Knowledge Growth and the Organization of Expert Work*

Maya Lozinski
University of Chicago

July 2024

Abstract

Many fields, such as computer science, molecular biology, and medicine, have a rapidly growing knowledge base. Do expert workers respond to growth in knowledge by becoming more specialized? We study this empirically in the context of oncology, where we document explosive growth in knowledge. Using Medicare claims data and novel panel of historical cancer treatment guidelines, we test if oncologists exposed to greater knowledge growth become more specialized. We proxy knowledge growth in each cancer subfield with the increase in the length of clinical guidelines. Exposure to knowledge growth causes ex-ante specialized oncologists to become even more specialized but does not affect general oncologists. This growth in specialization over time occurs only in large markets. Specialists also increasingly provide higher quality care: over time, they prescribe more up-to-date cancer drugs and have more experience with the drugs they prescribe than generalists. These patterns suggest that knowledge growth increases economies of scale in healthcare and leads to growing geographic inequality in local access to specialized care.

Keywords: healthcare, specialization, knowledge growth, economies of scale, technology adoption
JEL codes: I10, J24

*Thank you to Zarek Brot-Goldberg, Joshua Gottlieb, David Meltzer, Dan Black, Tamara Konetzka, and Pauline Mourot. Thanks to the NIH Medical Scientist Training Program (T32GM007281), the NIA Program in Medicine, the Social Sciences, and Aging (T32AG051146), and the Center for Research Informatics (2U54TR002389-06). 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. Expert workers often must stay at the frontier of this rapidly growing knowledge base. However, staying up to date can be quite costly. Specialization is one possible response to knowledge growth, as it reduces the cost of maintaining expertise. Several economic theories predict that workers should become more specialized as general knowledge increases (Becker and Murphy, 1992; Jones, 2009). We empirically test if expert workers respond to growth in knowledge by becoming more specialized.

We evaluate how specialization responds to growth in knowledge in the context of medical oncology. Medical oncologists are physicians who treat cancer patients medically, i.e., with drugs. They are extremely high human capital individuals; they must complete college, four years of medical school, and at least five years of supervised clinical training to practice in the United States. Entry into the field of medical oncology is highly regulated. However, further specialization within the field we study is not regulated by any governing bodies. In addition, the healthcare setting enables us to track subtle changes over time in the domain areas that individual oncologists work in. These features make medical oncology an ideal setting to study how specialization evolves.

In addition, cancer is a condition with a vast demand for treatment, leading to rapid innovation and knowledge growth. The lifetime risk of developing cancer is 40%(American Cancer Society, 2024)[1]. As a result, there has been substantial research and innovation in cancer over the last few decades. For example, the total number of FDA-approved anti-cancer drugs increased from 90 to 243[2]. As one oncologist told this author, "Some fields of cancer treatment are as complex now as the whole of oncology was 20 years ago"[3]. Oncologists openly acknowledge how challenging it is to stay up to date: "Incorporating this torrential

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

[2]Data from Pantziarka et al. (2021)

[3]He noted he had no evidence for this statement, but we can empirically corroborate it

evolution of knowledge is an impossible, Sisyphean task" (Wallis, 2023). As such, this setting enables us to study the effects of rapid technological evolution and knowledge growth in a context with material stakes.

We conceptualize the relationship between general knowledge growth and specialization in this context using the model from Becker and Murphy (1992). Oncologists can treat every type of cancer or collaborate with other oncologists to divide patients by cancer type. If they divide patients by cancer type, they will produce a higher quantity or quality of patient care. For example, an oncologist specializing in breast cancer will have more experience with breast cancer therapies, leading to higher quality of care or patient throughput. However, oncologists may also face coordination costs that limit their ability to specialize, including information frictions, agency issues, and contracting difficulties. The optimal team size and degree of specialization balance the benefits of specialization with the costs of coordination. Becker and Murphy (1992) postulate that increased general knowledge increases the returns to specialization and pushes the equilibrium worker towards a more specialized role. We estimate this relationship empirically.

We first quantify knowledge growth in oncology. We create a novel panel of historical cancer guidelines data using the National Comprehensive Cancer Network guidelines, the leading comprehensive cancer treatment guidelines. We use the page length of guidelines as a proxy of the size of the clinically relevant knowledge base for each type of cancer in each year. We proxy the growth in knowledge with the increase in the length of guidelines over time. We confirm that anecdotal observation that oncology knowledge has grown dramatically. The total length of National Comprehensive Cancer Network Guidelines, a "complete library" of cancer guidelines (NCCN, 2023*b*), has grown from 1,099 pages in 2002 to 5,760 pages in 2020.

Next, we measure oncologist specialization over time using a 21-year panel of Medicare claims data. Specialization is measured using an oncologist-level Herfindahl–Hirschman index (HHI), calculated over the shares of evaluation and management work for each cancer

type. The measure allows us to track subtle changes in worker specialization based on their day-to-day work within a person over time. We find that "top specialization", defined as the 90th percentile of specialization, has increased over time, implying specialized oncologists have become even more specialized. However, surprisingly, the typical oncologist has not become more specialized. Median specialization actually declines slightly over time despite substantial growth in medical knowledge over this period.

Next, we combine the guideline data with Medicare claims data to estimate how much knowledge growth different oncologists are exposed to. We construct a physician-level measure of exposure to knowledge growth. We measure exposure to knowledge growth as the increase in guideline length across all fields of cancer that the physician works in. This measure is a simulated instrument that 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 these subfields.

This simulated instrument suffers from omitted variables bias, as noted generally of formula instruments in Borusyak and Hull (2020*a*). In particular, oncologists who work in more fields are mechanically more exposed to higher levels of knowledge growth. To purge this bias, we recenter this simulated instrument using the general approach developed by Borusyak and Hull (2020*a*). Our recentering procedure eliminates the correlation between the simulated instrument and observable omitted variables, like market size, suggesting it also eliminates omitted variables bias.

We then test how exposure to knowledge growth influences oncologist specialization. Oncologists who start specialized become significantly more specialized in response to knowledge growth. For every 100-page increase in relevant guidelines, these oncologists increase their specialization by approximately 200 HHI points, a 4% increase in HHI for a 12% increase in relevant guidelines. In contrast, oncologists who start as general oncologists do not become more specialized in response to knowledge growth. We can reject an increase in HHI of greater than 30 points in response to a 100-page increase in relevant guidelines. If any-

thing, general oncologists seem to become less specialized; though not significant, the point estimates are consistently negative.

Lastly, we investigate the source of this heterogeneity through the lens of classic theories about the constraints to specialization in economics, including Becker and Murphy (1992). The median oncologist in both small and large markets is a generalist; this does not change over time. This pattern suggests that general oncologists face persistently high coordination costs. They cannot specialize in response to large increases in knowledge, even in large markets that could support greater specialization.

In contrast, specialist oncologists appear to face lower, albeit still constraining, coordination costs. These lower coordination costs enable them to specialize to an extent that market size constraints become relevant. Lower coordination costs are first evidenced by the prior result that specialized oncologists become even more specialized in response to knowledge growth. The 90th percentile of specialization, a measure of top specialization, is also rising over time. However, the increase in top specialization occurs only in the largest healthcare markets. These patterns are consistent with specialists facing an intersection of coordination cost and market size constraints.

Why does specialization matter? We find that specialists have different practice patterns than generalists. Specialists use fewer unique drugs, have more experience with each drug they prescribe, and prescribe newer drugs than generalists. These differences between specialist and generalist practice patterns have grown substantially over time. These patterns imply that, over time, specialists increasingly provide higher-quality cancer care than generalists.

The results have implications for economies of scale, geographic inequality, and healthcare access. We observe rising geographic inequality in oncologist specialization. This leads to rising geographic inequality in patients' local access to specialized healthcare. These results suggest that growth in knowledge contributes to growth in economies of scale in healthcare and high-skill labor markets.

Our work contributes to several threads of literature in economics. First, we contribute to the literature on the consequences of knowledge growth for specialization. Becker and Murphy (1992) theorized that specialization is limited by coordination costs and the extent of general knowledge. They predict that knowledge growth should increase specialization. Existing empirical evidence focuses on the effects of the historical stock or "depth" of knowledge and how that influences inventors (Jones, 2009). Our work differs by focusing on the effect of changes to the size of the currently relevant corpus of knowledge. We have yet to find any other work that measures or assesses the effects of a growing volume of currently relevant knowledge on workers' specialization in healthcare or other fields. We attribute this gap to challenges in measuring the size of knowledge, something we overcome with the use of a novel panel of clinical guidelines data.

Our findings here also contrast with the evidence on knowledge growth and specialization from Jones (2009). Jones (2009) documents empirically that a growing burden on knowledge on inventors leads to greater specialization. Similarly, the most specialized oncologists become even more specialized in response to knowledge growth. However, we find that most oncologists do not become more specialized. We attribute these contrasting results to differences between doctors and inventors. Inventors are creators of knowledge. In contrast, most oncologists, particularly general oncologists, are likely primarily users of knowledge (Jain et al., 2019). Users may bear the burden of knowledge more lightly than creators of knowledge, as the use of knowledge may be less difficult than the creation of it. In addition, as we document, users of knowledge have the option to fall behind the frontier (and continue to work) 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. Our findings suggest that knowledge growth may have different consequences for users of knowledge than previous evidence would suggest.

We also build on a small and emerging literature on economies of scale and agglomeration in healthcare. Dingel et al. (2023) document that healthcare has large market-level economies

of scale. We build on this by showing that economies of scale in healthcare appear to have grown over time and provide evidence that knowledge growth is a key mechanism.

Finally, our work contributes to an extensive empirical literature on the consequences of technological change in healthcare. Much of this literature focuses on single technologies, each with varied economic impacts on providers and patients (Skinner, 2011; Chandra, Malenka, and Skinner, 2014; Skinner and Staiger, 2015; Arrow, Bilir, and Sorensen, 2020). Our work considers the cumulative effect of technology. Existing evidence suggests that technological change in healthcare increases educational inequality in health outcomes (Glied and Lleras-Muney, 2008). We build on this by examining how technological change impacts healthcare workers and geographic inequality in care provision.

2 Data and Measurement

Our investigation combines multiple data sources to measure knowledge growth and specialization in medical oncology over time. We proxy knowledge growth with changes in the length of cancer treatment guidelines over time. We then link this to Medicare claims data to get physician-level estimates of exposure to knowledge growth. We also use Medicare claims data to measure how physician specialization evolves over time. We supplement our analysis with additional information about physicians from the Medicare Data on Provider Practice and Specialty.

2.1 Cancer guidelines data

To measure growth in knowledge, we create a novel dataset of historical cancer guidelines data from the National Comprehensive Cancer Network. The National Comprehensive Cancer Network (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*). The NCCN guidelines represent an expert consensus on approaches to cancer treatment and are updated multiple times a year to reflect the most up-to-date evidence. 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 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 oncologists have two key roles in cancer care. The first role is managing "systemic therapies", primarily drug treatments like chemotherapy and immunotherapy. The second role is to serve as the central node in a network of other specialists who may be involved in a patient's cancer care, e.g., cancer surgeons or pathologists. As such, oncologists are trained in comprehensive cancer care management and must stay up to date with the high-level recommendations for all types of cancer treatments (Popescu et al., 2014).

The NCCN guidelines are used as a proxy for the volume of relevant clinical knowledge by cancer type. The guidelines are organized anatomically, with separate documents for different cancer types, such as breast and colorectal cancer. We proxy the volume of knowledge in a given year for a type of cancer as the page length of the first set of guidelines per year for that cancer type.

This proxy relies on several key assumptions to be a valid measure of the volume of knowledge. First, this proxy assumes that each page represents an equal volume of knowledge. Reassuringly, over the main time period for our research design (2008-2014), the formatting of the guidelines was unchanged, suggesting that increases in page length result from more content, not changes in formatting. Second, the proxy assumes that any relevant knowledge not included in the guidelines is either exactly proportional to the volume of information the guideline pages contain or remains constant over time. For example, the guidelines may not include all relevant information about a chemotherapy drug, but we assume it contains a con-

sistent share of the information about each drug. Similarly, the cancer treatment guidelines are built on a foundation of pre-existing knowledge in fields like general internal medicine. We assume that the relevant volume of knowledge in general internal medicine does not increase rapidly over this time [4]. Therefore, the change in the length of guidelines reflects the full change in the volume of relevant knowledge.

2.2 Medicare claims data

We also use a 21-year panel of Medicare claims data. 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). We use a 20% sample of Medicare Part B Claims from 1999-2019.

Using this data, we construct an annual panel of oncologists. We restrict our sample to physicians whose most commonly listed specialty in a year is Hematology, Medical Oncology, or Hematology-Oncology. We further restrict our sample in a year to physicians who evaluated and managed at least 25 unique cancer patients in that year. This minimum patient count ensures that our measures of specialization are based on sufficient patient volume. For our main regression analysis, we subset to a balanced panel of oncologists from 2008 to 2014.

Next, we measure which types of cancer an oncologist manages. For this, we focus on the evaluation and management claim lines billed by each oncologist. As such, our attention is limited to instances where the oncologist managed care and decision-making for a patient, which requires high-level medical knowledge and judgment. We then assign each evaluation and management visit to a cancer type based on the patient's primary cancer type [5].

We also measure the time each oncologist spends managing each type of cancer, which we

[4] Anecdotally, standard treatments evolve much more slowly in general internal medicine than in oncology, so this assumption seems reasonable to us and the physicians we consulted

[5] We determine a patient's primary cancer as the most specific type of cancer for which they have the most diagnosis codes in a year. For example, if a patient has diagnosis codes for both "Unspecified" cancer and breast cancer, they are considered to have breast cancer. If an oncologist sees that patient, they are denoted as having managed a breast cancer patient.

later use to measure physician specialization. To measure time, we map each evaluation and management code 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, wRVUss reflect the expected time required. Accordingly, we use wRVUs Units to estimate the time that each oncologist spends managing each type of cancer.

2.3 Medicare Data on Provider Practice and Specialty

We also use the Medicare Data on Provider Practice and Specialty (MD-PPAS) to measure physician characteristics. This variable allows us to observe the physician's birth year and back out their age at the start of our panel period. It also contains the top two Tax ID numbers from which each oncologist bills. Multiple physicians can share the same tax ID number, which implies they are in the same organization or group practice. However, healthcare systems have multiple tax ID numbers or providers within the organization may all bill to personal tax ID numbers. If so, the number of physicians at the Tax ID may understate how many people are at the organization. As such, this variable provides useful but imperfect information about if a physician is a solo provider or in a bigger organization.

2.4 Measuring relevant knowledge growth for each oncologist

Next, we combine the Medicare claims and guidelines data to measure our key treatment variable, the growth in relevant knowledge for each oncologist. First, for each oncologist, we determine the set of cancer types they manage in a baseline year, in this case, 2008. The set of cancers for oncologist i in 2008 can denoted with C_i. Each cancer, c, has an increase in relevant guideline page count Δk_c over the following five years (2009 to 2014). We assume that the full guideline and change in guidelines for cancer c is relevant to a physician if they treat cancer c in the baseline year. With this assumption, we can calculate oncologist i's exposure to knowledge growth as $\Delta k_i = \sum_{C_i} \Delta k_c$.

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 would be exposed to 150 pages of knowledge growth over the following five years (2009 to 2014).

To link between cancers treated and guideline page lengths, we make a set of custom diagnosis code groupings. These groupings align with how cancers are grouped in guidelines. More detail about the creation of these groupings can be found in appendix section A.1.

2.5 Measuring physician specialization

Next, we discuss how we measure our outcome of interest: oncologist specialization. Our main measure is a Herfindahl–Hirschman Index (HHI) over the shares of time that an oncologist spends managing each type of cancer. We calculate HHI as follows, where T_{ic} is time (proxied by work RVUs) 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 the work that an oncologist performs in a year. This is contrast to the use of job title, a very common measure of specialization. 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.

3 Descriptive Patterns

In this section, we describe the characteristics of oncologists in our sample and the patterns of knowledge growth and specialization over time. We document that the volume of clinical knowledge and extreme specialization have grown over time in oncology.

3.1 Knowledge growth over time

Analysis of historical guidelines reveals that cancer treatment guidelines have become many times longer since 2002. Figure 1a displays the length of guidelines between 2000 and 2020 on average and for selected cancer types. Appendix Table C.1 lists the page count for each type of guideline in 2002 and 2020. The total length of guidelines has increased over five-fold in this period, from 1,099 pages in 2002 to 5,760 pages in 2020.

Many factors may drive this growth in guidelines, but at least one is an explosion in the number of drugs available to treat cancer. The number of FDA-approved anti-cancer drugs increased from 90 in 2002 to 243 in 2020. Figure 1b displays the number of U.S. Food and Drug Administration-approved anti-cancer drugs over time. Data on anti-cancer drugs is from Pantziarka et al. (2021) and uses Anatomical Therapeutic Chemical Level 5 classifications to define individual drugs. Not only has the total number of drugs been increasing steadily over time, but the rate at which new drugs are introduced has been accelerating.

3.2 Oncologist characteristics

Next, we describe summary statistics for our balanced panel of physicians from 2008 to 2014. Table 1 shows summary statistics for key variables of interest for a balanced panel of 4,861 oncologists. On average, they see 69 unique cancer patients in the baseline year, 2008. On average, they see 15 distinct types of cancer, but some see as few as one type of cancer, and others see as many as 27 types. Their average age at the beginning of the panel is 50 years

old. The mean oncologist HHI is 2,737. Over half of oncologists are in the largest one-fifth of hospital referral regions by population. On average, 20 other oncologists work at their Tax ID numbers, but this can be as low as zero and as high as 245.

In 2009, the total length of NCCN guidelines is 1789 pages, but only 1,159 pages are relevant to each oncologist, on average. Over the following five years, the total and relevant number of guidelines more than double, increasing by 1,745 and 1,087 pages respectively, to 3534 total and 2,246 relevant pages on average.

3.3 Trends in oncologist specialization over time

Next, we characterize the distribution of oncologist specialization over time. Figure 2 shows the 10th, 25th, 50th, 75th, and 90th percentile of oncologist HHI between 1999 and 2019. It also plots the HHI overall, using all evaluation and management services from oncologists. 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 our main balanced panel in 2014, which prevents coding changes from contaminating our estimates of change in HHI.

Second, the typical oncologist has not become substantially more specialized over time. Median specialization and below is flat or declining slightly over this period. The distribution's 10th, 25th, and 50th percentiles are all close to each other and the HHI of the field. This result suggests that most oncologists are "general" oncologists who see a similar distribution of cancer patients as the field overall. This lack of growth in specialization for the median oncologist is surprising, given the explosive 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 738 HHI points from 1999 to 2014 and another 328 points from 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. For our main estimation, we split our sample into generalist and specialist oncologists to test if they respond differently to knowledge growth. Of note, although we refer to specialist and generalist oncologists, all oncologists are considered medical subspecialists, highlighting the extremes of specialization we study here.

3.4 Differences between more and less specialized oncologists

Next, we describe the characteristics of specialist and generalist oncologists. We define generalists as those below the 75th percentile of physician HHI in the initial year of the panel. We define specialists as those at or above the 75th percentile of the distribution. Table 2 shows the characteristics of each group in our balanced panel.

Substantial differences are present between these two groups in the baseline year for the panel. In 2008, specialists have an average HHI of 4,637, over double that of the generalists, who have an HHI of 2,104 on average. Specialists are far more likely to be academic physicians. 32% of specialists appear to be academic physicians, compared to only 8% of generalists. (See Appendix A.2 for more details on how we proxy academic status). Specialists are also more concentrated in large markets: 63% of them are in the largest markets as compared to 50% of general oncologists. In addition, they work in larger organizations: On average, specialists work in places with 26 other oncologist colleagues, compared to 18 for generalists.

These empirical differences are consistent with large firms, large markets, and academic institutions all facilitating specialization. For example, only large markets and organizations may furnish a sufficient volume of kidney cancer patients to support a kidney cancer specialist. Large organizations may also lower coordination costs between oncologists. In speaking with oncologists, a key barrier to specialization is that some cancers are more profitable to treat than others, disincentivizing patient sharing. Large firms provide the financial and legal

structures to share patients and the resultant revenue. Academic institutions may provide additional non-pecuniary incentives to become specialized. For example, academic tenure and promotion systems often incentivize physicians to develop clinical and research expertise in a narrow domain.

4 Identification

The following section discusses our identification strategy to test if knowledge growth increases specialization. To test this, we regress changes in specialization on a proxy of exposure to knowledge growth at the physician level. Our proxy of exposure to knowledge growth is a formula instrument. We find it suffers from omitted variables bias, as is common with formula instruments. We remove this bias by recentering the instrument to account for expected exposure to knowledge growth.

4.1 Estimating Equation

First, we 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. The main estimating equation is as follows:

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

We estimate this equation 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 our 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), which is a simulated or "Bartik-style" instrument. 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. This variation is analogous to variation in industry shares in a Bartik instrument. 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. This variation is analogous to variation in industry-level shocks in a Bartik instrument.

4.2 Recentering instrument to correct for omitted variables bias

Importantly, simulated instruments generally suffer from omitted variables bias, which must be corrected for to estimate accurate causal effects (Borusyak and Hull, 2020*b*). In this context, the omitted variables bias arises from the fact that physicians who treat more types of cancer 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. The breast cancer oncologist may differ in many important ways from the oncologist who treats both cancers. For example, the breast cancer oncologist may be more likely to work at a large healthcare market because it has the patient volume to support a breast cancer oncologist. Oncologists in large markets may have different specialization trends than those in solo practice for reasons unrelated to knowledge growth. This variation in market size will be correlated with both exposure to knowledge growth (Δk_i) and with change in specialization (ΔHHI_i), leading to omitted variables bias in the

estimate of β.

We address this omitted variables bias using the recentering approach proposed by Borusyak and Hull (2020*b*). This approach removes the observed non-random exposure to knowledge growth that contributes to omitted variables bias. In our example, the non-random exposure results from the breast cancer specialist being predictably less exposed to knowledge growth than the generalist.

To remove the non-random variation, we estimate expected exposure to knowledge growth ($\mathbb{E}[\Delta k_i]$) and remove it from realized exposure to knowledge growth Δk_i. Our measure of realized exposure to knowledge growth, Δk_i, can be decomposed into two components.

$$\Delta k_i = \underbrace{\mathbb{E}[\Delta k_i]}_{\text{OVB}} + \underbrace{\Delta \tilde{k}_i}_{\text{Remaining Variation}}$$

The first component, $\mathbb{E}[\Delta k_i]$, is the expected exposure to knowledge growth and contains the omitted variables bias. The second component is the remaining variation and is assumed to be exogenous and uncorrelated with other factors that impact specialization except for knowledge growth.

The above expression can be rearranged as follows:

$$\begin{aligned}
\Delta k_i &= \mathbb{E}[\Delta k_i] + \Delta \tilde{k}_i \\
\Delta \tilde{k}_i &= \Delta k_i - E[\Delta k_i] \\
&= \sum_{C_i} (\Delta k_c - \mathbb{E}[\Delta k_c])
\end{aligned}$$

We can estimate the residual exposure to knowledge growth, $\Delta \tilde{k}_i$, by subtracting the expected exposure ($\mathbb{E}[\Delta k_c]$) from the realized exposure (Δk_i). Realized exposure (Δk_i) is calculated exactly from our data. Calculating expected exposure to knowledge growth ($\mathbb{E}[\Delta k_c]$) requires us to make assumptions about the expected knowledge growth per cancer subtype. In the next section, we will discuss how we construct this expectation and the

implications for the substance of the identification assumptions.

Once we have this expectation, we can then substitute our realized exposure, Δk_i, with the residual exposure, $\Delta\tilde{k}_i$, for an unbiased and efficient estimate of β, as below:

$$\Delta HHI_i = \beta\Delta\tilde{k}_i + \theta X_i + \epsilon_i$$

The formal identification assumption here is that this residual 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 we form our expectation of knowledge growth by cancer type ($\mathbb{E}[\Delta k_c]$).

4.3 Estimating Expected Knowledge Growth by Cancer Type

This section discusses different assumptions around expected knowledge growth by cancer type and the corresponding identification assumptions.

Our first approach assumes all cancer fields have the same expected knowledge growth. Specifically, we assume that $\mathbb{E}[\Delta k_c] = \hat{\mu}$ where $\hat{\mu}$ is the mean page increase across cancer types, weighted by the number of oncologists who manage that cancer type. Let n_i denote the number of cancer types that an oncologist treats. $\Delta\tilde{k}_i$ is estimated as follows:

$$\begin{aligned}\Delta\tilde{k}_i &= \sum_{C_i}(\Delta k_c - \hat{\mu}) \\ &= \Delta k_i - n_i\hat{\mu}\end{aligned}$$

This approach is equivalent to controlling for the number of cancers an oncologist treats (Borusyak and Hull, 2020*b*). Essentially, it compares doctors who treat the same number of cancers. The variation in $\Delta\tilde{k}_i$ arises because, holding cancer set size constant, oncologists vary in the composition of the set of cancers they treat, and some cancers have more innovation than others.

Critically, this approach assumes that cancer-level knowledge shocks are unpredictable and uncorrelated with physician sorting. We are particularly concerned that cancer-level knowledge shocks are not entirely unpredictable. A key concern is that more common cancers likely draw more investment in innovation (Acemoglu and Linn, 2004). In addition, more common cancers may also support more oncologist specialists since it is easier to find sufficient patients. Addressing this potential bias motivates further refinement in our approach.

Our second approach allows expected knowledge growth to vary with physician scope and cancer characteristics. Acemoglu and Linn (2004) find that conditions with more potential patients had more innovation. Budish, Roin, and Williams (2015) find that cancers with surrogate endpoints have more innovation. We incorporate both of these factors into our estimate of expected knowledge growth, $\mathbb{E}[\Delta k_c]$

Specifically, we predict the expected amount of knowledge growth for a cancer type based on these cancer characteristics using a Poisson regression. 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 surrogate endpoints per Budish, Roin, and Williams (2015) ($e_c = \mathbb{1}(\textit{surrogate endpoints})$). We include both of these in a Poisson maximum likelihood regression to predict the expected page count increase for each type of cancer, as follows:

$$E[ln\Delta k] = \delta_0 + \delta_1 e_c + \delta_2 p_c + \delta_3 e_c p_c$$

Note that this equation also controls for knowledge shocks shared across all cancer types (loaded on δ_0). Table 3 shows the estimate for the above regression.

Denote the predicted knowledge growth from this regression as $\Delta\hat{k}_c$. Then $\Delta\tilde{k}_i$ is estimated as follows:

$$\Delta\tilde{k}_i = \sum_{C_i}(\Delta k_c - \Delta\hat{k}_c) \tag{1}$$

This approach controls for the number of cancers an oncologist treats, weighting each cancer

by the expected innovation. As such, this recentered measure of exposure to knowledge growth removes variation from both physician scope and cancer characteristics.

This approach makes several key assumptions. First, it assumes that the volume of patients, presence of secondary endpoints, and shared knowledge shocks are the central factors correlated to highly predictable variation in innovation across cancer types. Empirically, this assumption is supported by the observation that these factors alone explain 75% of the variation in guideline page count increases (see Table 3). This approach also assumes that any residual variation in knowledge growth by cancer type is unrelated to oncologists' specialization decisions except through the direct effect on specialization. We attribute the residual variation to randomness and uncertainty in the research and development process, e.g., some clinical trials fail and others succeed.

4.4 Correlation of Exposure to Knowledge Growth with Omitted Variables

In the next section, we empirically test if this recentering approach purges measurable omitted variables bias. We find that the unadjusted measure strongly correlates with observed omitted variables. However, the recentered measure removes this correlation when it accounts for both physician scope and cancer characteristics. This recentered measure is also a strong instrument for the unadjusted measure. Both of these findings make this recentered measure our preferred measure.

First, we test if our measures of exposure to knowledge growth correlate with observed omitted variables. We follow the strategy in (Borusyak and Hull, 2020*b*) to test for such correlations. We regress our unadjusted measure and recentered measures of physician-level exposure to knowledge growth on key physician characteristics. These characteristics are 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), and count of unique patients seen, all in the baseline year. These variables are all sources of concern

about omitted variables bias.

Table 4 shows the results for regressions of the measures of exposure to knowledge growth on these omitted variables. Appendix tables C.2 and C.3 show the same regressions separately for specialist and generalist oncologists.

Column 1 in Table 4 shows the coefficient estimates for a regression of these observed omitted variables on the unadjusted measure of knowledge growth. 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 -12.8 pages ($\sigma = 1.7$) 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 that bias our estimate of β. Overall, the included omitted variables explain 38.5% of the variation in exposure to knowledge growth. These results substantiate concerns about omitted variables bias in formula instruments and suggest that the recentering procedure is necessary for unbiased estimation.

Columns 2, 3, and 4 in Table 4 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, our recentering ha purged all significant correlation between exposure to knowledge growth and the observed omitted variables. For example, the correlation with market size is now smaller and statistically indistinguishable from zero at .008 ($\sigma = .518$). Overall, the observed omitted variables now explain only about .5% of the total variation in the fully recentered measure as compared with 38.5% of the variation in the raw measure, a decrease of 99%.

Overall, the final recentering procedure essentially eliminates the correlation between the exposure to knowledge growth and observed omitted variables. The fact suggests that

the correlation with unobserved omitted variables has also largely been eliminated. This recentered measure, which accounts for physician scope and cancer characteristics, is our favored measure of exposure to knowledge growth and is the measure used in following analyses. Going forward, this measure is referred to as $\Delta\tilde{k}_i$, rather than the more cumbersome $\Delta\tilde{k}_i^{physician,cancer}$.

4.5 Instrument Strength

Next, we test if our recentered measure of knowledge growth is a "strong" instrument for the unadjusted measure. We have removed variation in this measure by recentering; how much remains?

To assess this, we regress our recentered measure ($\Delta\tilde{k}_i$) on our unadjusted measure. (Δk_i) in a "quasi-first stage" regression. This regression is not a true first-stage regression because we do not use the predicted values in a second-stage regression to estimate the coefficient of interest. Rather, we use the recentered measure ($\Delta\tilde{k}_i$) directly to measure the coefficient of interest, per Borusyak and Hull (2020*b*). However, the "quasi-first stage" regression can inform us about the extent of the variation that remains after recentering.

Figure 3 shows a binned scatter plot of the regression of the recentered measure ($\Delta\tilde{k}_i$) on the unadjusted measure (Δk_i). Appendix figures C.2a and C.2b show the same results separately for specialist and generalist oncologists. The recentered measure is a highly significant predictor of the unadjusted measure ($\beta = 0.77$; $\sigma = .04$). The overall R^2 of the regression is 5.6%, which suggests that the recentered measure removes 94.4% of the variation in the unadjusted measure. However, the F statistic is 306, suggesting that the recentered measure is a "strong instrument" for the unadjusted measure by traditional metrics.

5 Effect of Knowledge Growth on Specialization

In this next section, we test for the effect of knowledge growth on specialization for specialist and generalist oncologists. All estimates presented here use the recentered measure of exposure to knowledge growth in a balanced panel of oncologists. We estimate the effect of exposure to knowledge growth from 2009 to 2014, a period when total guideline pages doubled, from 1789 to 3534. We find that ex-ante specialized oncologists become more specialized, but general oncologists do not.

5.1 Effects for specialized oncologists

Figure 4 shows binned scatter plots and key regression estimates for the univariate relationship between the exposure to knowledge growth and change in specialization separately for specialist and generalist oncologists. Figure 4a shows this relationship for oncologists who start in the top 25% of the initial specialization distribution, or "specialist" oncologists. The relationship between exposure to knowledge growth and increased specialization is large and significant. For every 100-page increase in relevant guidelines, these oncologists increase their specialization by 219 HHI points ($\sigma = 75$). Relative to baseline means, this represents a 4% increase in HHI for a 12% increase in relevant guidelines. Table 5 shows the regression table of results, adding fixed effects for age bin, hospital referral region, and tax ID size. The coefficient estimates and standard errors are essentially unchanged after age and region-fixed effects are added. The coefficient estimate is slightly attenuated after adding fixed effects for Tax ID size, falling just below the threshold for significance at the 10% level (137.5, $\sigma =$ 84.0). This finding could suggest that organization size mediates some of this relationship between knowledge growth and specialization.

Notably, the increases in specialization are driven by intensive margin changes in specialization. We test the effect of knowledge growth on extensive margin specialization - the count of unique types of cancer treated. Appendix table C.4 displays the results for a regres-

sion of the change in the unique types of cancer treated by an oncologist on that oncologist's exposure to knowledge growth for specialized oncologists. The estimate is statistically indistinguishable from zero. The estimates are precise enough to rule out that exposure to 100 more pages of guidelines reduces the count of cancer types managed by one or more. Specialized oncologists do not become more specialized on the extensive margin; changes on the intensive margin must drive the observed changes in HHI. In other words, specialized oncologists focus their practice more on some types of cancer but do not entirely stop seeing other types. This pattern could reflect oncologists continuing to see established patients but limiting their new patients to those with cancers in a narrower area of specialization.

5.2 Effects for generalist oncologists

Next, we consider how knowledge growth impacts general oncologists. Figure 4b shows the relationship between exposure to knowledge growth and HHI for oncologists starting in the bottom 75% of the initial specialization distribution, or "general" oncologists. In contrast to the results for specialists, general oncologists do not become specialized in response to knowledge growth. The estimated effect on HHI is more precise than the effects for specialists and statistically indistinguishable from zero at $\beta = -17.4$ ($\sigma = 23.3$). Table 6 shows the regression table of results, adding fixed effects for age bin, hospital referral region, and tax ID size. After adding these fixed effects, the coefficient estimates and standard errors are again unchanged. The coefficient estimate is still statistically indistinguishable from zero. In addition, there is no effect on the count of cancer types treated (results in figure C.5)

5.3 Robustness

These relationships between knowledge growth and specialization, as measured by HHI, are generally robust to implementing a stricter threshold to count an oncologist as exposed to a field, changing the sample of evaluation and management codes, and equalizing measurement error across physicians.

In the main analysis, a physician is considered exposed to knowledge growth in a cancer field if they are observed to manage at least one patient with that cancer type in the baseline year in the 20% Medicare Part B sample. As a robustness check, the threshold is raised to two patients and then three patients with a cancer type. As before, knowledge growth significantly increases specialization for oncologists who are already specialized. The effect is closer to zero for oncologists who are not specialized (Table C.7).

In the main analysis, the sample consists of all evaluation and management codes. As a robustness check, we test if the results differ when considering only certain subtypes of evaluation and management codes. First, we rerun the main regression with only new patient evaluation and management codes. Changes to specialization may be reflected first as a change in the new patients that oncologists see. Unfortunately, this restriction shrinks the sample size to one-quarter of its original size (even after relaxing the patient count minimum to 11) and inflates the standard errors such that the estimate is not informative.

Next, we rerun the main regression with only outpatient evaluation and management codes (dropping inpatient codes). In academic centers, patients are typically better sorted to the appropriate physician in the outpatient setting than in the inpatient setting because of the non-urgent nature and higher volume of outpatient care. We expect specialization responses to persist (or even be driven) by outpatient codes. We find exactly that result for specialists. Generalists continue not to respond to knowledge growth.

Lastly, we consider how measurement error in our measure of specialization ($\hat{HHI}_i$) impacts our main findings. Appendix B addresses this issue in detail. Here, we summarize the key points. Recall that we do not observe all patients nationally in our 20% sample of Medicare data, so $\hat{HHI}_i$ is calculated over only a subset of the patients a physician treats in a year. Not observing all patients leads to measurement error in $\hat{HHI}_i$, our outcome variable. This measurement error has a positive expected value and is larger for physicians with smaller patient volumes. Fortunately, patient volume is balanced across our recentered measure of exposure to knowledge growth (recall Table 4), which implies that this measurement error

is uncorrelated with our treatment variable (recentered exposure to knowledge growth). As a robustness check, we replicate the key results with a measure of $\hat{HHI}_i$ 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. Even with this change, the time trends in specialization (Appendix Figure C.1) show the same rise in top specialization as Figure 2. In addition, the estimates of the effect of knowledge growth on specialization (Appendix Tables C.8 and C.9) are statistically indistinguishable from the estimates in Tables 5 and 6.

5.4 Discussion

In summary, specialist oncologists respond to knowledge growth by becoming more specialized. However, generalist oncologists do not become more specialized — they just manage with a larger volume of knowledge than before. Through the lens of Becker and Murphy (1992), these different responses imply that specialist oncologists face higher returns to specialization and/or lower coordination costs than generalists. Recall from section 3 that specialists are more concentrated in large markets, large firms, and academic institutions, all of which may lower coordination costs or raise returns to specialization. These differences in returns to specialization and/or coordination costs likely incentivized these oncologists to become specialists ex-ante and enabled them to respond to knowledge growth ex-post. In the next section, we evaluate the extent to which differing constraints to specialization drive these differential responses.

6 Constraints to Specialization

Why do specialist and generalist oncologists respond differently to knowledge growth? To unpack this, we assess the constraints to specialization for specialists and generalists in the context of classic theories of specialization. General oncologists are constrained by extremely

high coordination costs. In contrast, specialist oncologists face lower coordination costs but encounter market size constraints.

6.1 Theories of constraints to specialization

This section discusses possible reasons for the heterogeneous response to knowledge growth in the context of classic theories in economics about the constraints to specialization. We focus on two key potential constraints to specialization: market size and coordination costs.

Smith (1776) was the first in a long line of many economists to observe that specialization is often limited by the extent of the market. If the extent of the market limits specialization, then workers should not become more specialized in response to knowledge growth. Instead, oncologist specialization should primarily increase with market size. Figure 5a shows this theoretical relationship between market size and specialization.

Becker and Murphy (1992) highlight that coordination costs between specialized workers may also limit specialization. Specialization may make individual workers more productive at specific tasks. However, this splits tasks among more workers, leading to higher coordination costs, e.g., communication overhead or incentive compatibility challenges. These coordination costs may constrain specialization, regardless of market size. If coordination costs are a binding constraint in our setting, the oncologist specialization will remain flat across market size. Figure 5a shows this theoretical relationship between market size and specialization.

In the Becker and Murphy (1992) framework, workers should become more specialized in response to knowledge growth; knowledge growth raises the returns to specialization and thus increases the optimal degree of specialization. In sufficiently small markets, market size constraints may also bind. In that case, specialization will increase in response to knowledge growth only in large markets with sufficient market size to support the new, higher optimal level of specialization. Figure 5a also shows this theoretical prediction.

Lastly, even within the Becker and Murphy (1992) framework, workers may not respond

to knowledge growth appreciably if the returns to specialization are very low or coordination costs are very high. In that case, specialization will remain constant across space, even with substantial knowledge growth.

6.2 Evidence of constraints for general oncologists

Each of these constraints to specialization produces differing predictions about the degree of specialization over market size and time. We ascertain which theories are most consistent with the data and, thus, which constraints might drive the observed heterogeneous responses to knowledge growth. Market size is measured as the market population (using hospital referral regions) in the relevant year. We then graph the median and 90th percentile of specialization by market size in 1999 and 2014 in Figure 6. We also regress the median and 90th percentile of specialization on market size in Table 7.

The median oncologist appears highly constrained by coordination costs. Median specialization remains flat across the distribution of market size in 1999. The median oncologist is both large and small markets is a generalist. This pattern is highly consistent with coordination costs limiting the specialization for the typical oncologist, as in Becker and Murphy (1992). This pattern remains unchanged over time: it is the same in 2014 despite explosive growth in the volume of knowledge over this time period. That suggests that either the returns to specialization are quite low or coordination costs are quite high for general oncologists.

We find it somewhat unlikely that low returns to specialization drive oncologists' lack of response to knowledge growth. Top-ranked, academic cancer hospitals largely employ cancer specialists Gesme and Wiseman (2011); Graham et al. (2021). This pattern suggests that specialization has meaningful returns to quality. There may be low returns to specialization, but it seems unlikely to us, given that oncology is a rapidly changing and highly knowledge-intensive medical field. Instead, we suspect high coordination costs drive these patterns

[6].

6.3 Evidence of constraints for specialized oncologists

Notably, the patterns for highly specialized oncologists suggest a different set of constraints to specialization. Figure 6b shows the 90th percentile of physician specialization by market size in 1999 and 2014. We refer to the 90th percentile of specialization as the "top specialization" that a market can support. In 1999, top specialization increases with market size, from around 3000 HHI points in the smallest markets to just over 5000 HHI points in the largest. By 2014, this gradient has grown significantly steeper. Consequently, the largest markets support even greater specialization than before (over 6000 HHI points on average in the largest markets).

Patterns of specialization over time by market size are broadly consistent with specialist physicians facing an interaction of coordination cost and market size constraints. As knowledge grows, the optimal specialization increases when considering only coordination costs. However, this new optimal specialization may only be feasible in the largest markets. As a result, knowledge growth can lead to increasing specialization only in large markets, as observed. This pattern is also consistent with growing economies of scale.

The effect of knowledge growth on specialization also has implications for patient access to healthcare. Patients in large markets have increased local access to sub-specialist care, while local access has stagnated in smaller markets. These changes could worsen geographic differences in health outcomes if subspecialized oncologists provide higher-quality care.

[6]From talking with oncologists, the coordination costs arise primarily from incentive compatibility and contracting difficulties. For example, they report that general oncologists are incentivized to keep every patient that arrives at their office, regardless of cancer type, because marginal cancer patient is usually profitable.

7 Technology Use and Specialization

Does specialization impact the care that oncologists provide? We find that specialized oncologists use fewer unique drugs and have more experience in each drug they prescribe. They also tend to prescribe newer drugs to patients. Importantly, these differences between specialist and generalist oncologists have grown substantially over time.

7.1 Data and Measurement

In this section, we focus on the anti-cancer drugs that oncologists use. 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.

We 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], we restrict it to one for this analysis. We use data on oncologists' prescriptions from a 20% sample of Medicare Part B and Part D claims. Our data starts in 2008, the first year in which Medicare Part D claims use National Provider Identifiers (NPIs) to identify the majority of prescribers. To ensure these features are calculated over a meaningful volume of patients, 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.

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

7.2 Scope of Technology Use

First, we characterize the relationship between oncologists' specialization and the scope of technologies they use. The scope of technology used is proxied by the number of unique anti-cancer drugs that an oncologist is observed to prescribe in a year. We also document how that evolves over time when the number of technologies (in this case, new anti-cancer drugs) also rapidly increases.

Figure 7 shows the results. In the first year of the panel, differences in scope are significant but relatively small. The most specialized oncologists (75th—99th percentile of HHI) prescribe 12.2 unique drugs, while the most general oncologists (1st—50th percentile of HHI) prescribe 13.2 unique drugs per year on average.

Over time, these differences grow substantially. In 2019, the least specialized oncologists (1st to 50th percentile of specialization as measured by HHI) were observed to prescribe 12.4 unique anti-cancer drugs on average. Oncologists in the middle (50th to 75th percentile) prescribe 10.8 unique drugs on average. Finally, the most specialized oncologists (75th-99th percentile) prescribe 9.1 unique drugs on average, 27% fewer than the most general group. These differences are all significant. Overall, specialization narrows the scope of technologies that oncologists use.

One limitation is that we do not observe all the anti-cancer drugs each oncologist prescribes. We assume that if an oncologist prescribes more unique drugs in a 20% subsample of Medicare patients, they likely prescribe more unique drugs across all their patients. However, we observe fewer patients on average for specialized oncologists (see table 2). Fewer patients could lead to a reduced observed count of unique drugs due to a mechanical relationship between the number of patients seen and the number of drugs prescribed. We eliminate this mechanical relationship by selecting 11 random patients per oncologist without replacement and repeat the above analysis. The results are in C.3. The same patterns hold. In the early years of the panel, specialist and generalist oncologists prescribe similar counts of unique drugs. However, over time, they diverge, with specialist oncologists narrowing the scope of

drugs they use in a year relative to generalists.

7.3 Experience

Next, we ask if specialization enables oncologists to be more experienced with the drugs they prescribe. To answer this question, we estimate oncologists' experience with each drug, measured as the count of patients each oncologist prescribes a given drug to in a year. We focus on experience within a year since physician human capital appears to decay rapidly (Hockenberry and Helmchen, 2014; Facchini, 2022). We then normalize experience to relative experience, with a mean value of one within each year.

Figure 8 displays average relative experience by physician specialization over time. In 2008, the most and least specialized oncologists have similar levels of experience with the drugs they prescribed. By the end of the panel, the most specialized oncologists have significantly more (roughly 25% more) experience with the drugs they prescribe patients than the most general oncologists. As these drugs are typically complex and high-risk, more experience with a drug plausibly translates into better clinical management of the drug and potentially better patient outcomes.

7.4 Drug Age

Finally, we ask if specialization helps oncologists to keep up with new technology. To do so, we estimate the average age of the drugs oncologists prescribe by degree of specialization over time. Drug age is calculated as years since the first FDA approval for the ATC code for any indication, dosage, or delivery form[8]. We calculate the average age of the drugs that each oncologist prescribes, weighted by how many patients they prescribe the drug to in a year. We then chart the average drug age by oncologist by degree of specialization over time.

Figure 9 displays the results for all anti-cancer drugs. In the early years of the panel,

[8]This approach raises the average drug age relative to a strategy that accounts for new indications. However, we cannot observe which indication a drug was given for (or if it was given off-label, without an FDA-approved indication).

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, about 2.7 years older on average. These results are consistent with specialization enabling oncologists to adopt new drugs more quickly.

Of note, generalist oncologists initially appear not to update their drug bundle. 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 C.5 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. However, as before, specialists start to update their drug bundle faster over time than generalists, and pull ahead in the use of new technology.

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 outcomes? Unfortunately, we do not observe sufficient detail about each patient's cancer (e.g., stage, biomarkers) in our data to plausibly answer this question. However, several case studies from the medical literature are broadly consistent with specialists providing better care. 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, these results are consistent with specialists providing higher-quality care.

8 Conclusion

In this paper, we examine how this growth in knowledge impacts expert specialization. We study this in medical oncology, where we document large and rapid increases in anti-cancer drugs and the corpus of clinically relevant knowledge.

Studying a balanced panel of oncologists, we find that ex-ante specialized oncologists become significantly more specialized in response to knowledge growth. This growth in specialization is concentrated in the largest healthcare markets. As such, top specialization has increased substantially in the largest markets and stagnated in the smallest ones. These results imply that specialized oncologists face a combination of market size and coordination cost constraints to specialization.

We also observe that most oncologists are general oncologists and do not become more specialized in response to knowledge growth. The median oncologist in both large and small markets remains a generalist, even in the face of rapid growth in knowledge in the field of oncology. These findings suggest that coordination costs for many oncologists are very high, limiting their ability to specialize even in large markets and to achieve gains from specialization.

The growing dispersion in specialization has implications for 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. In addition, these differences in drug prescribing have grown substantially over time, consistent with growing quality differences between specialists and generalists over time.

Collectively, these results provide evidence that knowledge growth increases economies of scale and geographic inequality in specialized medical care and healthcare labor markets.

9 Appendix

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.

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.

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.

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. 2020*a*. "Non-random exposure to exogenous shocks: Theory and applications." National Bureau of Economic Research.

Borusyak, Kirill, and Peter Hull. 2020*b*. "Non-random exposure to exogenous shocks: Theory and applications." National Bureau of Economic Research.

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.

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

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

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

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

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

Graham, Laura S, Alexandra O Sokolova, Ali Raza Khaki, Qian "Vicky" Wu, and Nancy E Davidson. 2021. "Gender Differences in Faculty Rank and Subspecialty Choice among Academic Medical Oncologists." *Cancer investigation*, 39(1): 21–24.

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

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

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

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

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.

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.

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, volume 1.* Vol. 1, Oxford: Clarendon Press.

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.

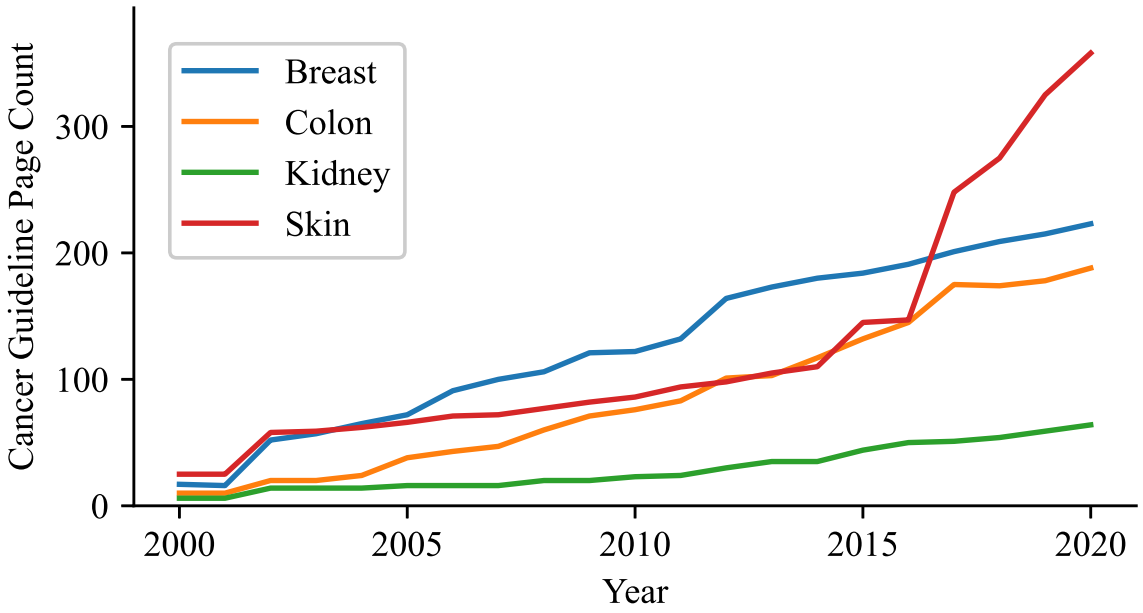
Voss, Rachel K, Yi-Ju Chiang, Keila E Torres, B Ashleigh Guadagnolo, Gary N Mann, Barry W Feig, Janice N Cormier, and Christina L Roland. 2017. "Adherence to National

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.

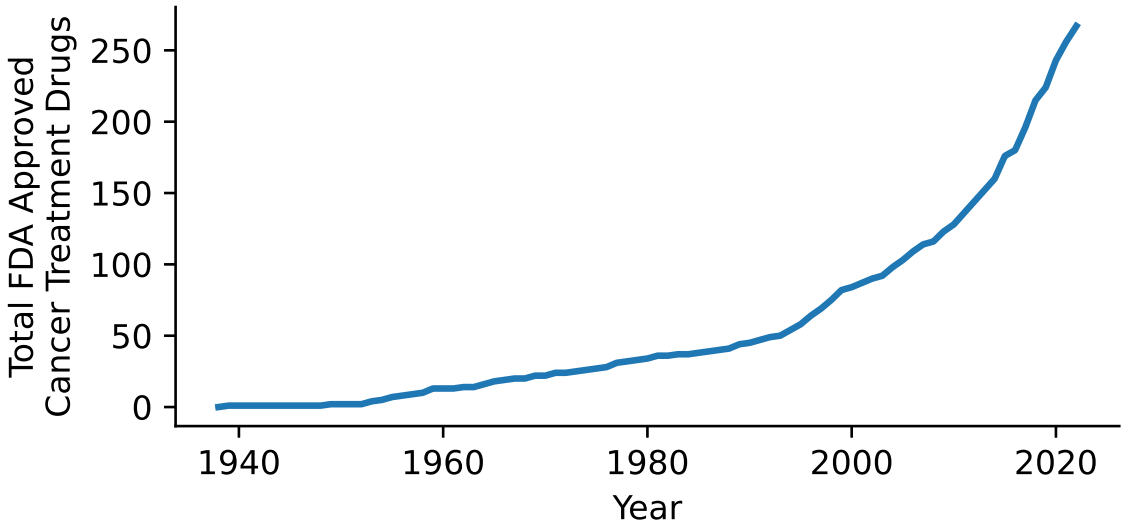
Wallis, Claudia. 2023. "A Wave of New Cancer Treatments Challenges Community Oncologists to Keep Up." *Scientific American*. Accessed: 2024-06-07.

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

(a) Oncology treatment guidelines have become considerably longer

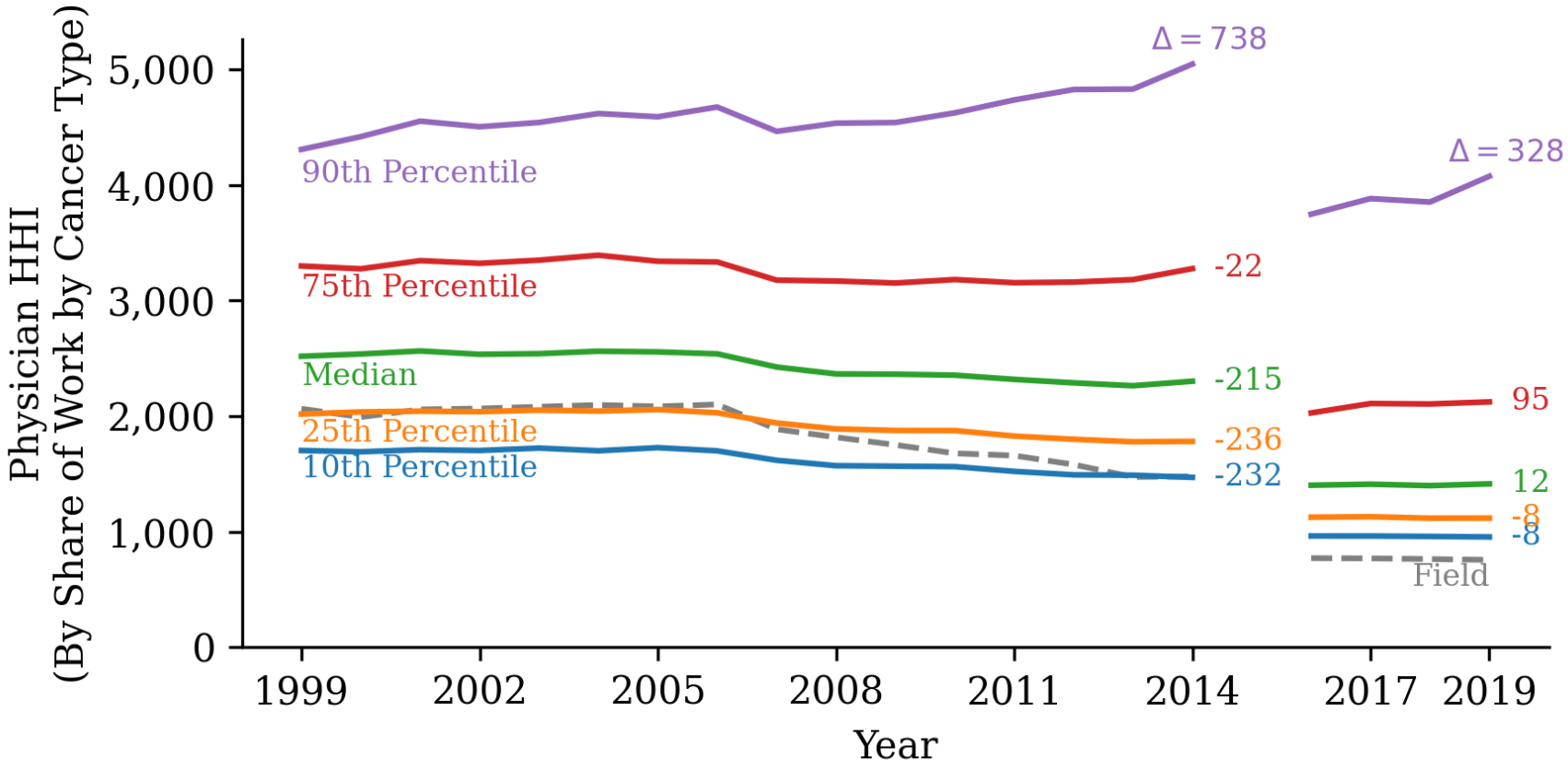


(b) The number of FDA-approved anti-cancer drugs has also increased substantially



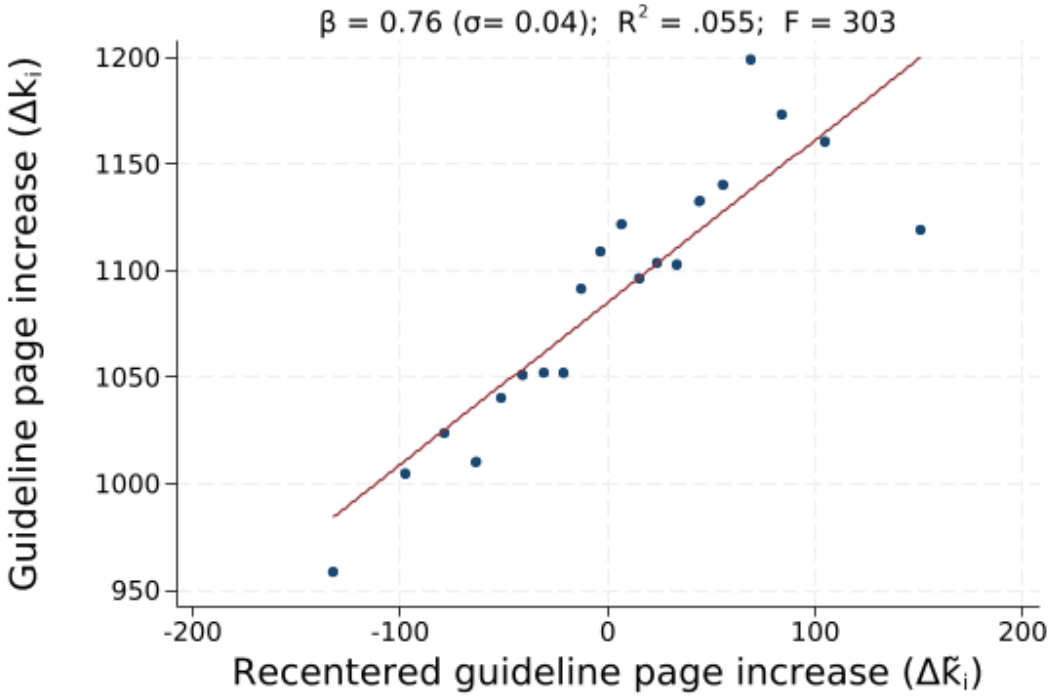
Notes: Panels (a) and (b) show the growth in cancer guideline length and pharmaceutical treatments over time. Panel (a) displays the page count of the first set of treatment guidelines in each year from the National Comprehensive Cancer Network Clinical Practice Guidelines in Oncology. Panel (b) displays the number of unique anti-cancer drugs, where unique 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 this panel is from Pantziarka et al. (2021).

Figure 2: Specialized Oncologists are Growing More Specialized 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, oncologists, or hematologists who saw at least 25 unique cancer patients in the year in a 20% sample of Medicare Part B claims. 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. The 10th, 25th, 50th, 75th, and 90th percentile are displayed and labeled. The dashed grey line represents the HHI of the field overall. 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. Overall, the 10th, 25th, and 50th percentile of oncologists see a roughly representative distribution of patients, suggesting they are "general" oncologists. As such, most oncologists remain generalists over this period. However, the most specialized oncologists, those near the 90th percentile of the distribution, become substantially more specialized over time.

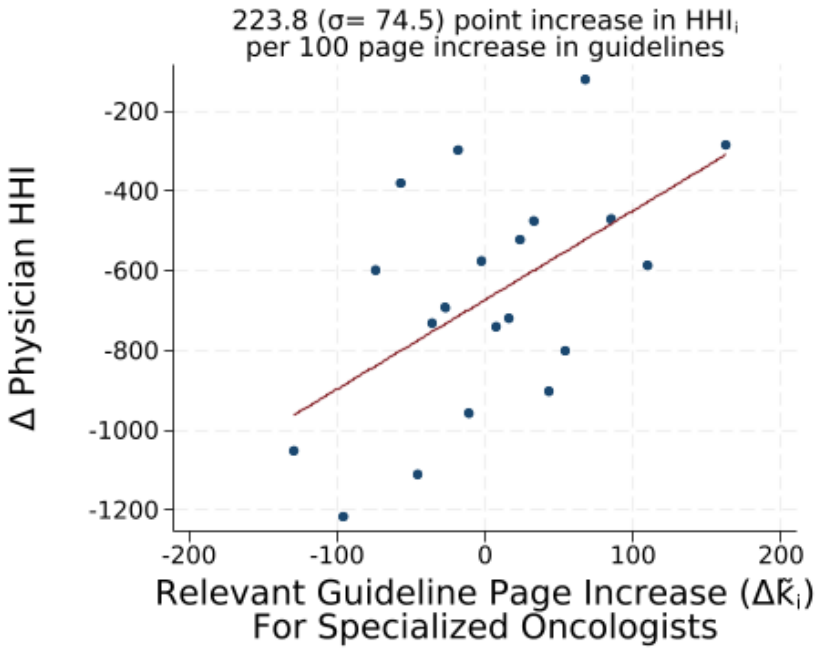
Figure 3: Quasi first stage suggests recentered instrument ($\Delta\tilde{k}_i$) is a strong "instrument" for unadjusted instrument (Δk_i)



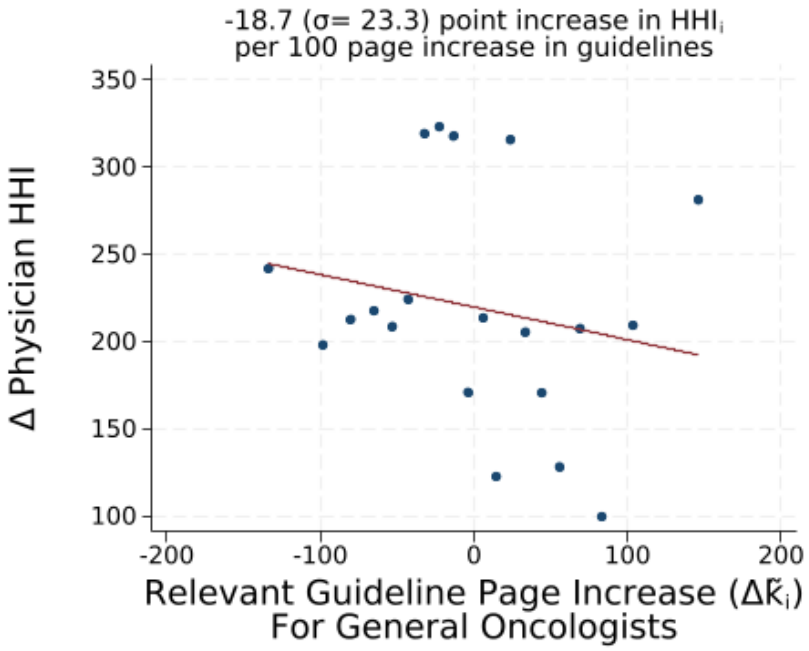
Notes: This figure shows the results for a quasi first stage regression of the recentered measure of exposure to knowledge growth on the unadjusted measure. It displays a binned scatter plot and key regression outputs for the relationship between the raw and recentered measure of exposure to knowledge growth for physicians. Regression standard errors are robust. The recentered measure of exposure to knowledge growth controls for variation in physician scope and field expected knowledge growth. In the recentered measure, the predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. Raw exposure to knowledge growth is calculated using the cancer types a physician treated in 2008 and the subsequent growth in knowledge in those fields over the next five years (2009-2014), based on realized changes in guideline lengths. 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 evaluated and managed at least 25 unique cancer patients in a 20% sample of Medicare Part B claims in 2008, 2009 and 2014.

Figure 4: Knowledge growth increases specialization only for those who start specialized

(a) Knowledge growth increases specialization for oncologists who start specialized



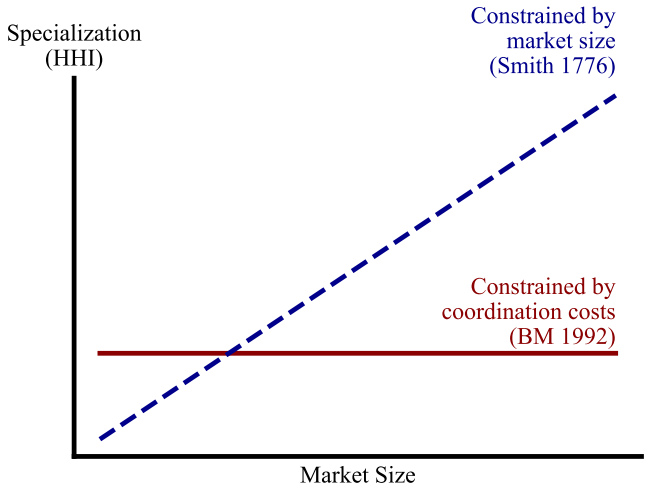
(b) But not for general oncologists



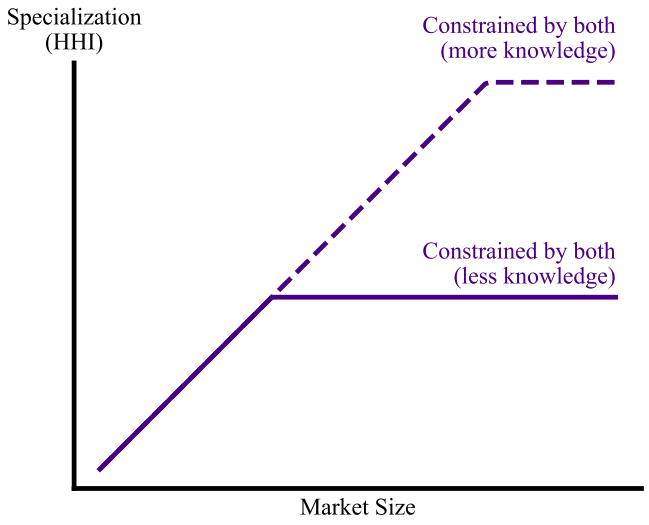
Notes: Panels (a) and (b) shows binned scatterplots and key regression estimates for the relationship between the Poisson recentered exposure to knowledge growth and change in specialization. Panel (a) shows this relationship for oncologists who start in the top 25% of the initial specialization distribution based on a Herfindahl—Hirschman index (HHI) over types of cancer managed in 2009. Panel (b) shows this relationship for oncologists who start in the bottom 75% of the initial specialization distribution based on a Herfindahl—Hirschman index (HHI) over types of cancer managed in 2009. Exposure to knowledge growth. In the Poisson 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 evaluated and managed at least 25 unique cancer patients in a 20% sample of Medicare Part B claims in 2008, 2009 and 2014.

Figure 5: Growth in knowledge (Δk) can increase market-level economies of scale

(a) Market size contraints and coordination cost constraints to specialization lead to different patterns of specialization by market size

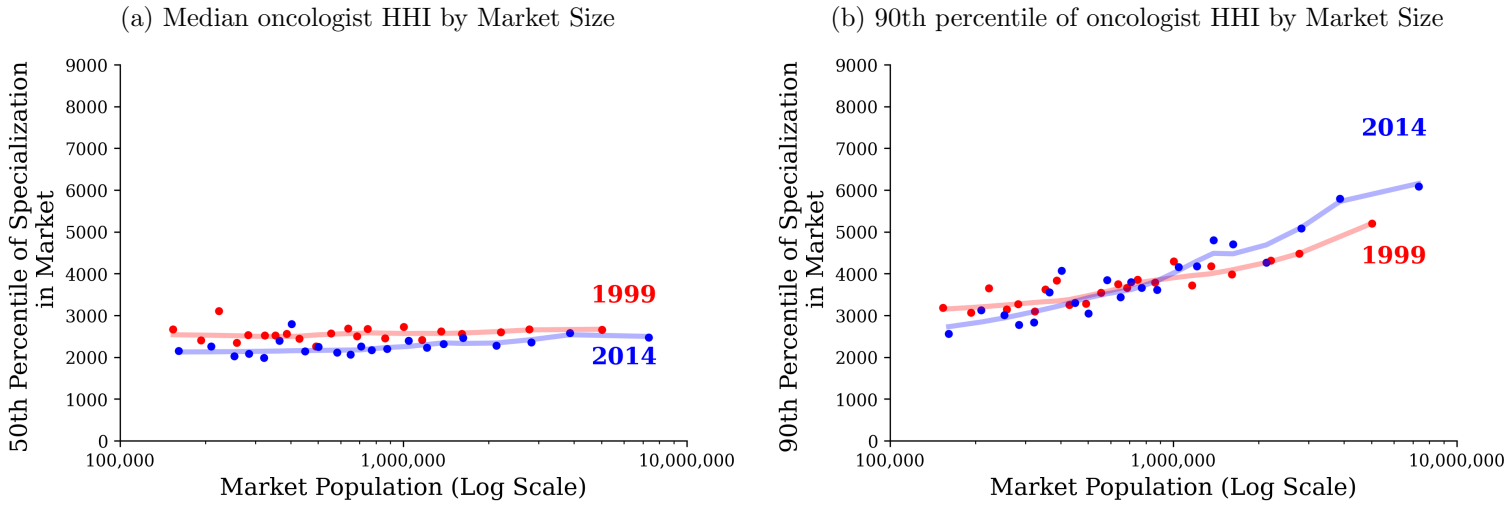


(b) When both constraints exist, more knowledge can lead to larger differences in specialization by market size



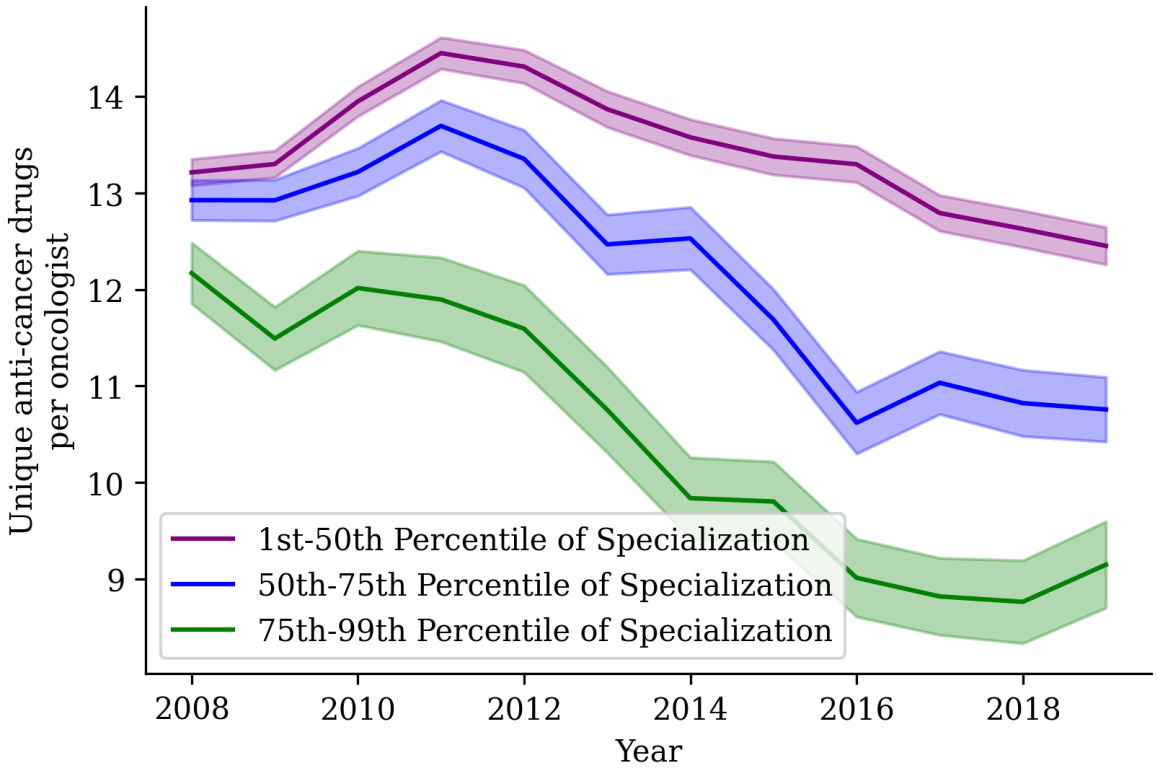
Notes: Panels (a) and (b) show theoretical predictions. Panel (a) shows predicted relationship between specialization and market size if specialization is constrained by coordination costs and constrained by market size. Panel (b) shows the change in the relationship between specialization and market size when knowledge increases in a setting where both coordination costs and market size constrain specialization.

Figure 6: Median specialization is stable over time and across market size, but the 90th percentile of specialization is increasing in the largest markets (1999-2014)



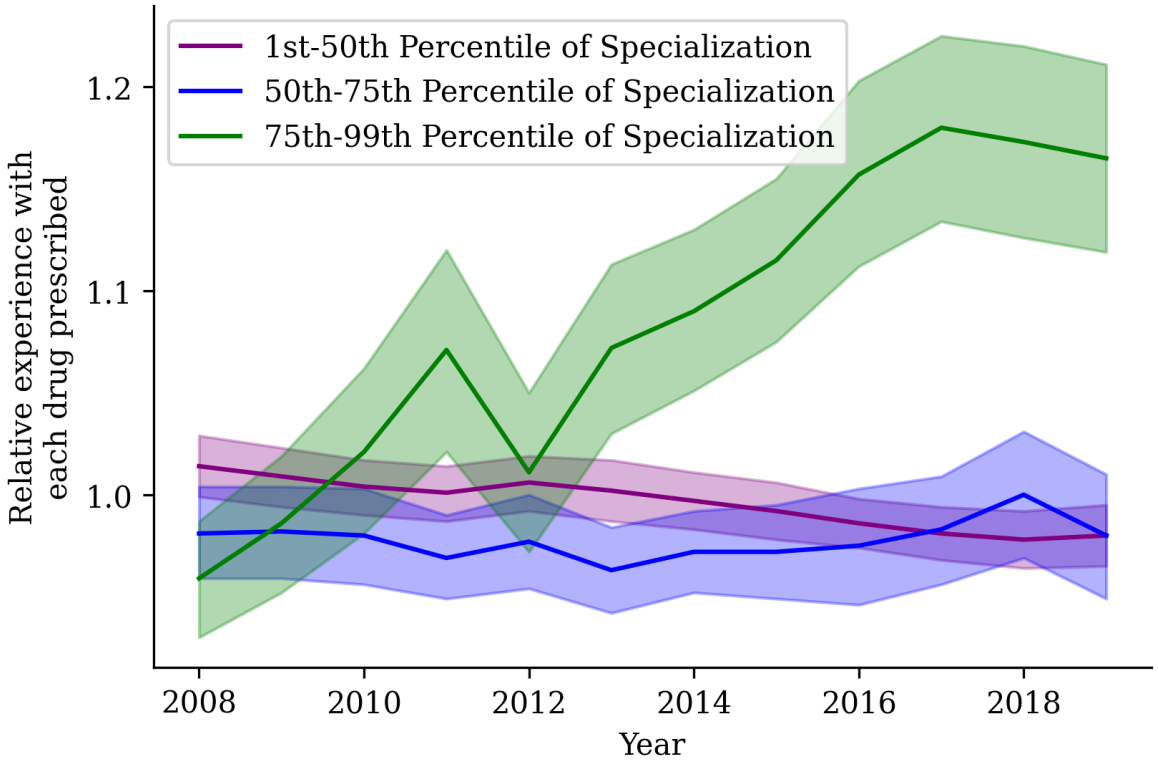
Notes: Panels (a) and (b) show a binned scatterplot of the 50th and 90th percentile of oncologist specialization by market size in 1999 and 2014. The x-axis shows the market size, measured as population, in the relevant year. Markets are defined using Hospital Referral Regions (HRR). The y-axis shows oncologist specialization, as measured using a Herfindahl—Hirschman index (HHI) over the types of cancer that the oncologist evaluated and managed in that year. In panel (a), each point represents the local average of the 50th percentile of HHI in an HRR in a year. In panel (b), each point represents the local average of the 90th percentile of HHI in an HRR in a year. The lines represent locally weighted scatterplot smoothing (LOWESS) over these points. All physicians in the sample are hematologist-oncologists, oncologists, or hematologists who evaluated and managed at least 25 unique cancer patients in the year in a 20% sample of Medicare Part B claims data.

Figure 7: Over time, specialists begin to prescribe fewer unique anti-cancer drugs than generalists in a year



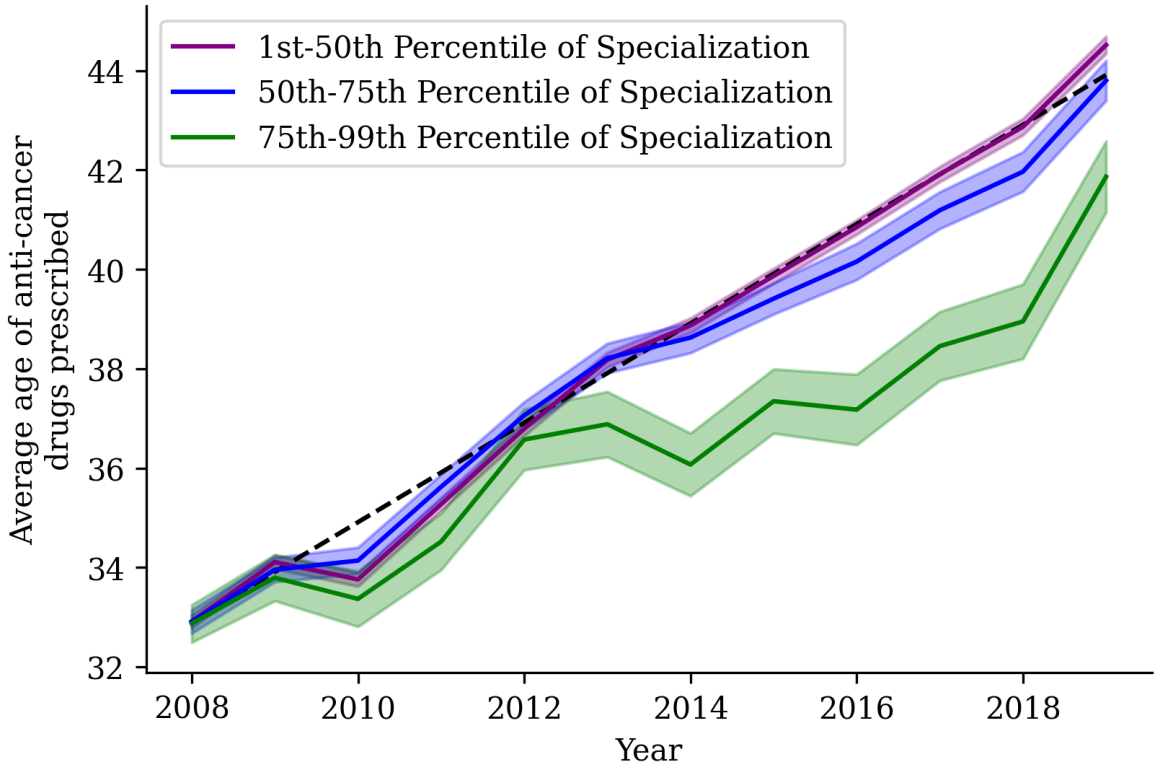
Notes: This figure shows the average number of unique anti-cancer drugs that oncologists prescribe over time by level of specialization. In the early years of the panel, specialized and generalist oncologists prescribe a similar number of unique drugs. However, by the end of the panel, the more specialized 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 claims.

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



Notes: This figure shows oncologists' average relative experience with the anti-cancer drugs that they prescribe over time by level of specialization. Experience is calculated as how many times an oncologist prescribed a drug in a year. Relative experience normalizes these values to have a mean of one in a year. In the early years of the panel, specialized and generalist oncologists are similarly experienced with drugs. However, by the end of the panel, more specialized oncologists are about 20% more experienced with 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). 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 claims.

Figure 9: Over time, specialists begin to prescribe newer drugs than generalists



Notes: This figure shows the average age of 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 claims.

Table 1: Summary Statistics for Balanced Panel of Oncologists (2008-2014)

	Mean	SD	Min	Max	N
N	4,861				
Unique Patients (2008)	69	33	25	297	4,861
Count of Cancer Types (2008)	15	4	1	27	4,861
HHI (2009)	2,739	1,453	954	10,000	4,861
Guideline Length (2009)	1,789				
Relevant Guideline Length (2009)	1,159	242	42	1,702	4,861
Change in Length of Guidelines (2009-2014)	1,745				
Change in Length of Relevant Guidelines (2009-2014)	1,087	226	24	1,612	4,861
Share in Largest 20% of Regions	1	0	0	1	4,861
Count Oncologists at Main Tax ID (2008)	19	40	1	219	4,861
Physician age (2008)	49	9	30	77	4,861

Notes: This table shows summary statistics for a balanced panel of oncologists from 2008 to 2014. All physicians in the sample are hematologist-oncologists, oncologists, or hematologists who evaluated and managed at least 25 unique cancer patients in a 20% sample of Medicare Part B claims in 2008, 2009 and 2014. More information about each variable is calculated is detailed in section 2.

Table 2: Specialized Oncologists Work in Larger Regions and Organizations than General Oncologists (2008-2014)

	Top 25% Most Specialized Docs		Other 75%	
	Mean	Std	Mean	Std
N	1,213		3,637	
Unique Patients	59	(29)	73	(34)
Count of Cancer Types	13	(5)	16	(3)
HHI (2009)	4,637	(1,729)	2,104	(460)
Academic	0.32	(0.47)	0.08	(0.27)
Share in Largest 20% of Regions	0.63	(0.48)	0.50	(0.50)
Count Oncologists at Main Tax ID	25	(41)	17	(40)
Physician age	50	(9)	49	(9)
Relevant Guideline Length (2008)	860	(279)	1,065	(165)
Change in Length of Relevant Guidelines (2009-2014)	932	(283)	1,139	(174)

Notes: This table compares key summary statistics for "specialist" and "general" oncologists. 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 2009. Specialist oncologists are defined as those in the bottom 75% of the specialization distribution as measured by a Herfindahl—-Hirschman index (HHI) over types of cancer managed in 2009. All physicians in the sample are hematologist-oncologists, oncologists, or hematologists who evaluated and managed at least 25 unique cancer patients in a 20% sample of Medicare Part B claims in 2008, 2009 and 2014. More information about how each variable is detailed in section 2.

Table 3: Guidelines grow faster for cancer types with more patients and surrogate endpoints

	(1) Page Count Increase	(2) Page Count Increase	(3) Page Count Increase
Patient Count in Medicare (Log)	1.95e-05	2.07e-05	1.58e-05
	(6.18e-06)	(5.39e-06)	(2.96e-06)
Surrogate Endpoints		0.986	-4.306
		(0.388)	(1.285)
Surrogate x Pat Count (Log)			0.569
			(0.120)
Observations	40	40	40
R-squared	0.143	0.433	0.747
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 National Comprehensive Cancer Network (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 A.1.

Table 4: Physician characteristics become substantially less predictive of exposure to knowledge growth after recentering

	(1) Unadjusted exposure to knowledge growth (Δk_i)	(2) Remove physician scope variation ($\Delta\tilde{k}_i$)	(3) Also remove cancer prevalence variation ($\Delta\tilde{k}_i$)	(4) Also remove cancer endpoint type variation ($\Delta\tilde{k}_i$)
Market population (millions)	-12.78	1.454	-0.574	0.00813
	(1.670)	(0.726)	(0.662)	(0.518)
Count oncologists at main org	-0.451	0.0446	0.0875	0.0169
	(0.0757)	(0.0370)	(0.0320)	(0.0240)
Patient Count	4.022	-1.635	-0.282	0.0247
	(0.0878)	(0.0445)	(0.0348)	(0.0283)
Observations	4,861	4,861	4,861	4,861
R-squared	0.385	0.263	0.017	0.005
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 Tax IDs 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. Raw exposure to knowledge growth is measured as the total increase in guideline page counts for cancer types from 2009 to 2014 for cancer types the physician managed in the baseline year (2008). In Column 2, the outcome variable is a recentered measure of exposure to knowledge growth, where the recentered measure controls for variation in physician scope. In this measure, predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. Predicted exposure is estimated as the number of cancer types treated times the mean knowledge growth per cancer type. This procedure removes variation in exposure to knowledge growth that arises become some physicians have a broader scope i.e. treat more types of cancer. In Column 3 and 4, the outcome variable is a recentered version of the measure of exposure to knowledge growth that controls for variation in physician scope and cancer field expected knowledge growth. As in Column 2, predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. However, in column 3 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 and an indicator for if the cancer type has well-established surrogate endpoints. Column 4 is similar, but predicted knowledge growth also accounts for the interaction between the number of Medicare patients and surrogate endpoints. This 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 evaluated and managed at least 25 unique cancer patients in a 20% sample of Medicare Part B claims in 2008, 2009 and 2014. All regression standard errors are robust.

Table 5: Knowledge growth increases specialization for oncologists in the top 25% of the initial specialization distribution

	(1) Knowledge Growth Recentered ($\Delta\tilde{k}_i$)	(2) Restrict Sample	(3) Add Age FE	(4) Add Market FE	(5) Add Org Size FE
Increase in Guideline Length (100 Pages)	223.8	202.0	200.8	194.1	165.1
	(74.55)	(73.30)	(73.59)	(81.17)	(82.47)
Observations	1,216	1,164	1,164	1,164	1,164
R-squared	0.008	0.007	0.008	0.139	0.152
Age FE			Yes	Yes	Yes
Market FE				Yes	Yes
Org Size 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 in 2009. 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 Tax IDs. 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 Tax IDs, 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 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 evaluated and managed at least 25 unique cancer patients in a 20% sample of Medicare Part B claims in 2008, 2009 and 2014.

Table 6: Knowledge growth does not increase specialization for oncologists in the bottom 75% of the initial specialization distribution

	(1) Knowledge Growth Recentered ($\Delta\tilde{k}_i$)	(2) Restrict Sample	(3) Add Age FE	(4) Add Market FE	(5) Add Org Size FE
Increase in Guideline Length (100 Pages)	-18.68	-18.88	-18.49	-10.75	-12.18
	(23.27)	(23.33)	(23.40)	(24.98)	(24.84)
Observations	3,647	3,629	3,629	3,629	3,629
R-squared	0.000	0.000	0.000	0.077	0.079
Age FE			Yes	Yes	Yes
Market FE				Yes	Yes
Org Size 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 in 2009. 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 Tax IDs. 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 Tax IDs, 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 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 evaluated and managed at least 25 unique cancer patients in a 20% sample of Medicare Part B claims in 2008, 2009 and 2014.

Table 7: Median specialization does not increase with market size, but top specialization does, and increases more steeply over time

	(1) Median HHI	(2) 90th Pct HHI
Current Population (Log)	6.426	492.2
	(37.48)	(79.38)
Current Population (Log) x 2014	90.32	399.3
	(48.62)	(114.1)
Observations	601	601
R-squared	0.090	0.252

Notes: Table 7 shows a regression of market-level oncologist specialization on market size, year, and an interaction of market size and year. 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, oncologists, or hematologists who evaluated and managed at least 25 unique cancer patients in the year in a 20% sample of Medicare Part B claims data.

10 Appendix

For Online Publication

Knowledge Growth and Specialization
Maya Lozinski
July 2024

A Data Appendix

A.1 Cancer Groupings

We create a set of custom diagnosis code groupings that align with how cancers are grouped in guidelines. This section discusses why and how we create these custom groupings.

We 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. We wish to group these into the category of "breast cancer" to align with the NCCN guidelines for breast cancer treatment.

To create the grouping, we 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 we rectify with our custom grouping. First, we drop the CCS diagnosis codes for cancer treatment and consider only those for specific types of cancer. Second, we make our groupings more specific. For example, the CCS groupings classify colon cancer of "unknown behavior" as an "unknown" cancer type. We re-classify this as colon cancer. Third, we make our custom grouping more granular to align with the guidelines. For example, we divide the CCS category of "Leukemia" into "Myeloid Leukemia" and "Lymphocytic Leukemia" (guidelines are even further split by acute and chronic within each subtype of leukemia). Lastly, we rearrange groupings that are misaligned with clinical groupings. For example, the CCS groupings group rectal cancer with anal cancer. However, guidelines group rectal cancer with colon cancer, as these are more biologically similar. Our custom grouping follows the guidelines' approach, with a single category for colon and rectal

cancer.

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, we map from ICD-10 to ICD-9 codes, then to our custom cancer grouping.

Notably, our 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. We cannot observe these other dimensions in the diagnosis codes. However, we do not need to observe them as NCCN guidelines are also organized based on the anatomical origin of tumors.

In addition, we are reassured that anatomic origin is the appropriate high-level grouping because NCCN guidelines are also organized anatomically. This observation suggests that our 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".

A.2 Labeling Academic Physicians

This section discusses how we labeled physicians as academic or non-academic. Our measurement 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 our data.

We 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. We 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"). We 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. We 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. We use affiliation in May 2024 rather than 2008 as we have more limited tools to determine past affiliation status. Finally, we 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. Our measure of academic status prioritizes sensitivity (including all plausibly academic physicians) over specificity (definitively excluding all non-academic physicians from the label of "academic").

B Measurement Error in Physician Specialization

This section discusses measurement error in our 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). In addition, this measurement error does not change the paper's key findings.

B.1 Characterizing Measurement Error

Measurement error occurs because we 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 we observe only one patient; mechanically, we will estimate that $\hat{HHI}_i = 10000$. Suppose, we observe only two patients; mechanically, the minimum HHI we 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 we observe only a 20% sample of Medicare patients treated by a physician rather than all patients. As such, we 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 C.4 demonstrates that $\hat{HHI}_i$ declines with volume and that some of the decline is attributable to decreases in measurement error. We 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 25 random patients per doctor. Using 25 random patients inflates the finite sample measurement error for physicians who see more than 25 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 HHI calculated with 25 patients is consistently higher than the HHI calculated with all patients. Second, the relationship between volume and HHI changes over time. In 1999, low-volume and high-volume doctors have similar HHIs once we 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.

B.2 Effect of equalizing volume-dependent measurement error in main results

Next, I investigate the impact of volume-dependent measurement error on the main findings.

First, I consider the impact on the observed trends over time in specialization. Figure C.1 plots the distribution of HHI over time, similar to figure 2, 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. However, the 90th percentile of specialization increases substantially over time, consistent with rising top specialization.

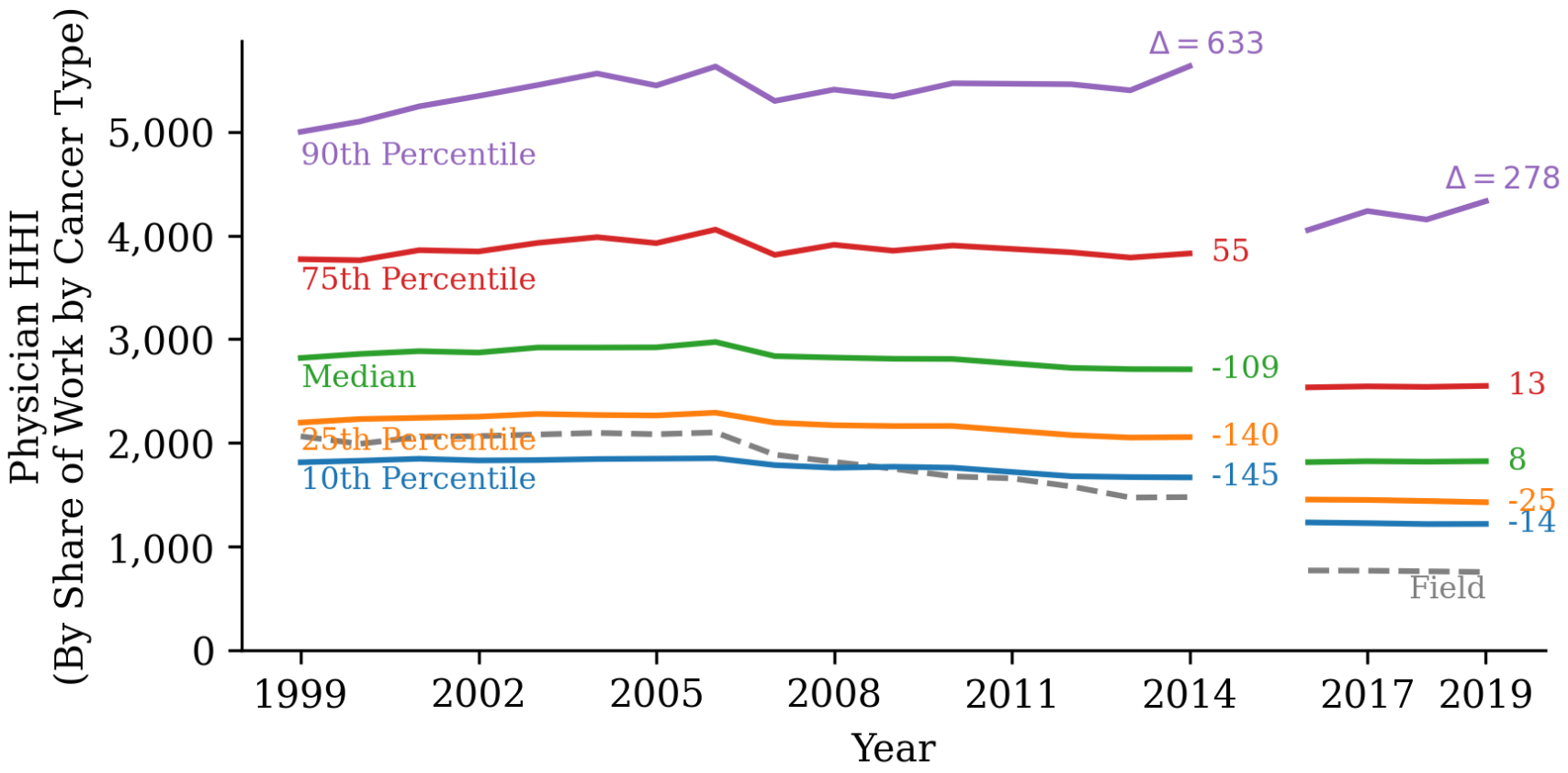
Next, I consider the impact of volume-dependent measurement error on my main identification strategy. Physician volume is a potential source of omitted variables bias. Table 4 shows that the raw measure of exposure to knowledge growth is correlated with patient

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.

Finally, we replicate the main regressions using only 25 patients per physician. Table C.8 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 C.9 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.

C Additional exhibits

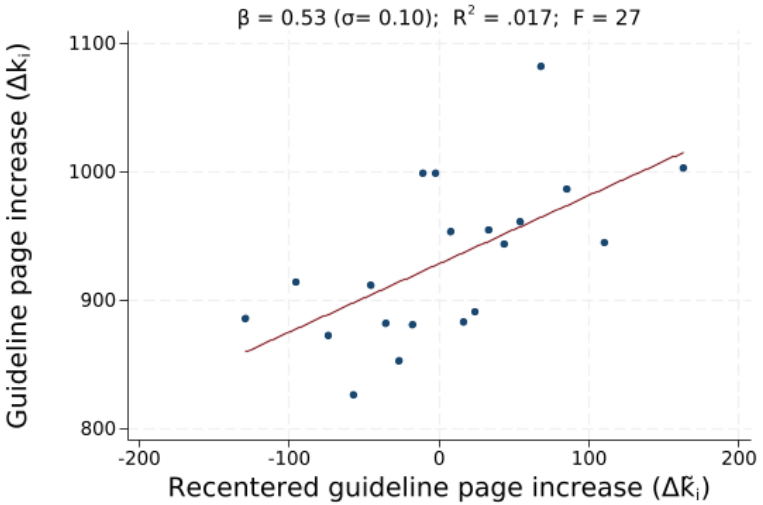
Figure C.1: Specialized Oncologists are Growing More Specialized Over Time (25 patients used to calculate HHI)



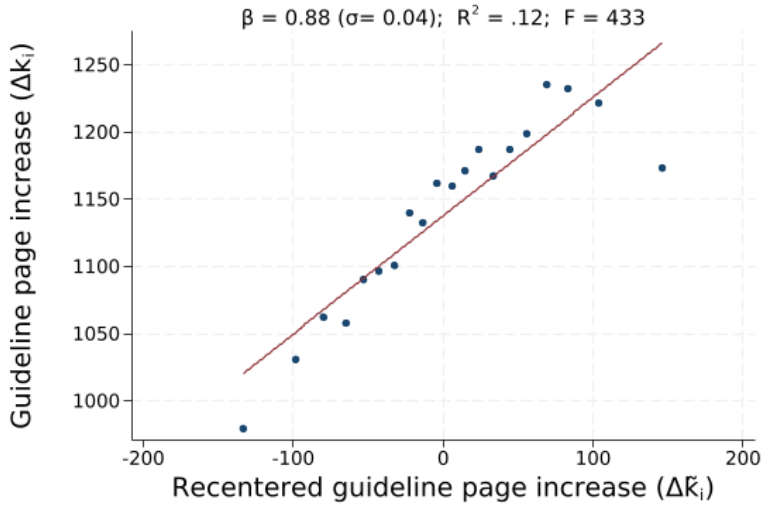
Notes: This figure shows the distribution of physician-level Herfindahl—Hirschman index (HHI), a measure of specialization, from 1999-2019, **using only 25 random patients per physician per year to calculate HHI**. All physicians in the sample are hematologist-oncologists, oncologists, or hematologists who saw at least 25 unique cancer patients in the year in a 20% sample of Medicare Part B claims. 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. The 10th, 25th, 50th, 75th, and 90th percentile are displayed and labeled. The dashed grey line represents the HHI of the field overall. 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. Overall, the 10th, 25th, and 50th percentile of oncologists see a roughly representative distribution of patients, suggesting they are "general" oncologists. As such, most oncologists remain generalists over this period. However, the most specialized oncologists, those near the 90th percentile of the distribution, become substantially more specialized over time.

Figure C.2: Quasi first stage suggests recentered instrument ($\Delta\tilde{k}_i$) is a strong "instrument" for unadjusted instrument (Δk_i)

(a) Specialized Oncologists

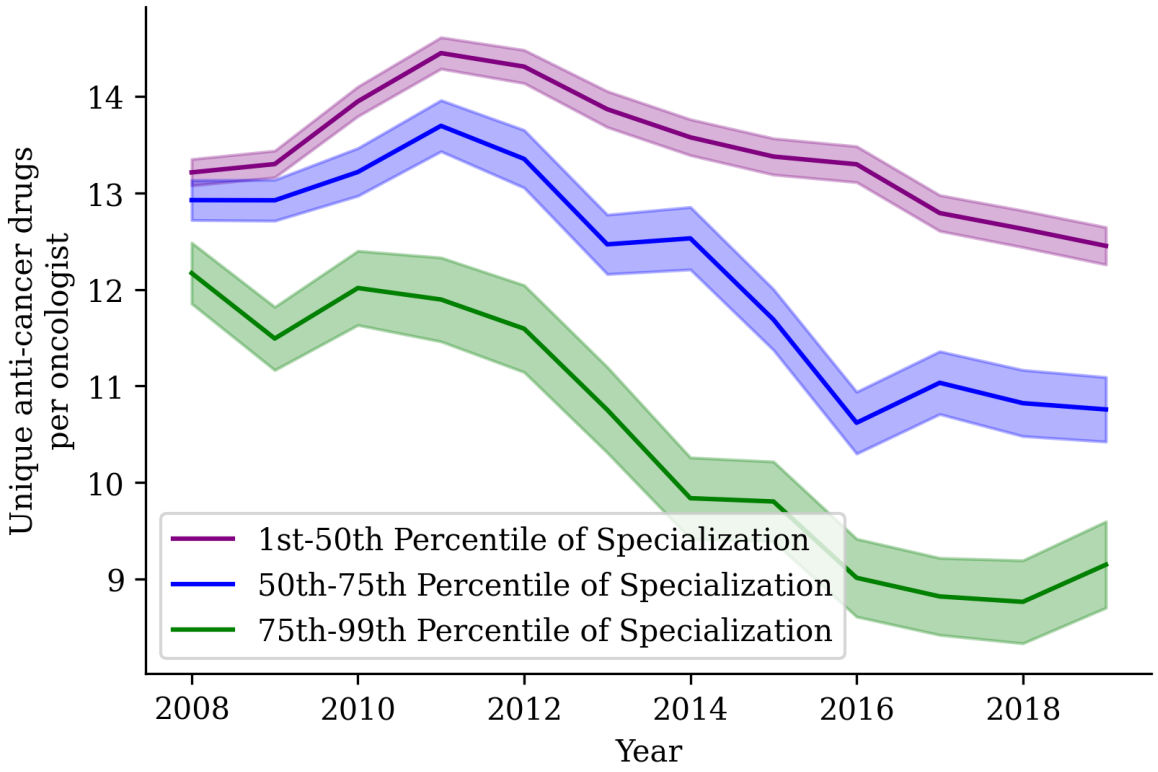


(b) General Oncologists



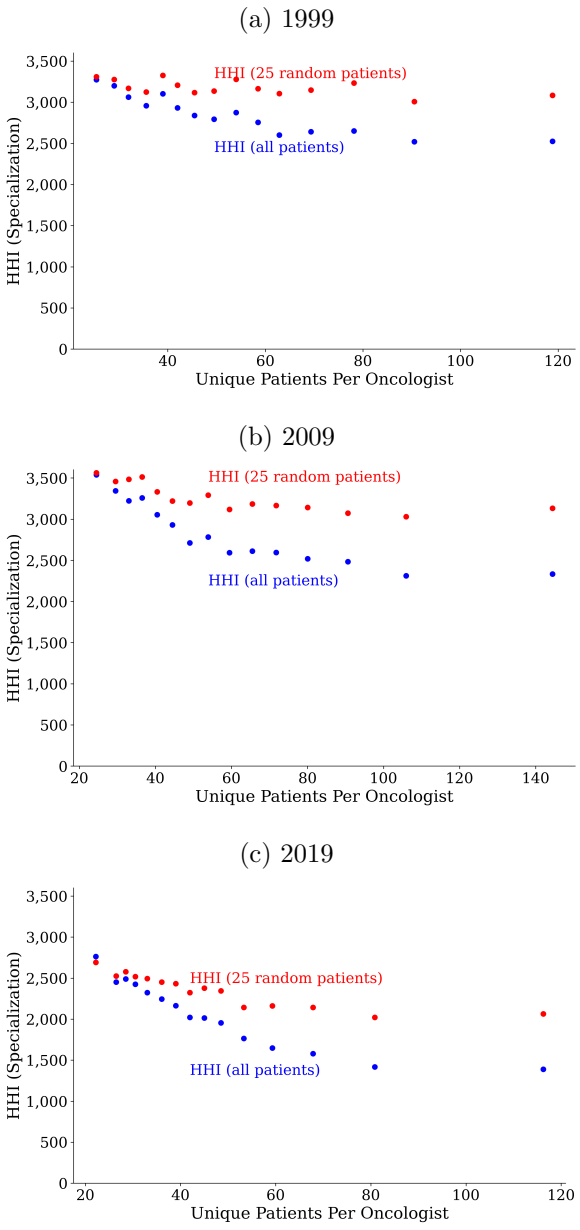
Notes: Panels (a) and (b) show the results for a quasi first stage regression of the recentered measure of exposure to knowledge growth on the unadjusted measure. Panel (a) displays the results for specialized oncologists, defined as those in the top 25% of the specialization distribution in 2009, as measured by a Herfindahl—Hirschman index over types of cancers managed. Panel (b) displays the results for general oncologists, defined as those in the bottom 75% of the specialization distribution. Both figures display a binned scatter plot and key regression outputs for the bivariate relationship between the raw measure and the recentered measure of physician-level exposure to knowledge growth. The raw measure of physician level exposure to knowledge growth is calculated using the cancer types a physician treated in 2008 and the subsequent growth in guideline length in those fields over the next five years (2009-2014). 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 treated in 2008 and the predicted growth in knowledge for each cancer type 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 evaluated and managed at least 25 unique cancer patients in a 20% sample of Medicare Part B claims in 2008, 2009 and 2014.

Figure C.3: Over time, specialists begin to prescribe fewer unique anti-cancer drugs than generalists in a year after removing variation in patient volume



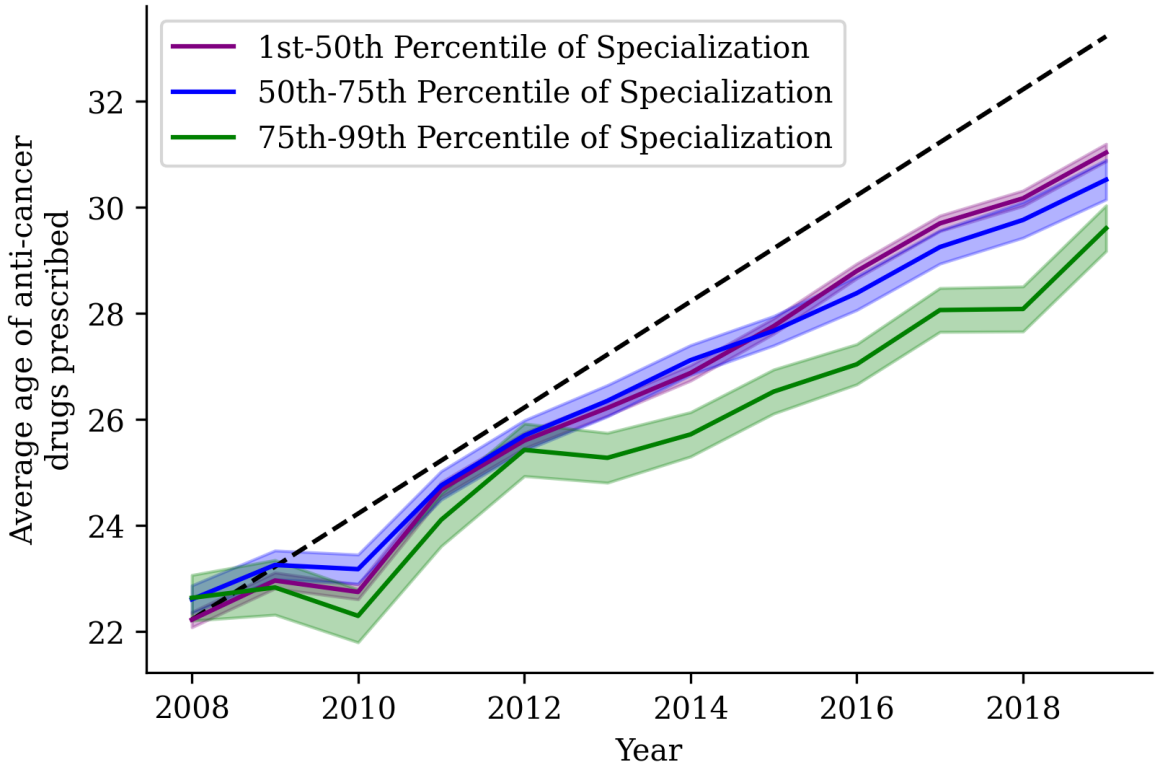
Notes: This figure shows the average number of unique anti-cancer drugs that oncologists prescribe over time by level of specialization. This figure differs from Figure 7 in that the average number of unique anti-cancer drugs is calculated over 11 random patients, rather than all patients, per oncologist per year. This change in sample removes systematic differences in patient volume that may exist by level of oncologist specialization. The patterns found in 7 continue to hold after accounting for possible underlying differences in patient volume by degree 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 claims.

Figure C.4: Physician level HHI declines with patient volume



Notes: Figure displays binned scatter plots of physician-level Herfindahl—Hirschman index (HHI), calculated using both all patients and 25 random patients per physician, by patient volume. Figures , , show the relationship for 1999, 2009 and 2019 respectively. 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 claims. 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.

Figure C.5: Over time, specialists begin to prescribe newer drugs than generalists (excluding steroids)



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 claims.

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

Year Site	2002	2020
ALL	0	117
AML	23	138
Anal	9	55
Bladder	40	113
Bone	25	96
Breast	52	223
Breast Risk	22	64
Breast Screening	33	77
CML	48	80
CNS	51	165
Cervical	24	86
Colon	20	188
Esophageal	19	154
GTN	0	40
Gastric	18	125
Hairy Cell	0	24
Head And Neck	86	233
Hepatobiliary	32	151
Hodgkins	30	85
Kaposi	0	39
Kidney	14	64
Lung	68	294
MDS	22	91
MPN	0	101
Mastocytosis	0	67
Mesothelioma	0	46
Myeloma	27	95
NHL	49	658
Neuroendocrine	46	141
Ovarian	26	371
Pancreatic	23	160
Penile	0	49
Prostate	18	167
Prostate Detection	22	63
Rectal	18	167
Sarcoma	31	146
Skin	58	358
Small Bowel	0	48
Testicular	34	78
Thymic	0	41
Thyroid	69	136
Uterine	42	108
Vulvar	0	58
All	**1,099**	**5,760**

Notes: Figure C.1 displays the page count of the first set of treatment guidelines in 2002 and 2020 from the National Comprehensive Cancer Network (NCCN) Clinical Practice Guidelines in Oncology. 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 C.2: For specialized oncologists, physician characteristics become substantially less predictive of exposure to knowledge growth after recentering

	(1) Unadjusted exposure to knowledge growth (Δk_i)	(2) Remove physician scope variation ($\Delta \tilde{k}_i$)	(3) Also remove cancer prevalence variation ($\Delta \tilde{k}_i$)	(4) Also remove cancer endpoint type variation ($\Delta \tilde{k}_i$)
Market population (millions)	-29.01	2.047	-1.371	-0.103
	(4.353)	(1.676)	(1.442)	(0.988)
Count oncologists at main org	-1.028	0.137	0.172	0.0107
	(0.217)	(0.0811)	(0.0713)	(0.0440)
Patient Count	4.709	-1.565	-0.420	-0.0796
	(0.233)	(0.110)	(0.0848)	(0.0634)
Observations	1,215	1,215	1,215	1,215
R-squared	0.346	0.170	0.023	0.011
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 Tax IDs 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. Raw exposure to knowledge growth is measured as the total increase in guideline page counts for cancer types from 2009 to 2014 for cancer types the physician managed in the baseline year (2008). In Column 2, the outcome variable is a recentered measure of exposure to knowledge growth, where the recentered measure controls for variation in physician scope. In this measure, predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. Predicted exposure is estimated as the number of cancer types treated times the mean knowledge growth per cancer type. This procedure removes variation in exposure to knowledge growth that arises become some physicians have a broader scope i.e. treat more types of cancer. In Column 3 and 4, the outcome variable is a recentered version of the measure of exposure to knowledge growth that controls for variation in physician scope and cancer field expected knowledge growth. As in Column 2, predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. However, in column 3 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 and an indicator for if the cancer type has well-established surrogate endpoints. Column 4 is similar, but predicted knowledge growth also accounts for the interaction between the number of Medicare patients and surrogate endpoints. This 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 evaluated and managed at least 25 unique cancer patients in a 20% sample of Medicare Part B claims in 2008, 2009 and 2014. All regression standard errors are robust.

Table C.3: For general oncologists, physician characteristics become substantially less predictive of exposure to knowledge growth after recentering

	(1) Unadjusted exposure to knowledge growth (Δk_i)	(2) Remove physician scope variation ($\Delta \tilde{k}_i$)	(3) Also remove cancer prevalence variation ($\Delta \tilde{k}_i$)	(4) Also remove cancer endpoint type variation ($\Delta \tilde{k}_i$)
Market population (millions)	-1.416	0.283	-0.680	-0.0825
	(1.155)	(0.740)	(0.728)	(0.614)
Count oncologists at main org	-0.0263	-0.0208	0.0472	0.0121
	(0.0534)	(0.0394)	(0.0346)	(0.0289)
Patient Count	3.435	-1.564	-0.218	0.0626
	(0.0791)	(0.0483)	(0.0376)	(0.0321)
Observations	3,646	3,646	3,646	3,646
R-squared	0.434	0.275	0.012	0.005
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 Tax IDs 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. Raw exposure to knowledge growth is measured as the total increase in guideline page counts for cancer types from 2009 to 2014 for cancer types the physician managed in the baseline year (2008). In Column 2, the outcome variable is a recentered measure of exposure to knowledge growth, where the recentered measure controls for variation in physician scope. In this measure, predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. Predicted exposure is estimated as the number of cancer types treated times the mean knowledge growth per cancer type. This procedure removes variation in exposure to knowledge growth that arises become some physicians have a broader scope i.e. treat more types of cancer. In Column 3 and 4, the outcome variable is a recentered version of the measure of exposure to knowledge growth that controls for variation in physician scope and cancer field expected knowledge growth. As in Column 2, predicted exposure to knowledge growth is subtracted from raw exposure to knowledge growth. However, in column 3 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 and an indicator for if the cancer type has well-established surrogate endpoints. Column 4 is similar, but predicted knowledge growth also accounts for the interaction between the number of Medicare patients and surrogate endpoints. This 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 evaluated and managed at least 25 unique cancer patients in a 20% sample of Medicare Part B claims in 2008, 2009 and 2014. All regression standard errors are robust.

Table C.4: For specialized oncologists, knowledge growth does not reduce the number of types of cancers managed

	(1) Knowledge Growth Recentered ($\Delta\tilde{k}_i$)	(2) Restrict Sample	(3) Add Age FE	(4) Add Market FE	(5) Add Org Size FE
Increase in Guideline Length (100 Pages)	-0.0656	-0.0801	-0.118	-0.0389	-0.0542
	(0.139)	(0.142)	(0.143)	(0.157)	(0.157)
Observations	1,216	1,164	1,164	1,164	1,164
R-squared	0.000	0.000	0.007	0.158	0.160
Age FE			Yes	Yes	Yes
Market 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 their Tax IDs. 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 Tax IDs, 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 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 evaluated and managed at least 25 unique cancer patients in a 20% sample of Medicare Part B claims in 2008, 2009 and 2014.

Table C.5: For general oncologists, knowledge growth does not reduce the number of types of cancers managed

	(1) Knowledge Growth Recentered ($\Delta\tilde{k}_i$)	(2) Restrict Sample	(3) Add Age FE	(4) Add Market FE	(5) Add Org Size FE
Increase in Guideline Length (100 Pages)	-0.107	-0.106	-0.139	-0.135	-0.129
	(0.0774)	(0.0775)	(0.0772)	(0.0812)	(0.0813)
Observations	3,647	3,629	3,629	3,629	3,629
R-squared	0.000	0.000	0.019	0.127	0.129
Age FE			Yes	Yes	Yes
Market 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 Tax IDs. 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 Tax IDs, 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 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 evaluated and managed at least 25 unique cancer patients in a 20% sample of Medicare Part B claims in 2008, 2009 and 2014.

Table C.6: Knowledge growth largely increases specialization for oncologists in the top 25% of the initial specialization distribution, after raising patient count threshold for exposure to knowledge growth and varying evaluation and management code sample

	(1)	(2)	(3)	(4)	(5)
Increase in Guideline Length (100 Pages)	223.8	284.6	292.6	25.52	293.3
	(74.55)	(81.82)	(97.28)	(155.8)	(80.96)
Observations	1,216	1,216	1,216	326	988
R-squared	0.008	0.010	0.007	0.000	0.015
Min Pat Count per Field	1	2	3	1	1
HHI Sample	All pat	All pat	All pat	All pat	All pat
Claims	All EM	All EM	All EM	New EM	Outpatient EM

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 in 2009. 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 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 evaluated and managed at least 25 unique cancer patients in a 20% sample of Medicare Part B claims in 2008, 2009 and 2014.

Table C.7: Knowledge growth does not increase specialization for oncologists in the lower 75% of the initial specialization distribution, after raising patient count threshold for exposure to knowledge growth and varying evaluation and management code sample

	(1)	(2)	(3)	(4)	(5)
Increase in Guideline Length (100 Pages)	-18.68	-10.45	63.27	44.24	22.26
	(23.27)	(24.99)	(27.90)	(53.17)	(26.91)
Observations	3,647	3,647	3,647	978	2,962
R-squared	0.000	0.000	0.001	0.001	0.000
Min Pat Count per Field	1	2	3	1	1
HHI Sample	All pat	All pat	All pat	All pat	All pat
Claims	All EM	All EM	All EM	New EM	Outpatient EM

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 in 2009. 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 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 evaluated and managed at least 25 unique cancer patients in a 20% sample of Medicare Part B claims in 2008, 2009 and 2014.

Table C.8: Knowledge growth typically increases specialization for oncologists in the top 25% of the initial specialization distribution, estimated with 25 patients per oncologist

	(1)	(2)	(3)
Increase in Guideline Length (100 Pages)	223.8	135.8	135.8
	(74.55)	(83.22)	(83.22)
Observations	1,216	1,216	1,216
R-squared	0.008	0.002	0.002
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 in 2009. 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 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 evaluated and managed at least 25 unique cancer patients in a 20% sample of Medicare Part B claims in 2008, 2009 and 2014.

Table C.9: Knowledge growth typically increases specialization for oncologists in the lower 75% of the initial specialization distribution, estimated with 25 patients per oncologist

	(1)	(2)	(3)
Increase in Guideline Length (100 Pages)	-18.68	24.66	24.66
	(23.27)	(29.72)	(29.72)
Observations	3,647	3,643	3,643
R-squared	0.000	0.000	0.000
Knowledge Exposure Sample	25 pat	All pat	25 pat
HHI Sample	All pat	25 pat	25 pat

Notes: This table shows regression estimates for the relationship between exposure to knowledge growth and change in specialization for 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 in 2009. 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 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 evaluated and managed at least 25 unique cancer patients in a 20% sample of Medicare Part B claims in 2008, 2009 and 2014.