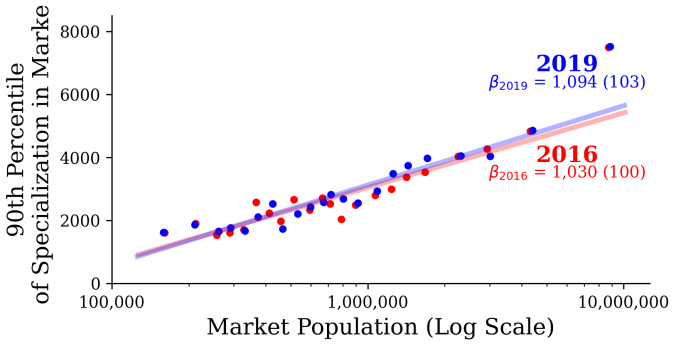
Notes: Figure A.4 shows a binned scatterplot and key regression estimates for the relationship between the recentered exposure to knowledge growth and change in specialization for all oncologists. In the recentered measure, the predicted exposure to knowledge growth is subtracted from raw exposure to 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.5: No detectable change in specialization by market size between 2016 and 2019

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



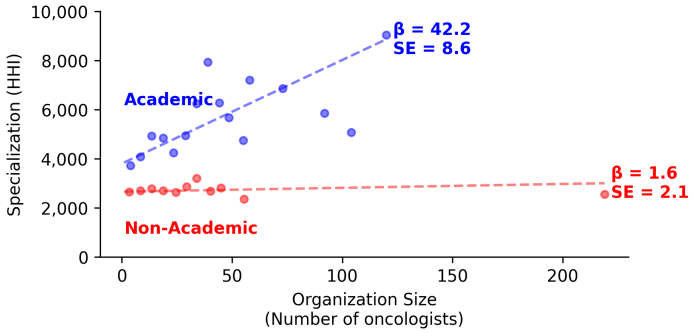
(b) 90th percentile of specialization by market size over time markets (2016 and 2019)



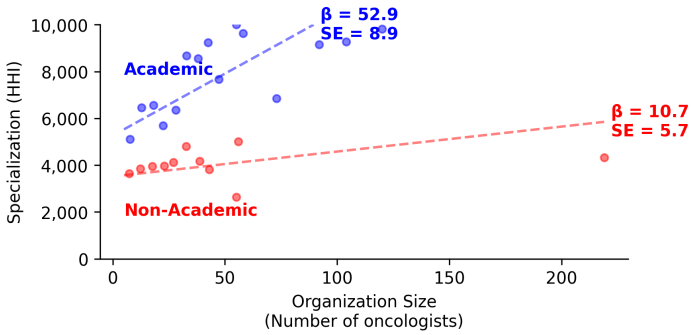
Notes: Panel (a) shows the predicted relationship between specialization and market size if specialization is constrained by market size (Smith, 1776) versus coordination costs (Becker and Murphy, 1992). Panels (a) and (b) show binned scatterplots of the 50th and 90th percentile of oncologist specialization 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.6: Academic organizations facilitate specialization while non-academic ones do not (2008)

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



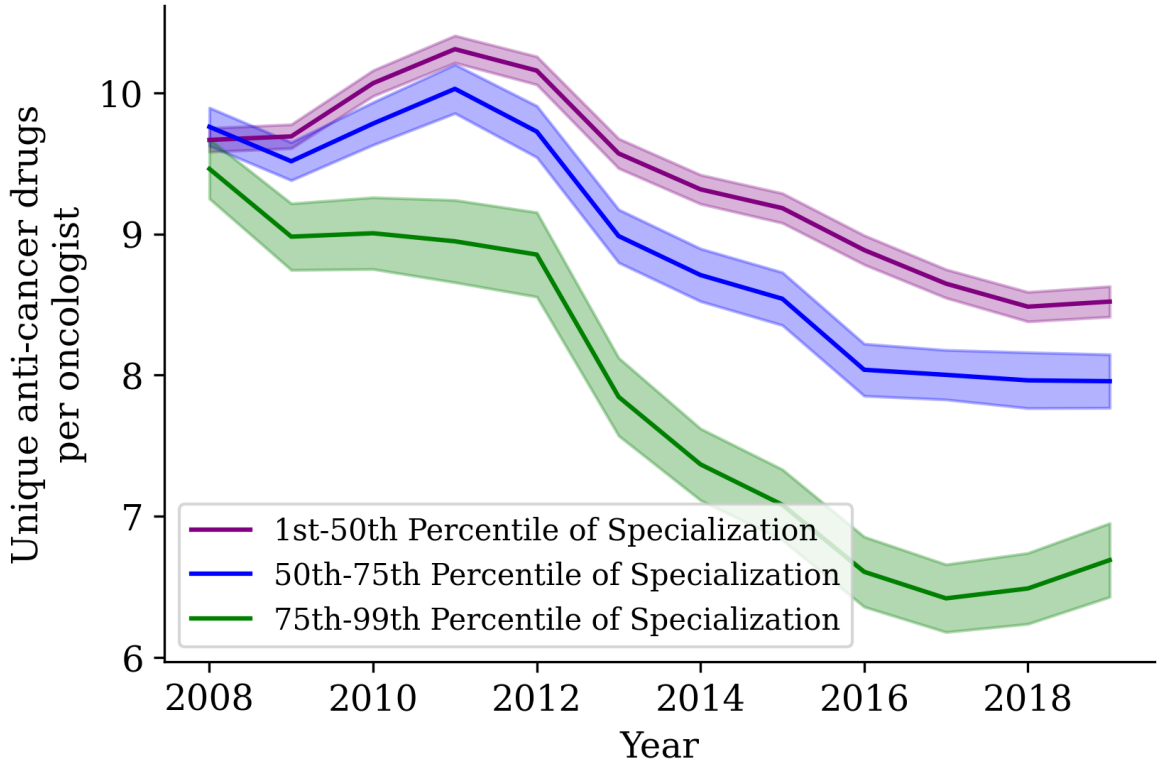
(b) 90th percentile of specialization by organization size and type (2008)



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.7: Over time, specialists begin to prescribe fewer unique anti-cancer drugs than generalists in a year (11 random patients per oncologist)

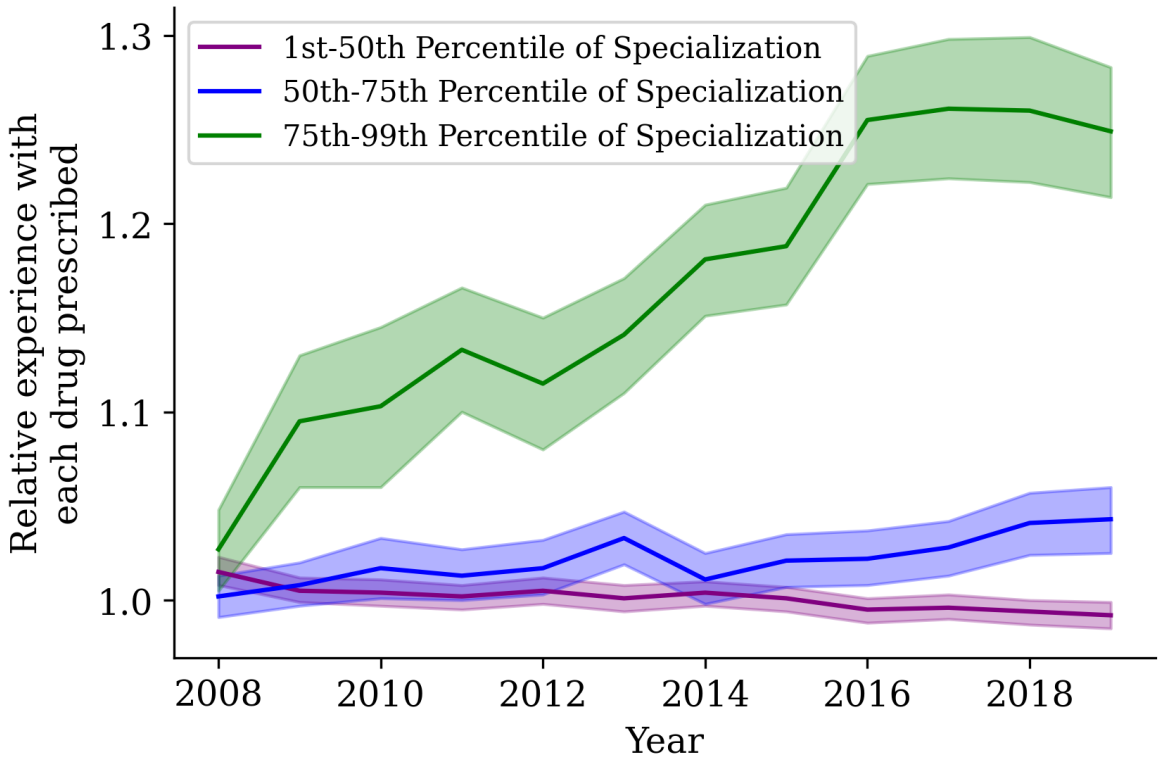
(a) Count of unique anti-cancer drugs prescribed by year and degree of specialization (11 random patients per oncologist)



Notes: This figure shows the average number of unique anti-cancer drugs that oncologists prescribe over time by level of specialization. This figure differs from Figure 7 in that the average number of unique anti-cancer drugs is calculated over 11 random patients, rather than all patients, per oncologist per year. This change in sample removes systematic differences in patient volume that may exist by level of oncologist specialization. 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 A.8: Over time, specialists become relatively more experienced with the drugs they prescribe than generalists (11 random patients per oncologist)

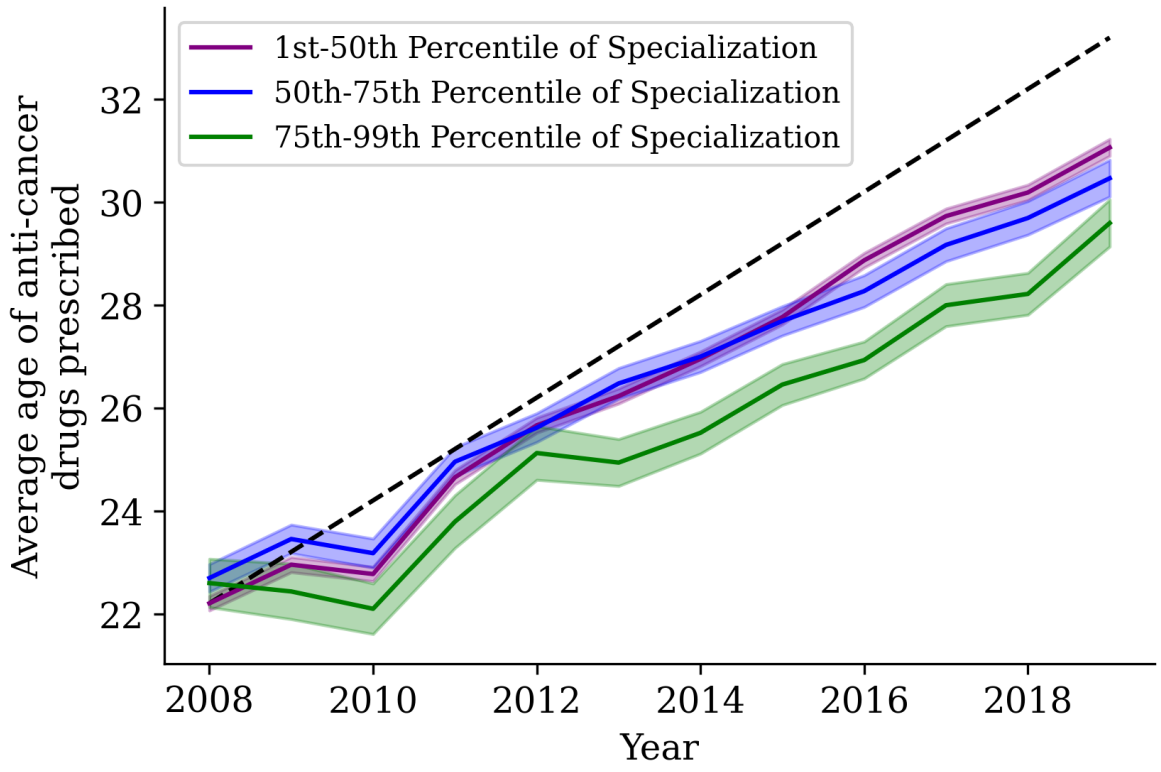
(a) Relative experience with anti-cancer drugs prescribed by year and degree of specialization (11 random patients per oncologist)



Notes: This figure shows oncologists' average relative experience with the anti-cancer drugs that they prescribe over time by level of specialization. This figure differs from Figure 8 in that average relative experience is calculated over 11 random patients, rather than all patients, per oncologist per year. This change in sample removes systematic differences in patient volume that may exist by level of oncologist specialization. Experience with a drug is calculated as the number of patients an oncologist prescribed a drug to in a year. Relative experience normalizes these values to have a mean of one in a year. 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 A.9: Over time, specialists begin to prescribe newer drugs than generalists (excluding steroids)

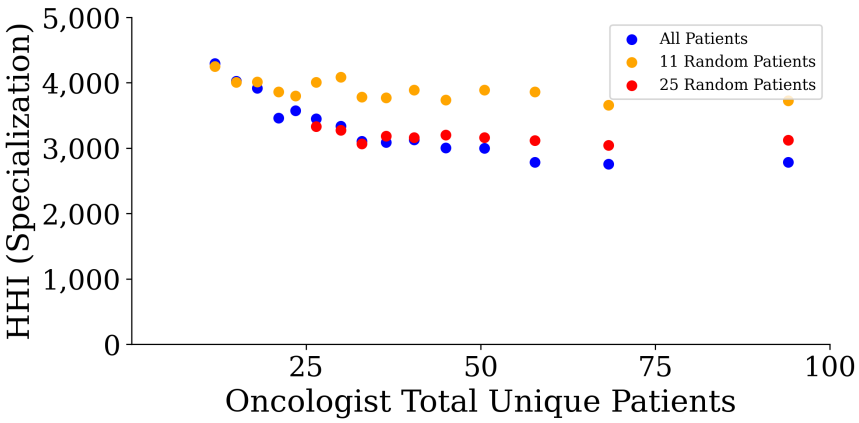
(a) Average age of non-steroid anti-cancer drugs prescribed by year and degree of specialization



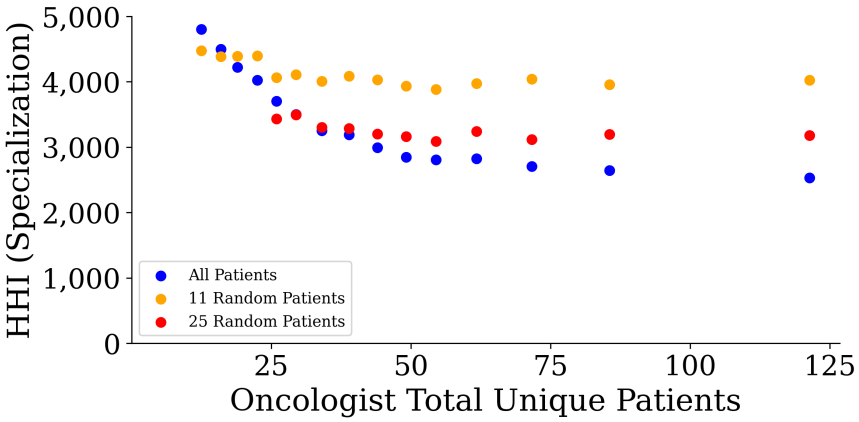
Notes: This figure shows the average age of non-steroid anti-cancer drugs that oncologists prescribe over time by level of specialization. In the early years of the panel, specialized and generalist oncologists prescribe drug bundles of a similar age. However, by the end of the panel, the more specialized an oncologist is, the newer the drugs they prescribe. 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. 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. 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 A.10: Physician level HHI declines with patient volume

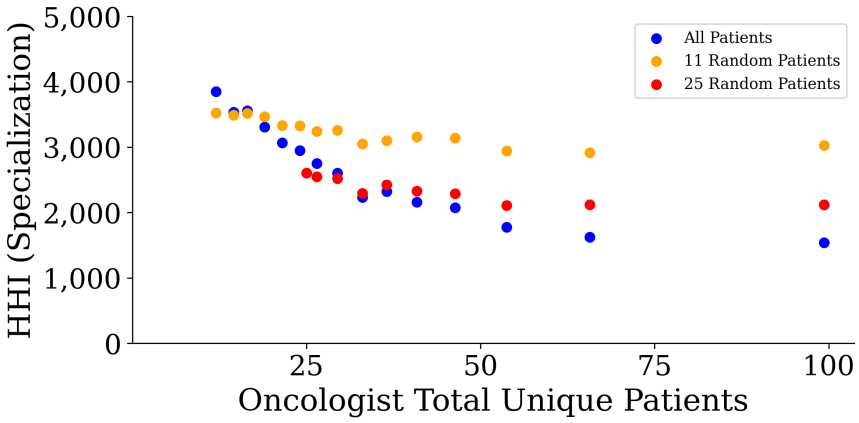
(a) HHI by patient volume (1999)



(b) HHI by patient volume (2009)



(c) HHI by patient volume (2019)



Notes: Figure A.10 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.10a, A.10b, 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

multicolumn7cDependent Variable: HHI	(1) Knowledge Growth Recentered ($\Delta\tilde{k}_i$)	(2) Restrict Sample	(3) Add Age FE	(4) Add Market FE	(5) Add Pat Count FE	(6) Add Org Size FE
Increase in Guideline Length (100 Pages)	73.74	71.95	71.62	87.93	88.79	89.36
	(31.27)	(31.43)	(31.41)	(32.72)	(32.71)	(32.67)
Observations	3,950	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

multicolumn7cDependent Variable: HHI	(1) Knowledge Growth Recentered ($\Delta\tilde{k}_i$)	(2) Restrict Sample	(3) Add Age FE	(4) Add Market FE	(5) Add Pat Count FE	(6) Add Org Size FE
Increase in Guideline Length (100 Pages)	-0.111	-0.135	-0.172	-0.197	-0.222	-0.248
	(0.169)	(0.174)	(0.174)	(0.188)	(0.188)	(0.190)
Observations	988	924	924	924	924	924
R-squared	0.000	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

multicolumn7cDependent Variable: HHI	(1) Knowledge Growth Recentered ($\Delta\tilde{k}_i$)	(2) Restrict Sample	(3) Add Age FE	(4) Add Market FE	(5) Add Pat Count FE	(6) Add Org Size FE
Increase in Guideline Length (100 Pages)	-0.165	-0.182	-0.180	-0.236	-0.244	-0.236
	(0.0909)	(0.0909)	(0.0908)	(0.0935)	(0.0937)	(0.0935)
Observations	2,962	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

multicolumn7cDependent Variable: HHI	(1) Knowledge Growth Recentered ($\Delta\tilde{k}_i$)	(2) Restrict Sample	(3) Add Age FE	(4) Add Market FE	(5) Add Pat Count FE	(6) Add Org Size FE
Increase in Guideline Length (100 Pages)	-0.141	-0.145	-0.156	-0.213	-0.233	-0.233
	(0.0800)	(0.0801)	(0.0798)	(0.0817)	(0.0812)	(0.0813)
Observations	3,950	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

multicolumn5cDependent 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

multicolumn5cDependent 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

multicolumn4cDependent Variable: HHI	(1)	(2)	(3)
Increase in Guideline Length (100 Pages)	213.1	-13.24	-113.9
	(81.07)	(93.34)	(100.2)
Observations	988	988	988
R-squared	0.007	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

multicolumn4cDependent Variable: HHI	(1)	(2)	(3)
Increase in Guideline Length (100 Pages)	12.28	58.99	3.659
	(31.01)	(42.84)	(43.86)
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

multicolumn4cDependent Variable: HHI	(1)	(2)	(3)
Increase in Guideline Length (100 Pages)	48.73	35.44	-27.85
	(30.77)	(39.95)	(41.21)
Observations	3,950	3,950	3,950
R-squared	0.001	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

multicolumn7cDependent Variable: HHI

	(1) Knowledge Growth Recentered ($\Delta\tilde{k}_i$)	(2) Restrict Sample	(3) Add Age FE	(4) Add Market FE	(5) Add Pat Count FE	(6) 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

multicolumn7cDependent Variable: HHI

	(1) Knowledge Growth Recentered ($\Delta\tilde{k}_i$)	(2) Restrict Sample	(3) Add Age FE	(4) Add Market FE	(5) Add Pat Count FE	(6) 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

multicolumn7cDependent Variable: HHI	(1) Knowledge Growth Recentered ($\Delta\tilde{k}_i$)	(2) Restrict Sample	(3) Add Age FE	(4) Add Market FE	(5) Add Pat Count FE	(6) 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: 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.16 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.3a) 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.10 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.3a 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.