Biased beliefs and stigma as barriers to treatment and innovation adoption

Laura Grigolon, Laura Lasio*

February 22, 2023

Abstract

Lung cancer is associated with smoking and is characterized by low treatment rates and research funds. We estimate a model of treatment choice where patients internalize societally biased beliefs on the effectiveness of treatment and stigma, basing their treatment decision on the treatment decisions of their reference group. Identification rests on the exogenous variation in the treatment propensity of physicians. Placing all patients in a neighborhood characterized by low social discrimination increases treatment rates by 4% and the use of innovative therapies by 3%. Social effects account for around 2% of the gap in research funding for lung cancer.

Laura Grigolon: University of Mannheim, Department of Economics, MaCCI and CEPR. (Email: laura.grigolon@uni-mannheim.de). *Laura Lasio:* McGill University, CIREQ, CIRANO, and CEPR. (Email: laura.lasio@mcgill.ca). This paper would not have been possible without the guidance of Dr. William Evans, who generously shared his oncological expertise with us. We thank Claudia Allende, Andrew Bibler, Bryan Bollinger, Chris Conlon, Michael Dickstein, Geert Dhaene, Pierre Dubois, Daniel Garcìa, Leonardo Giuffrida, Martin Hackmann, Ben Klopack, Margaret Kyle, Rafael Lalive, Fernando Luco, Jürgen Maurer, Thomas Neuber, Elena Prager, Nicolas Schutz, Michelle Sovinsky, Keith Teltser, Frank Verboven, Daniel Xu, Nicolas Ziebarth, Christine Zulehner, and participants at seminars and conferences for their helpful comments. Tania Guerra, Nazli Inay, Naemi Reiss, and Olivia Yu provided excellent research assistance. We thank Stefana Jovanovska, Keng Liu, and Refik Saskin at ICES, Jessica McCutcheon at the Canadian Hub for Applied and Social Research, and Brooke Filsinger at Ontario Health. We gratefully acknowledge financial support from SSHRC (Insight Development Grant), the German Research Foundation (DFG) through CRC TR 224 (Project A02), and the McGill Observatory on Health and Social Services Reforms.

But I think that's how you associate it. Because the first thing they ask—even me, the first thing I would ever ask somebody was, "Did you smoke?" (Female lung cancer patient, recent quitter)

...people who are diagnosed with lung cancer, they have feelings that it's their fault or feelings that people will think that they're using up their health resources and they don't somehow deserve them as much (Healthcare professional)

Quotes from Hamann et al. (2013) and Dunn et al. (2016), respectively

1 Introduction

Lung cancer is the most commonly diagnosed cancer and the leading cause of cancer-related deaths worldwide: it accounts for 13% of all new cancer cases and has the lowest survival rate among leading cancers. Fortunately, the advent of targeted and immunotherapy agents has revolutionized our understanding of the disease in the past decade. These therapies significantly improve patient survival, are often administered orally (instead of intravenously), and are associated with milder side effects. Unfortunately, and surprisingly, many patients have not taken full advantage of these innovations: lung cancer patients access treatment at much lower rates than patients affected by cancers with similar (untreated) survival rates. Furthermore, these striking differences in adoption are not fully explained by heterogeneity in the diseases or patients (Sacher et al., 2015).

One explanation for the lack of adoption lies in the nature of the disease and, more specifically, in the negative social effects associated with having lung cancer. As is to be expected, most lung cancer patients have a smoking history. The strong association of lung cancer with smoking can result in *biased beliefs* and *stigma* connected with the disease. In other words, patients incorrectly believe that therapy is ineffective (biased beliefs) or feel shame about having lung cancer as conferred by the social representation of lung cancer as self-inflicted (stigma). In addition, as biased beliefs and stigma constitute barriers to accessing treatment, they may also hinder the adoption and diffusion of innovative therapies for cancer patients. In turn, a lower number of treated patients impacts the number and

value of investments made in innovative therapies. While lung cancer is responsible for 32% of cancer deaths, it receives only 10% of cancer research funding: Kamath et al. (2019) report an average spending of USD 2,229 in research per lung cancer death, compared to USD 24,442 for breast cancer.

In our paper, we tackle the question: to what extent may social effects, such as biased beliefs and stigma, hinder access to treatment, the adoption of innovative therapies, and investment in innovation? While the current literature has explored a variety of motives to investigate heterogeneity in adoption patterns, from learning and uncertainty about side effects (Crawford and Shum, 2005, Gong, 2019), to healthcare culture (Cutler et al., 2019), we are the first to explore the connection between disease stigmatization and innovation.

We combine a unique collection of micro-level datasets, including treatment modalities and health and socio-demographic information, for the population of patients diagnosed with lung cancer in the Canadian province of Ontario between 2008 and 2018. We start with a linear-in-means specification to identify social effects in the probability of treatment. The share of untreated patients living in the same neighborhood is our measure of social effects, which exploits the granular geographic information available in the data and captures the role of a patient's reference group in the decision to seek treatment. Following the literature on social norms, as well as the health policy literature, we identify the community in which the patient lives as the relevant reference group. Causal social effects are hard to identify empirically because of simultaneity and correlated effects. We address simultaneity effects by focusing on the choice of newly diagnosed patients whose decision to pursue treatment may be influenced by patients from the same neighborhood diagnosed in previous years, but not vice versa. To disentangle social effects from correlation in unobserved attributes, we isolate the variation in treatment choices of fellow patients living in the same community independently of unobservables. In particular, we rely on quasi-random variation in treatment rates of the reference group. In a system with universal healthcare, such as that in Ontario, patients do not access secondary care directly and do not choose their oncologists. In addition, those clinicians work in regional cancer centers and do not have ties to a specific neighborhood. We construct the (risk-adjusted) average treatment propensity of physicians treating the

patients in the reference group in the previous years, and use it as an instrumental variable for treatment rates in the neighborhood. In other words, we exploit an exogenous shifter of treatment rates in a research design that manipulates the characteristics of the reference group in a manner unrelated to a patient's characteristics: past treatment propensity of physicians should not otherwise influence an individual after controlling for the patient's own physician (Angrist, 2014). Placebo tests using other cancer types (for which stigma is less of a concern) also confirm the effectiveness of our identification strategy.

We find that a one percentage point increase in the share of untreated patients in the neighborhood reduces the individual's probability of accessing treatment by around 0.2 percentage points. We also find a positive association between our measure of social effects and the severity of symptoms at diagnosis, suggesting that social stigma causes patients to delay seeking medical care. As timely medical attention is critical for this disease, awareness campaigns would be extremely valuable in order to educate the public about the possible symptoms and encourage earlier encounters with treatment providers.

In order to confirm that the share of untreated patients living in the neighborhood is a good proxy of societally biased beliefs and stigma, we conduct a survey of around 400 adults across Ontario to elicit a direct measure of attitudes towards lung cancer. The survey suggests that 20 to 23 percent of Ontarians feel less sympathy for lung cancer patients than for patients affected by other tumors. Notably, the variation in the degree of stigma across communities in Ontario positively correlates with the measure that we have constructed in our data.

Having established the presence of social effects in access to treatment, we model treatment choices as a nested sequence of decisions: at the upper level, the choice is between pursuing treatment or not; at the bottom level, the choice is between the different treatment options, including the innovative therapies. At the upper level, our econometric model allows for social effects in the choice of pursuing treatment. We find that placing all patients in a neighborhood characterized by low social stigma (corresponding to a risk-adjusted share of untreated patients equal to 45 percent) decreases the share of untreated patients by 4 percent. In particular, it increases the use of innovative therapies by 3 percent.

Following a cost-effectiveness approach that typically guides policy decisions when evaluating a given therapy, we compare the additional costs from treatment with its benefit, measured by the incremental quality-adjusted life year. We find that mitigating social stigma would imply additional overall spending of CAD 1.3 million in innovative drugs alone. However, the gain in survival is also high, which justifies the use of innovative therapies, whereby each additional patient would imply an extra annual spending of CAD 23,000 compared to the "no treatment" option, which is lower than CAD 65,000 (USD 50,000) per year of longer quality life (the de facto standard used by the Canadian medical agency to decide on the public coverage of drugs or medical procedures). The average spending for a patient treated with innovative therapies is equal to CAD 149,104, which is higher than all other treatment options. However, it is important to note that innovative therapies generate far greater health benefits in terms of survival. Our work corroborates, with precise patient-level cost information, the literature on the role of pharmaceutical treatments in improving outcomes in cancer care: see Lakdawalla et al. (2010), Lichtenberg (2010), Lichtenberg (2015), Dubois and Kyle (2016).

Finally, we quantify the impact of biased beliefs and stigma on R&D investment in cancer care. When looking at the relationship between innovation and market size, reverse causality is a potential issue: a higher number of treated patients may stimulate innovation, while, at the same time, innovation may increase the number of treated patients. To instrument for the effective market size, namely the number of treated patients, we use an accurate measure of potential market size; that is, the total number of patients affected by the disease. Our estimated elasticity suggests that a 10 percent increase in market size is associated with a 3.4 to 5.6 percent increase in R&D spending. Back-of-the-envelope calculations indicate that social stigma and biased beliefs about the effect of treatment are responsible for around 2 percent of the gap in research funding for lung cancer with respect to other common cancers; this amounts to $7 million every year in US public funding alone.

Related Literature A substantial medical literature documents the undertreatment and stigma associated with lung cancer. Clinical studies reporting a low level of adherence

to treatment guidelines (with no treatment or less intensive treatment than recommended) include Davidoff et al. (2010), Sacher et al. (2015), Cassidy et al. (2018), Walter et al. (2019), Blom et al. (2020), and Pham et al. (2021). According to these studies, the aggressiveness of lung cancer compared to other tumors, the fact that most patients are elderly and cannot tolerate toxic treatment, and the diagnosis when the cancer is already at an advanced stage only partially explain the lowest treatment rates for lung cancer among the leading cancers. In parallel, the medical and psychological literature examines the negative attitudes towards lung cancer: see Chapple et al. (2004), Chambers et al. (2012), Hamann et al. (2013), Carter-Harris (2015), Dunn et al. (2016), Riley et al. (2017). Most of these are qualitative studies based on interviews with patients, physicians, and oncology social workers; they all describe health-related stigma as part of the experience of having lung cancer. Feelings of stigma are closely connected to beliefs about lung cancer causation, poor prognosis and the perception of the futility of treatment (biased beliefs); many of these studies highlight the link between the internalization of such guilt and the reluctance to seek care.

Societally biased beliefs and stigma are an example of social conformity effects occurring when the utility of a given behavior is affected by others making the same choice. Economic studies have linked social stigma to the limited use of welfare programs: Moffitt (1983), Stuber et al. (2000), Bertrand et al. (2000). More generally, our work relates to two strands of the literature on social interactions. The first documents the effect of social interactions on program participation, including Duflo and Saez (2002), Aizer and Currie (2004), Chetty et al. (2013), and Grossman and Khalil (2020). The second emphasizes the role of social interactions in the diffusion of innovation. Since the seminal work by Granovetter (1978), several studies have shown the importance of social learning in technology adoption in different contexts, from medical innovation (Agha and Molitor (2018), Burke et al. (2007)) to agriculture in developing countries (Munshi (2004), Bandiera and Rasul (2006), Conley and Udry (2010), Beaman et al. (2020)). Most of these studies highlight how social networks facilitate the adoption and diffusion of technology via the acquisition or transmission of information. Social interactions in our context may also operate through the direct information channel but predominantly emerge as a more general form of social norms, namely stigma

and shared biased beliefs. With the exception of recent work applied to sanitation investment by Guiteras et al. (2019), we are not aware of any other work documenting this mechanism. In sum, neither the medical nor the economic literature has empirically investigated the link between stigma, access to treatment, and innovation.

We also contribute to the literature on the relationship between innovation and market size in the pharmaceutical industry. The most recent studies include Dubois et al. (2015) and Agarwal and Gaulé (2021). The literature has produced a wide range of elasticity estimates, partly because of the variety of measures employed for market size and innovation. These elasticities range from 4-6 across therapeutic classes in Acemoglu and Linn (2004), to estimated values for cancer of 0.53 in Lichtenberg (2007) and 0.38 in Dubois et al. (2015).[1] Thanks to our specific focus on the relationship between R&D spending and market size in cancer treatment, we are able to retrieve accurate measures for both market size and public R&D efforts.

Finally, our work relates to the literature on the role of physicians and patients in treatment decisions: see Coscelli (2000); Hellerstein (1998); Finkelstein et al. (2016); Cutler et al. (2019), and especially to the studies investigating heterogeneity in the adoption of innovative treatments: see Crawford and Shum (2005), Gong (2019), Currie and MacLeod (2020); Chan et al. (2022).

The remainder of the paper is organized as follows. Section 2 describes the institutional setting, the data, and some motivating facts documenting the dispersion in risk-adjusted treatment rates across neighborhoods. Section 3 discusses the identification strategy and the results of the linear specification. Section 4 builds and estimates a structural model of the treatment choice in lung cancer. Section 5 presents the counterfactual exercise. Section 6 links social barriers to market size and R&D investments, and Section 7 concludes.

[1] Measures of market size are: (i) income-weighted potential consumers in Acemoglu and Linn (2004); (ii) number of patients in Lichtenberg (2007); and (iii) global revenue of pharmaceutical products in Dubois et al. (2015). Measures of innovation are: (i) new molecular entities in Acemoglu and Linn (2004) and Dubois et al. (2015); and (ii) chemotherapy regimens in Lichtenberg (2007). Ward and Dranove (1995) and Giaccotto et al. (2005) use R&D spending as a measure of innovation effort. For a systematic review of the literature, see Agarwal and Gaulé (2021).

2 Cancer Care in Ontario

2.1 Institutional Background

Cancer care in Ontario Healthcare in Ontario is publicly funded through provincial and federal income taxation. The Ontario Health Insurance Plan (OHIP) guarantees coverage for all necessary diagnostic and physician services. Public funding programs cover the provision of cancer drugs. In particular, all approved intravenous drugs administered in outpatient settings are fully covered by the New Drug Funding Program, while oral drugs may qualify for either the Exceptional Access Program or the Ontario Drug Benefit Program (which may incur a small co-payment). Some less expensive, supportive drugs and non-essential services are not covered by OHIP but are either covered by hospital budgets or funded by private insurers and specific programs. Finally, all medical oncologists are part of alternative funding plans, and the choice of pursuing treatment (or the treatment type chosen) does not affect their compensation: agency issues are unlikely to arise in our setting.

Regional cancer programs in Ontario Cancer care is provided through 14 regional cancer programs, which are networks of hospitals. Our data identify the Local Health Integrated Networks (LHINs), which are the administrative authorities responsible for Ontario's regional provision of healthcare where patients are treated. Each LHIN hosts a regional cancer center, where all radiation treatments and a substantial proportion of systemic therapy are provided.[2] Some systemic therapy (chemotherapy, immunotherapy, and targeted therapy) is also provided at partner hospitals (affiliate and satellite facilities), but consultations with oncologists are mainly conducted at the regional cancer centers. Table A.5 in the Appendix provides the list of LHINs and related regional cancer centers.

Innovation in lung cancer treatment and R&D funding All metastatic cancers are incurable but treatable. Indeed, clinical studies have demonstrated the clear survival benefits of systemic therapy for lung cancer patients: see Davidoff et al. (2010), Arenberg (2012),

[2]The LHIN of Toronto Central is an exception with two cancer centers: Odette (Sunnybrook Health Sciences Centre) and Princess Margaret (University Health Network).

Sacher et al. (2015). Clinical evidence shows that patients with significant comorbidities can receive therapy that preserves their quality of life while substantially prolonging survival. The guidelines of Cancer Care Ontario, the agency responsible for cancer services in Ontario, follow the recommendations issued by the American Society of Clinical Oncology. These recommendations state that metastatic patients should be offered systemic treatment; in addition, therapeutic options exist for patients who may not be fully active.

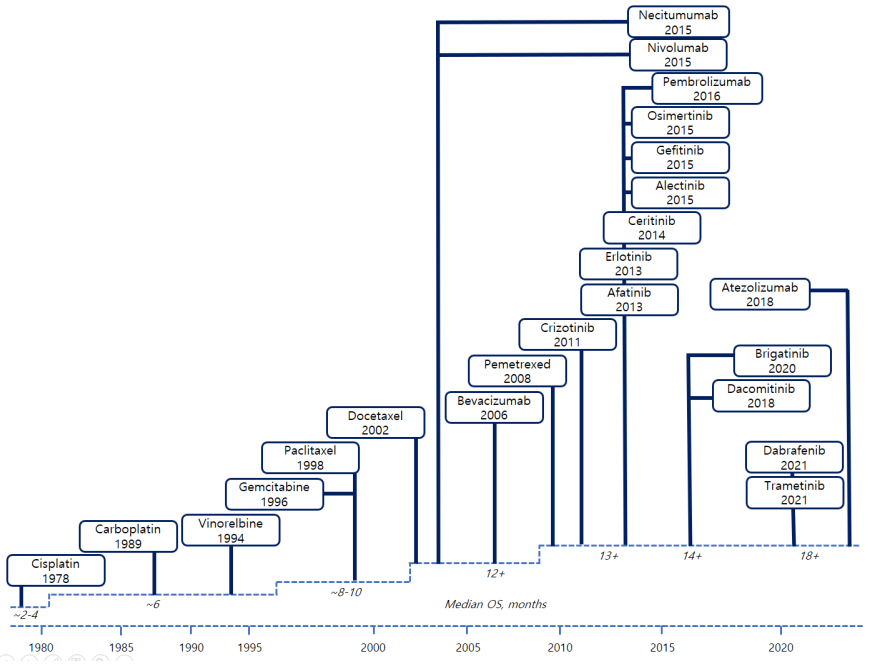
In recent years, the treatment of lung cancer has offered a substantial improvement in survival rates (Howlader et al., 2020); for example, in our data, one-year survival increases from 25% at the beginning of the sample to around 35% at the end of the sample. Such an increase is mainly attributable to new therapies, as screening programs for lung cancer remain uncommon and patients are diagnosed symptomatically.[3] In the past two decades, major therapeutic innovations have been introduced in lung cancer treatment with the advent of targeted therapy and immunotherapy. Figure 1 illustrates the therapeutic revolution in lung cancer treatment, with the number of *targeted* and *immunotherapy* drugs expanding greatly over the last decade; Table A.2 in the Appendix provides the list of all publicly funded therapeutic options available to the patients in our sample period (regimens). *Targeted* therapies exploit genetic changes that cause cancer (mutations) to find the right match between patients and treatment, while *immunotherapy* recruits the immune system to attack cancerous cells. These new therapies present health and economic advantages, especially compared to the standard of care based on aggressive and toxic chemotherapy. Specifically, they significantly improve patient survival, they are often administered orally, with cost savings relative to intravenous drugs, and they tend to involve fewer and milder side effects. Regimens often combine several chemotherapy drugs; Table A.2 reports the drugs contained in the regimen and the relevant dates of approval by the FDA and Health Canada (for drugs) and the Ontario health authority (for regimens). See Appendix C for more background on the therapeutic evolution in lung cancer.

The development of targeted therapies has been facilitated by cheap genome sequencing.

[3] Cancer Care Ontario. Cancer Fact: Lung cancer mortality differences between men and women influenced by smoking trends. April 2015. Available at cancercareontario.ca/cancerfacts.

Immunotherapy was initially developed for malignant melanomas; only later it has been used for lung cancer patients. Recent medical literature shows that up to 70% of lung cancer patients have an alteration targetable by existing drugs or drugs currently under development: see Suh et al. (2016). Research on novel immunotherapy agents is also advancing to extend their applications: see Zhang and Chen (2018). However, lung cancer is poorly funded compared to how common it is and how many deaths it causes. Kamath et al. (2019) report that while lung cancer is responsible for 32% of cancer deaths, it receives only 10% of cancer research funding; the average spending in research per lung cancer death is USD 2,229, compared to USD 24,442 for breast cancer. Lower research spending also appears to translate into fewer clinical trials. For example, panel (b) of Figure D.1 in Budish et al. (2015) shows that the ratio of the number of clinical trials to incidence is much lower for metastatic lung cancer with respect to the other leading cancers.

Figure 1: FDA approvals in advanced lung cancer - First line



The figure shows a timeline of FDA drug approvals for stage IV lung cancer - first line - since 1980. OS = overall survival (in months). Source: fda.gov.

2.2 Data

Cohort definition We use administrative data held at the Institute for Clinical Evaluative Sciences (ICES), a data repository consisting of record-level, linkable health datasets encompassing much of the publicly funded administrative health services records for the Ontario population. Table A.1 in the Appendix provides an overview of the datasets and the relevant variables that we extract. The main dataset is the Ontario Cancer Registry, which reports the diagnosis date and tumor characteristics, including the stage, for each patient diagnosed with cancer in Ontario. We select all patients diagnosed with stage IV (metastatic) non-small cell lung cancer with known disease stage from 2008 to 2018, with follow-up to the end of 2019. We match each patient to the primary caregiver and restrict our sample to physicians with a minimum number of five patients over the sample. Our final cohort comprises 15,761 patients. The cohort selection is motivated by three main reasons. First, this population presents a desirable setting for our study because the treatment decisions for this cancer stage are made by one primary physician, while, in non-metastatic stages, there may be other variables at play, including complementarities between radiology, surgical interventions and systemic therapy. Second, many innovative cancer drugs introduced in recent years were initially approved for the metastatic stage of the disease and only later approved for the treatment of earlier stages. Third, by restricting our sample to physicians with a minimum number of five patients over the sample, we address the concern of estimation error while focusing only on specialists who work in the regional cancer centers and are unrelated to specific neighborhoods. As detailed below, this is crucial for our identification strategy.[4]

In parallel, we select three other cohorts of cancer patients for the same years and following the same criteria: (i) stage IV colorectal cancer; (ii) stage IV prostate cancer; and (iii) stage IV female breast cancer. Colorectal, prostate and breast cancers are the most common

[4]Selection on the outcome (treated or untreated) is unlikely to be concerning in our setting. The vast majority of physicians matched to a handful of patients are local general practitioners providing palliative or end-of-life care to patients who are unlikely candidates for treatment (92% of those patients are untreated). In the full sample, 58% of patients are untreated; in the selected sample, 55% of patients are untreated. The filter decreases the sample size by 1,767 patients.

cancer types in Canada after lung cancer. We use these three cohorts for placebo tests: these patients are unlikely to face the same degree of social discrimination that characterizes lung cancer. We, therefore, perform our empirical analysis on these cohorts, in parallel with the main analysis, as a falsification check, with the expectation that social effects are irrelevant in the context of these cancers. We mainly focus on the cohort of colorectal cancer patients as the most appropriate comparison group. In a similar way to lung cancer, therapeutic decisions at this cancer stage are taken mainly by the oncologist. At stage IV, radiology is only used for supportive care (symptom management), survival probabilities are similar if the disease is left untreated (as highlighted in the survival analysis presented below), and therapies present comparable side effects: Table B.17 in the Appendix presents a qualitative comparison between the two cancers in terms of treatment toxicity. Further details on the three cohorts (colorectal, prostate, breast cancers) are presented in Appendix B.

Treatment plans Combining hospital claims for systemic treatment from the New Drug Funding Program database and the Activity Level Reporting System, we are able to reconstruct all treatment plans (regimens), if any, administered to each patient. Regimens often combine several chemotherapy drugs. Details on how we have reconstructed which regimens are administered to each patient are reported in Appendix A. The Activity Level Reporting System also includes information on the administration of radiation therapy, which helps achieve palliation and symptom controls in patients with metastatic disease. We classify treatment plans into three macro-categories: (i) no treatment; (ii) standard of care; and (iii) innovative therapies. No treatment means that the patient does not receive any systemic therapy (chemotherapy or innovative therapy). We identify as the standard of care both platinum doublet chemotherapy regimens based on combinations of cytotoxic agents (cisplatin or carboplatin) and third-generation agents (such as gemcitabine and pemetrexed), as well as single agents (for a complete list see Table A.2 in the Appendix). Innovative therapy includes all approved oral agents for first-line treatment (such as afatinib, crizotinib, erlotinib, and gefitinib) and immunotherapy drugs (pembrolizumab).

Patient characteristics We merge the cohort using anonymized patients' identifiers with the ICES datasets listed in Table A.1. We extract detailed health information on the patients, including measures of utilization at diagnosis (treatment, hospitalization, prescription drugs, care at home), outcomes (mortality), patient and disease characteristics (tumor morphology and histology, stage, patient sex, age, and income). Section A.3 in the Appendix details how comorbidities, cancer-related surgery, and other patient characteristics are constructed. Table A.3 in the Appendix provides a complete overview of the characteristics of the patients, their definition, and source.

Table 1 reports summary statistics for selected patient characteristics; Table A.4 in the Appendix reports summary statistics for the full set of variables. After excluding patients with incomplete records and those diagnosed via autopsy, we observe 15,761 patients and 334 physicians. Only 7,150 patients (45% of our sample) receive treatment; 78% of the treated patients receive the standard of care, and 22% receive innovative treatments. Innovative therapies steadily gained market share during the period thanks to the approval of new agents: around 4% of treated patients received innovative treatment in 2010 (almost entirely gefitinib), with the share increasing to 37% at the end of the sample. After the approval of new agents, we observe that their adoption rate is high and relatively stable, with no evidence of physicians' learning. Our setting differs from those explored by the literature on learning in pharmaceuticals, where physicians need to learn the matching between the drug and the patient in the absence of clear guidelines (Crawford and Shum, 2005), or can exploit spillovers across patients in a context of a large potential market (Coscelli and Shum, 2004). Two features of our setting explain this. First, oncologists are aware of new drugs well before their approval since cancer drugs must complete lengthy clinical trials showing evidence of safety and effectiveness, and prescriptions are offered as soon as the drug is cleared for provincial reimbursement; second, innovative drugs usually target specific mutations, as clearly indicated in the guidelines, with little substitutability among them or to the standard of care, which limits physicians' discretion.

Columns 2 to 4 of Table 1 compare the characteristics of patients who do not receive treatment (0) to patients receiving the standard of care (1) and innovative therapy (2),

while the last three columns report the results of a test on the equality of means for each subsample. Untreated patients tend to be male, older, more likely to present a tumor with squamous histology, less likely to undergo surgery, and present more comorbidities (as measured by the Charlson index) than patients who receive any systemic therapy. Among those who are treated, patients receiving innovative therapy are healthier beyond cancer (usually adenocarcinoma) and more likely to be women. Moreover, they are significantly less likely to be smokers at the time of diagnosis.[5]

We report the same set of summary statistics for colorectal, breast, and prostate cancer patients in Tables B.8, B.9, and B.10 of Appendix B.

[5] We observe the self-reported smoking status of the patient only for patients diagnosed after 2014, when the Ontario smoking cessation program was introduced; see Appendix B.

Table 1: Sample Summary Statistics: Patient characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Cohort	Treatment type				$p-value$	
		untreated	SOC	innovative			
		(0)	(1)	(2)	(0)=(1)	(0)=(2)	(1)=(2)
	Health-related attributes at diagnosis						
Charlson index	1.04	1.18	0.89	0.76	0.00	0.00	0.00
Active smoker (0/1)	0.32	0.36	0.37	0.16	0.36	0.00	0.00
Surgery (0/1)	0.03	0.02	0.04	0.03	0.00	0.01	0.02
Preventive care (%)	0.48	0.43	0.50	0.60	0.00	0.00	0.00
Home care (%)	0.26	0.34	0.17	0.19	0.00	0.00	0.11
	Cancer-related attributes						
Adenocarcinoma (0/1)	0.75	0.70	0.77	0.91	0.00	0.00	0.00
Squamous cell (0/1)	0.20	0.25	0.18	0.04	0.00	0.00	0.00
Multiple tumors (0/1)	0.01	0.01	0.02	0.03	0.00	0.01	0.00
	Socio demographics attributes						
Male (%)	0.52	0.54	0.53	0.41	0.20	0.00	0.00
Age	[65-69]	[70-74]	[65-69]	[65-69]	0.00	0.00	0.00
Distance to hospital (km)	31.26	30.98	33.61	24.65	0.00	0.00	0.00
Income quintile	2.81	2.71	2.92	2.97	0.00	0.00	0.27
Education terciles	1.91	1.87	1.92	2.04	0.00	0.00	0.05
Employment (0/1)	0.48	0.46	0.49	0.52	0.00	0.06	0.00
Minority (0/1)	0.50	0.49	0.48	0.61	0.06	0.00	0.00
	Health outcomes						
1-year survival prob.	0.29	0.12	0.45	0.68	0.00	0.00	0.00
Survival days	337	188	484	627	0.00	0.00	0.00
Tot. patients	15,761	8,611	5,545	1,605			

The table reports the summary statistics of selected variables in our sample related to patients. (Table A.4 in Appendix presents the summary statistics for the full set of patient characteristics.) The first column includes health-related attributes, tumor attributes, health care utilization measures, and a set of characteristics related to the three-digit zip code of the patient's residence for the whole sample. Columns 2-4 compare those characteristics between (i) untreated patients; (ii) patients treated with the standard of care (SOC or chemotherapy); and (iii) patients treated with innovative therapies. Columns 5-7 report the results of a Welch $t-$test across the subsamples.

Geographic characteristics The data reports the patient's place of residence at a very granular level; that is, the three-digit zip code (FSA, Forward Sortation Area). Canadian

postal codes identify a fine geographic unit: an FSA is roughly equivalent to a five-digit US zip code.[6] In our sample, we have 486 FSAs. In the urban context, the median FSA has an area of 19 square kilometers, with one-third of them below ten square kilometers, and 11,600 households.

We geocode the FSA to the census tract and block to add socio-demographic information combining the census and survey data from the Canadian Statistical Institute. We supplement our data with FSA-level information on income, employment, education, immigration, smoking and drinking habits, and pollution (particulate matter concentration, $PM_{2.5}$). We also include the Ontario marginalization index: the index measures multiple axes of deprivation in Ontario, including economic, ethnic-racial, age-based, and social marginalization.[7] Finally, we exploit the geographic dimension of our data to compute the distance between the centroid of the FSA of residence of the patient and both the nearest regional cancer center (should the patient decide not to be treated) and the center the patient chooses to attend.

Table A.5 in the Appendix presents an overview of the characteristics at the FSA level and their definition. As our neighborhood-level dataset contains a vast set of potential predictors of treatment, some of which are highly collinear, we select them via LASSO, and use the selected variables to estimate the model: see Belloni and Chernozhukov (2013). In practice, we use all the covariates to predict treatment rates by neighborhood, splitting the data into a training set for model development and a hold-out set for validation; the LASSO tuning parameter is selected using cross-validation.[8] Table 2 presents summary statistics at

[6] In Canada, six-digit postal codes may consist of a block face (one side of a city street between consecutive intersections), a community mailbox, an apartment building, or a mail delivery route: see Grubesic (2008).

[7] The index was developed by researchers at the Centre for Urban Health Solutions at St. Michael's Hospital in Toronto to explicitly capture inequalities in various measures of health and social well-being, either between population groups or between geographical areas: see Matheson et al. (2012). It combines a wide range of demographic indicators from the census into four distinct dimensions of marginalization: residential instability (percent of renters and those living alone); material deprivation (percent of low-income and solo parent families); dependency (percent of seniors and employment); and ethnic concentration (percent of recent immigrants and visible minority).

[8] We also fit a different machine learning model (that is, the random forest algorithm) to predict treatment rates and to ascertain which covariates affect treatment using variable-importance scores. The covariates selected as the most important predictors by the random forest algorithm are broadly consistent with those selected via LASSO.

the FSA-level for the selected variables. Lung cancer patients who do not receive systemic treatment tend to come from disadvantaged areas and live further away from a regional cancer center; in contrast, those receiving innovative therapy are more likely to live in urban areas and closer to a regional cancer center.

Table 2: Sample Summary Statistics: Neighborhood characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Cohort	Treatment type			$p-value$		
		untreated	SOC	innovative			
		(0)	(1)	(2)	(0)=(1)	(0)=(2)	(1)=(2)
Urban (%)	0.83	0.84	0.82	0.88	0.01	0.00	0.00
Population density	2,333	2,429	2,058	2,769	0.00	0.00	0.00
Median income	30,583	30,481	30,785	30,435	0.00	0.78	0.04
% income from welfare payments	22.29	22.62	22.29	20.49	0.01	0.00	0.00
Unemployment rate	8.26	8.31	8.18	8.26	0.00	0.32	0.15
Pollution (pm2.5)	28.98	27.47	33.45	21.65	0.01	0.03	0.00
Quintiles of marginalization index:							
instability	3.05	3.15	2.96	2.81	0.00	0.00	0.00
deprivation	3.28	3.34	3.20	3.24	0.00	0.00	0.33
ethnic concentration	3.01	2.98	2.94	3.42	0.08	0.00	0.00
Share of population:							
with high school degree	0.27	0.27	0.27	0.26	0.00	0.00	0.00
South-Eastern Asian immigrants	0.05	0.05	0.05	0.08	0.13	0.00	0.00
heavy smokers	0.14	0.14	0.14	0.12	0.05	0.00	0.00
heavy drinkers	0.36	0.36	0.36	0.34	0.00	0.00	0.00
Tot. patients	15,761	8,611	5,545	1,605			

The table reports the summary statistics of variables in our sample related to neighborhood characteristics. Columns 2-4 report summary statistics for the variables related to: (i) untreated patients; (ii) patients treated with the standard of care (SOC or chemotherapy); and (iii) patients treated with innovative therapies. Columns 5-7 report the results of a Welch $t-$test across the subsamples.

2.3 Survival Analysis

The raw statistics presented in Table 1 and Tables B.8, B.9, and B.10 in the Appendix suggest a shorter survival of lung cancer patients compared to patients with other cancers. However, these figures cannot be compared across patients or cancer types, as they are af-

fected by patients' characteristics. For example, within a cancer type, untreated patients tend to be older and in poorer health. Across cancer types, untreated patients share similar attributes, but differ along some important dimensions: for instance, (untreated) lung cancer patients tend to have more comorbidities than colorectal cancer patients, although they are, on average, younger. For an accurate comparison of survival across cancer types, we estimate a flexible parametric Royston-Parmar survival model for lung and colorectal cancer patients: Danesh et al. (2019). Our rich specification includes all the demographic and health-related patient characteristics, treatment modality (no treatment, chemotherapy, innovative therapy), histology of the tumor, year of diagnosis, and cancer care center of treatment or catchment area (if untreated), together with interactions between (i) age group and histology, (ii) treatment modality, and (iii) year of diagnosis. In addition, age group, treatment modality, and year of diagnosis are included as time-dependent variables.

We plot the survival curves for each treatment modality based on the coefficient estimates. The curves all refer to a hypothetical female patient with adenocarcinoma, aged 65-69 and with a low Charlson index (healthy), receiving palliative radiation but no surgery, diagnosed in 2018 and treated at Toronto Central. Figure B.3 shows that, when left untreated, this patient has a significantly worse expected survival rate. We estimate the same model using the sample of colorectal cancer patients. After controlling for patient characteristics, the survival probability between cancers is similar: the survival curves reported in Figures B.3 and B.4 in the Appendix show that the female lung cancer patient has a 12.9% [10.2-16.3] one-year survival probability if left untreated, compared to 14.4% [9.3-22.5] for a colorectal cancer patient with the same baseline observables. We also observe similar gains in survival coming from treatment: the one-year survival probability for a lung cancer patient treated with the standard of care equals 44.8% [40.2-50.0] and 65.4% [61.6-69.5] if treated with innovative therapy, compared to 64.9% [59.2-71.2] for a colorectal cancer patient with the same baseline observables treated with the standard of care.

We draw three conclusions from our results. First, treatment is effective: systemic therapy significantly increases survival rates for both lung and colorectal patients. Second, our clinical data is rich enough to obtain unbiased estimates of the effect of treatment: our

estimates are in line with the gains in survival from clinical trials, reporting that patients treated with innovative therapies (targeted and immunotherapy) can achieve an overall survival longer than two years, compared to an average nine months for those treated with standard chemotherapy: see de Castro-Carpeño et al. (2011). Third, the similarity in survival probabilities for lung and colorectal cancer across treatment types confirms the comparability of these two cancers for our placebo analysis.

2.4 Motivating empirical facts

Geographic variation in treatment rates We document some empirical facts about treatment variation across neighborhoods. Although average treatment rates increase over time, from 43 percent at the beginning of the sample to 50 percent at the end, these intertemporal differences are dwarfed by the spatial differences. Figure A.1 illustrates the spatial heterogeneity across the 14 administrative health regions in Ontario (panel a), denominated LHIN, and 486 neighborhoods (FSAs) (panel b). Figure A.2 visualizes the variation in incidence by LHINs (panel a) and FSAs (panel b). Following Duflo and Saez (2002), we compare the empirical variance in treatment rates observed in the data with the variance under the hypothesis that treatment rates are independent. The empirical variance in treatment rates across FSAs in 2018 equals 1.35; this number cannot be generated by independent behavior, which would give rise to a variance of only 0.05.

To further represent the variation in treatment rates, we follow Chandra and Staiger (2020) and estimate a random effect logit model of whether a patient receives treatment on the rich set of covariates describing the patient health (measures of utilization at diagnosis, patient and disease characteristics) and neighborhood-level random intercepts. We retrieve the Bayesian posterior (shrinkage) estimates of the random effects and add these to the fixed portion of the model to obtain the variation in treatment propensity at the patient level for observationally similar patients. The empirical Bayesian estimates account for the estimation error caused by the small sample of patients in each neighborhood, which would attenuate the estimated amount of variation. We also estimate the benefit of treatment; in particular, we estimate a random coefficient logit model of whether a patient survives after 90 days on

the treatment dummy and the patient covariates; that is, we allow for a neighborhood-level random intercept and a correlated random coefficient on treatment. The shrinkage estimates of the random coefficient on treatment capture the variation in the benefit of treatment at the neighborhood level.

Panel A of Figure 2 reports the histogram of risk-adjusted treatment propensity across the 486 neighborhoods for lung cancer patients, with the average neighborhood normed to zero. The histogram visually illustrates the sizable variation across neighborhoods in treatment rates for observationally similar patients. In Panel B of Figure 2 we overlay the treatment propensity for colorectal cancer: lung cancer exhibits a greater variation across neighborhoods with respect to colorectal cancer. Figure B.5 in the Appendix shows that this also holds for the other cancer types (breast and prostate). Panel C of Figure 2 is a binned scatter plot of treatment propensity across neighborhoods against the effect of treatment. The figure shows that treatment is beneficial (always positive), and that neighborhood-level treatment propensity and treatment benefit are slightly negatively correlated (-0.20). In other words, patients coming from a neighborhood with a low-propensity of treatment would benefit more from treatment. Why is lung cancer unique among top cancers in the heterogeneity of preference for accessing treatment? One answer may be the social discrimination connected to the disease: due to the social entrenchment of negative beliefs and stigma surrounding lung cancer, patients that would benefit from treatment may be left behind.

Physician variation in treatment rates We match patients' records with physicians' claims to identify the primary physician treating the patient. Details on the matching algorithm are presented in Appendix A. As we restrict our sample to physicians with a minimum number of five patients over the sample, we focus only on specialists who work in the regional cancer centers and are unrelated to specific neighborhoods. Our sample includes 192 medical oncologists, who are matched to 81% of the patients; the remaining specialists are radiation oncologists.

We construct a measure of risk-adjusted treatment propensity at the physician level. Again, we estimate a random effect logit model of whether the patient receives treatment

on patient covariates with physician-level random effects. The Bayesian posterior (shrinkage) estimates of the random logit intercepts capture the variation in treatment propensity across physicians. Shrinkage techniques adjust for estimation error in our physician-specific estimates.

A critical feature of the medical system in Ontario is that individuals can choose the hospital where they are treated but not a specific oncologist within the hospital. Notably, the allocation to a physician is random from the patient's perspective. Ontario's guidelines do not allow for a referral to a specific oncologist within the chosen cancer center, and conversations with medical oncologists also confirm that direct referral is not possible.[9] Our data confirm that physicians are not related to a specific neighborhood: patients from the same neighborhood share the same physician, on average, only 7.5% of the times. Appendix Table D.18 verifies the quasi-random assignment of physicians to neighborhoods by reporting the coefficient estimates from an ordinary least squares regression of physician treatment propensity on neighborhood characteristics. Almost all coefficients are not statistically significantly different from zero.[10] A joint F-test fails to reject the null of quasi-random assignment at the neighborhood level.

Team decisions or group practices are uncommon during the period covered by our sample, so spillovers across physicians are unlikely. From our billing data, we observe that only 5.4 percent of patients receive consultations from multiple oncologists in a group practice setting between the diagnosis date and the start of treatment, or 60 days after the diagnosis if the patient does not undergo treatment.

Finally, while patients can choose the hospital (LHIN), sorting at the hospital level has limited scope. Because of the severity of symptoms caused by the disease, most patients (71 percent) receive treatment at the closest cancer center, and 83 percent do not travel to a hospital more than 100 km away.[11]

[9] In other contexts, patients with specific characteristics may pursue physicians with a higher propensity to treat: see Dubois and Tunçel (2021).

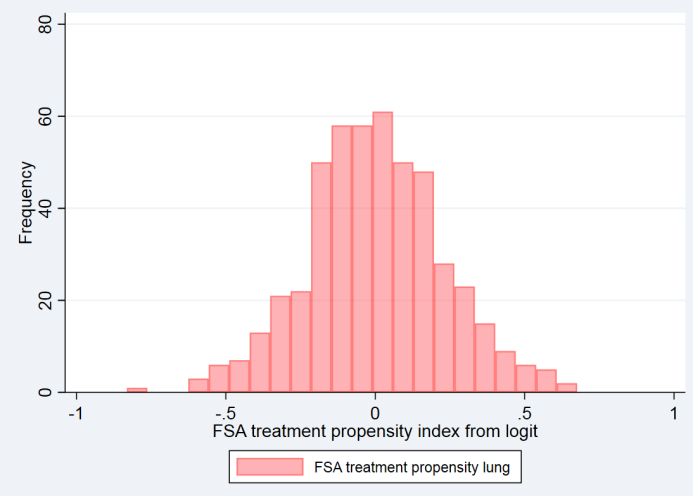
[10] An exception is the variable "share of the population of South-Eastern Asian origin". Medical research (Shi et al., 2014) shows that patients of South-East Asian ethnicity are 50% more likely to present the EGFR oncogenic mutation in lung cancer. Unfortunately, we do not have information on the patient's ethnicity, and this variable likely captures this patient's attribute.

[11] We implement the Kolmogorov-Smirnov equality test in the distribution of physician treatment propen-

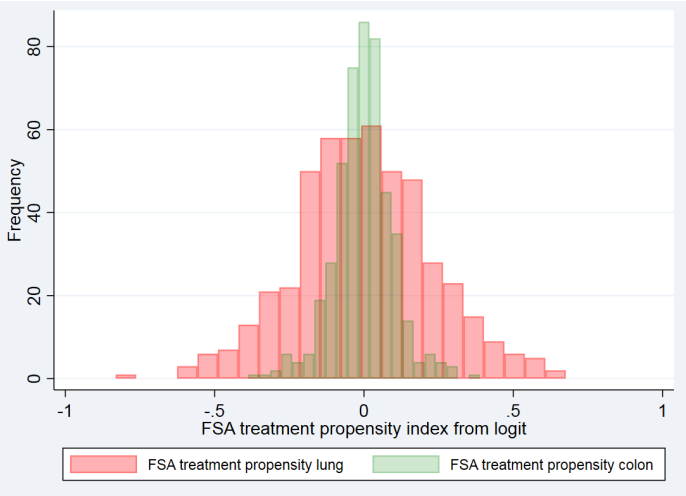
Figure 3, Panel A, documents the wide variation in the treatment propensity across physicians: the distribution is multimodal, with two main peaks, corresponding to high and low-propensity physicians. Overlaying the histograms of risk-adjusted physician propensity to treatment with colorectal cancer (see Panel B) illustrates that physicians exhibit substantially more variation in treatment propensity for lung cancer with respect to colorectal cancer.

sity for each pair of cancer centers. In 86 percent of the cases, we fail to reject the hypothesis that the two cancer centers have the same distribution of physician treatment propensity. In the few instances in which we reject the hypothesis of equal distributions, those cancer centers are located in catchment areas that are not contiguous, and only a handful of patients seek treatment between the pairs of hospitals.

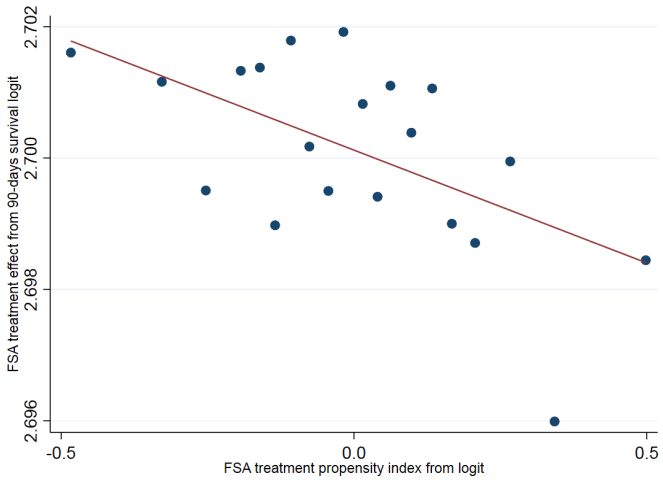
Figure 2: Geographic variation in treatment rates



Panel A



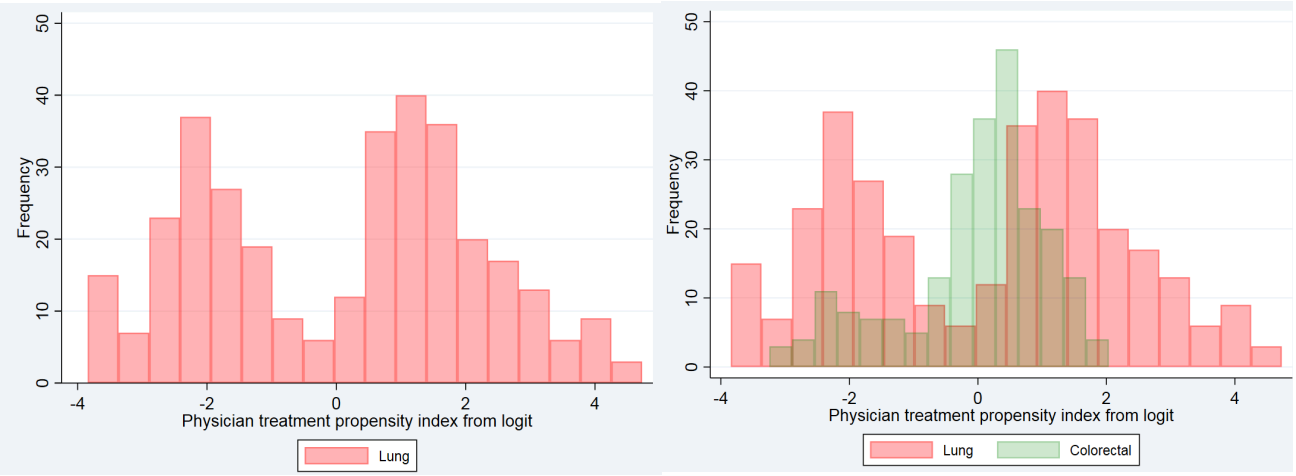
Panel B



Panel C

Panels A and B show the risk-adjusted treatment rate at the FSA (three-digit zip code) level; the rate is an empirical Bayesian estimate of a FSA-level intercept from a random effect logit model of whether a patient receives treatment regressed on patient and tumor characteristics and a FSA-level random intercept. In panel C, the survival benefit of treatment at the FSA level is an empirical Bayesian estimate of the FSA-level coefficient on treatment from a random-coefficient logit model of whether a patient survived 90 days after diagnosis regressed on whether the patient received treatment, controlling for patient and tumor characteristics. We allow for a FSA-level random intercept and (possibly correlated) random coefficient on treatment.

Figure 3: Physician variation in treatment rates



Panel A Panel B

Panels A and B show the risk-adjusted treatment rate at the physician level. This rate is an empirical Bayesian estimate of the physician-level intercept from a random effect logit model of whether a patient receives treatment regressed on patient and tumor characteristics and a physician-level random intercept.

3 Social effects in access to treatment

3.1 A simple empirical specification

We consider biased beliefs and stigma as a form of social effects. Empirically identifying social effects is notoriously challenging because the decisions of the reference group are endogenous. We start by using a linear specification to illustrate the three main empirical issues affecting our setting: first, the definition of the appropriate reference group; second, the reflection problem; and third, correlated effects. Let i index the patient and t the diagnosis year; $r(i)$ denote the relevant reference group of patient i and $p(i)$ the physician treating patient i. The variable y_{it} is a binary indicator representing patient i's decision to pursue treatment; the decision is determined by the treatment decision of other patients belonging to the patient's reference group, $\overline{d}_{it}$; the individual observable attributes related to health (x_{it}) and socio-demographics (z_{it}); the contextual effects of the reference group (neighborhood) ($\eta_{r(i)t}$);

supply-side determinants of treatment choice captured by $\eta_{p(i)}$, which denotes the fixed effect of the primary physician treating patient i; and unobservable individual attributes (ε_{it}):

$$y_{it} = \beta_1 \overline{d}_{it} + x_{it}\beta_2 + z_{it}\beta_3 + \eta_{r(i)t} + \eta_{p(i)} + \varepsilon_{it}, \tag{1}$$

where $\overline{d}_{it}$ is the share of *untreated* patients living in the same neighborhood and diagnosed in the previous periods:

$$\overline{d}_{it} = \frac{1}{\Sigma_{l=1}^{T} |\Re_{i,t-l}|} \sum_{l=1}^{T} \sum_{k \in \Re_{i,t-l}} d_{k,t-l},$$

where $d_{k,\tau}$ is the decision of patient k in period τ to take treatment and $\Re_{i,\tau}$ the set of patients living in individual i's neighborhood in period τ; specifically, $d_{k,\tau}$ is a decision indicator equal to one if patient k decides *not* to take treatment in period τ, and zero otherwise.

The key identification concern arises from disentangling *endogenous* effects (which refer to an individual's propensity to behave in a way that varies with the prevalence of the behavior in the group) from *correlated* effects (which refer to the similarity of behavior coming from similar environments or individual characteristics). In our empirical strategy, we use the treatment propensity of physicians associated with the reference group to exogenously shift the average treatment rate of patients living in the same neighborhood; the allocation of a physician to a patient is quasi-random, from the patient's perspective, as direct referrals are not allowed in Ontario. In other words, we exploit an exogenous shifter of treatment rates, consistent with the suggestion of Angrist (2014) to manipulate peer characteristics in a manner unrelated to individual characteristics.

We now discuss the empirical issues and how we solve the potential identification concerns.

Reference group The first difficulty with models of social interactions is the correct identification of the reference group: see Manski (1993). Previous works have emphasized the role of geographic proximity in the prevalence of social norms, including social stigma. Most of the literature on social norms, as well as the medical and health policy literature, uses an individual's community - often identified as the neighborhood of residence - as the relevant reference group, where social and work-level interactions tend to occur: see Bertrand et al.

(2000), Aizer and Currie (2004), Bayer et al. (2008), Topa and Zenou (2015), Baranov et al. (2015), Stewart et al. (2015) and Elliot et al. (2018). Bailey et al. (2018) use data from social networking services to develop a Social Connectedness index. They find that the intensity of friendship links is strongly declining in geographic distance: on average, 63% of friendship links are to individuals living within 100 miles and that the geographic concentration of the social ties tends to be higher in areas with worse socioeconomic outcomes (lower income, education, and social mobility).

Following the literature, we treat members of the neighborhood (FSA) where the patient resides as the main reference group. Patients from the same community are likely to be subject to similar degrees of social discrimination. Hence, the choice of fellow patients may play a direct role in an individual's choice to seek treatment, as well as serve as a proxy for the degree of empathy that the community feels for lung cancer patients. We leverage the rich information in our data on the geographic proximity between patients diagnosed with the same disease and exploit the variation in treatment rates that we observe at this granular level. In robustness checks, we note the appropriate axes to situate our patients in the social space.

The reflection problem First recognized in a seminal paper by Manski (1993), the reflection problem is the failure of identification that may arise from the interdependence in individuals' choices. A patient may choose whether or not to access treatment on the basis of the choices of patients in the reference group; choices of the reference group may in turn be affected by the individual's choice.[12]

We address the simultaneity or reflection problem by exploiting the panel dimension in our data. The measure we use to proxy for social stigma as a barrier to access treatment is the share of patients living in the same neighborhood who were diagnosed in previous periods and did not access treatment.[13] In our setting, the choice of using the decision of

[12] Interdependence in patients' decisions does the following: (i) generates simultaneity bias, as the mean outcome in the reference group is influenced by the patient's choice; and (ii) impedes the use of standard maximum likelihood methods to estimate the parameters of interests, as independence in individual choice probabilities may be violated.

[13] The approach of using the lagged outcome in the reference group was initially proposed by Brock and

past patients is intuitive: the effect of social stigma is naturally unidirectional as new patients may be affected by the decisions of previously diagnosed patients, but not vice versa.

Correlated effects Correlated effects are essentially a problem of omitted variables; they arise because the researcher is unable to observe all possible determinants of the behavior, including those that may be correlated within neighborhoods. Our main challenge is distinguishing social effects (endogenous effects) from correlated effects, which would lead to the same observational outcomes, but would not qualify as a social phenomenon. Patients in the same reference group may behave similarly because they share similar characteristics, some of which may be unobserved by the researcher. Correlation in the treatment decisions among patients in the same neighborhood may, therefore, not necessarily arise from social stigma but, for example, from similar socio-demographic factors, sharing the same doctor, or a similar attitude towards medical advice.

To identify social effects in treatment choices, we seek to isolate variation in treatment choices of fellow patients living in the same neighborhood, independently of ε_{it}. We construct the instrument as the average treatment propensity of physicians treating the patients in the reference group as follows:

$$\overline{S}_{it} = \frac{1}{\sum_{\tau=1}^{t-1} |\Re_{i,\tau}|} \sum_{\tau=1}^{t-1} \sum_{k \in \Re_{i,\tau}} S_{k,\tau},$$

where $S_{k,\tau}$ is the treatment propensity of the physician treating patient k in period τ and $\Re_{i,\tau}$ the set of patients living in individual i's neighborhood in period τ. The risk-adjusted measure of physician treatment propensity, $S_{k,\tau}$, is calculated by estimating a random effect logit model of whether the patient receives treatment on patient covariates with physician-level random effects on the sample of patients diagnosed in period t. The Bayesian (shrinkage) estimates of the random logit intercepts capture the variation in treatment propensity across physicians. The resulting risk-adjusted treatment propensity is a continuous variable where the average treatment propensity is normed to zero. By computing the risk-adjusted measure

Durlauf (2001) and applied in Aizer and Currie (2004) and Sorensen (2006).

of physician treatment propensity on the sample of patients diagnosed in all periods before patient i's diagnosis, we eliminate the bias originating from patient i's own case entering into the instrument. The identification assumption is that the past treatment propensity of physicians should not otherwise influence an individual's treatment decision after controlling for the patient's own physician $p(i)$.

Finally, we use two additional instruments that exploit the idiosyncratic variation in cancer diagnosis: the percentage of patients in the reference group affected by adenocarcinomas and synchronous multiple lung cancers.

The first stage equation is:

$$\overline{d}_{it} = \gamma_1 Z_{it} + x_{it}\gamma_2 + z_{it}\gamma_3 + \theta_{r(i)t} + \theta_{p(i)} + u_{it}, \tag{2}$$

where Z_{it} denotes the set of instruments including the past average treatment propensity of physicians and the tumor-specific attributes of patients in the reference group, $\theta_{r(i)t}$ neighborhood characteristics, $\theta_{p(i)}$ physician fixed effects, and u_{it} the error term. We use X_{it} to denote all the observable patient's attributes. For the identification of β_1, we need the following conditions to be satisfied:

Assumption 1 *Independence* $E\left(\varepsilon_{it}|Z_{it}, X_{it}\right) = E\left(\varepsilon_{it}|X_{it}\right)$

Assumption 2 *Relevance* $\gamma_1 \neq 0$

First, we discuss evidence that independence is satisfied in our setting. The main concern is the possibility that the instruments proxy for some shared unobservables at the neighborhood level that affect the probability of a patient accessing treatment. As the diagnosis itself and the attributes of the disease are essentially idiosyncratic (not all smokers develop the disease), we are confident that the requirement of independence is satisfied for cancer-related attributes of the reference group. Regarding the average treatment propensity of physicians, four features of our setting, documented in Section 2.4, allow us to establish independence: (i) medical and radiation oncologists work in regional cancer centers and do not have ties to specific neighborhoods; (ii) patients can choose the hospital where they are

treated but not a specific oncologist within the hospital; (iii) all hospitals exhibit substantial heterogeneity in the propensity to treatment across physicians and patients are limited in their choice of hospital by the characteristics of the disease; and (iv) we do not find evidence of team decisions or group practices regarding treatment. After controlling for patient i's physician, a direct effect of other physicians on patient i's probability of accessing treatment seems extremely unlikely in our setting. Finally, the timing assumption helps us to exclude simultaneity effects in the first stage.

Second, we determine the relevance of the instrument by estimating the first-stage Equation (2) in the next section.

3.2 Baseline results

We begin by estimating Equation (1): first, we determine that, in our data, the optimal number of periods in calculating the share of *untreated* patients is $T = 3$, and second, we focus on the estimation error given the relatively small number of patients in a neighborhood.[14] Aggregating the shares over the three years partially addresses this concern. We also restrict our sample to neighborhoods with at least ten patients; increasing the cut-off threshold does not meaningfully impact our results (it only reduces the sample size). Finally, we estimate risk-adjusted treatment rates and apply hierarchical modeling techniques to those rates for reliability: see Dimick et al. (2010).

Table 3 presents the results for the OLS and instrumental variable estimations. In all specifications, we control for the baseline attributes related to the patient (health and socio-demographics), the disease, and the neighborhood. In the baseline specification, we also use fixed effects at the year, two-digit zip code, and physician level. Both year and physician fixed effect control for supply-side drivers of access to treatment. As, in our sample, physicians do not move across hospitals, physician fixed effects absorb any hospital effects. We also construct a proxy of hospital congestion: the lag between the diagnosis and first consultation with an oncologist equals, on average, 29 days. Our results are essentially unchanged by the

[14] Both AIC/BIC criteria and a Likelihood Ratio test indicate that the optimal lag length equals three. To avoid the loss of too many observations, we use $T = 2$ for the year 2010.

inclusion of this control.

Column 1 reports the OLS specification, which does not instrument for the share of untreated patients living in the same neighborhood. The result suggests that a one percentage point increase in the share of untreated patients is associated with a 0.07 percentage point decrease in the patient's probability of treatment. Column 2 presents results when we instrument for the share of untreated patients using the average treatment propensity of physicians, and column 3 shows the first stage results. The first stage (column 3) shows that relevance is high, as the average treatment propensity of physicians is negatively correlated with the share of patients left untreated; the $F-$statistic equals 98.27, suggesting that we do not have a weak instrument problem. The estimated effect of biased beliefs and stigma on the probability of treatment using our IV estimator is over two times as large as the OLS estimate, and is statistically significant at the five percent level. A one percentage point increase in the share of untreated patients decreases the probability of accessing treatment by 0.17 percentage point. For comparison, moving from an area of low to high treatment (from the 10^{th} to the 90^{th} percentile in the distribution) increases the treatment probability by three percentage points; moving from the first to the fifth quintile of the income distribution increases the treatment probability by seven percentage points. Intuitively, health and demographic attributes are stronger drivers of treatment probabilities; for example, holding all variables at their mean values, the treatment probability decreases from 62% for the 45-49 age group to 30% for the 80-84 age group.[15]

That the IV estimates predict more negative effects than OLS has three possible concurrent explanations. First, social effects may be measured with error, so that OLS understates the effect relative to IV. Second, because of heterogeneous effects, IV and OLS are not directly comparable, as OLS estimates the average treatment effect and IV estimates a weighted local average effect for the patients whose latent unobserved sensitivity to social stigma is

[15] As we demean the data to remove the fixed effects, we implicitly assume that future period values of the share of untreated patients are uncorrelated with the current period error term. We perform a diagnostic test similar in spirit to the one proposed by Wooldridge (2010) and add the lead share of untreated patients as an additional regressor. The only reason to find statistically significant results from the lead share of untreated patients is the presence of correlated trends that are influencing both the reference group and the focal patient. The estimated coefficient of the regressor "lead share" is practically zero and statistically insignificant.

triggered by the treatment propensity of the physician. Third, correlated effects that work within a neighborhood may affect the OLS estimates.

In columns 4 to 6 of Table 3, we estimate risk-adjusted treatment rates and apply hierarchical modeling techniques to those rates for reliability following Dimick et al. (2010). Using risk-adjusted "shrunk" rates, we find a larger coefficient of social effects. This result is consistent with some degree of classical measurement error and, as a consequence, the attenuation bias in our measure of treatment rates; we, therefore, consider our estimates of social effects as conservative.

Table 4 illustrates the robustness of our results to a variety of checks. In column 2 of Panel A, we use the subsample of patients for which we have more detailed tumor characteristics, including the size, the presence, and location of metastases; our results do not change. In column 3 of Panel A, we test whether or not the effects we find are driven by a patient reacting to the health outcomes of fellow patients. Observing health outcomes may also deter access to treatment as the focal patients would Bayesian-update the negative prior that lung cancer is a death sentence. However, when we control for the observed average survival of past patients, the coefficient of the share of untreated patients becomes more negative, while the coefficient of past patients' survival is practically zero. The result suggests that Bayesian updating on the basis of observed outcomes does not play a role in our setting. Column 1 of Panel B presents a specification with a rich set of fixed effects at the three-digit zip code level, in addition to year and physician; again, our results hold. Columns 2 and 3 focus on the role of the reference group. We run our specification on subsamples defined by the intensity of social ties, as proxied by the Social Connectedness Index developed by Bailey et al. (2018). We find that stigma is only a barrier to access treatment when social ties are intense within a community. When social relations are intense in the neighborhood (the Social Connectedness Index is equal to or above quintile 3 of its distribution), the coefficient of stigma is more negative and statistically significant (-0.34); on the other hand, when social ties are loose (quintiles 1 and 2), the coefficient of social stigma is not statistically different from zero.

Placebo tests Table 5 provides a set of placebo tests; we apply the same identification strategy to patients affected by other cancers who should not feel the same degree of social discrimination or hold biased beliefs about the effectiveness of their treatment. Panel A contrasts the sample of lung and colorectal cancer patients; we use hospital fixed effects rather than physician fixed effects because the sample size of colorectal patients is severely reduced when we restrict our sample to neighborhoods with a minimum number of ten patients. For colorectal patients, both the OLS and IV estimates reveal no statistically significant relationship between the treatment rate of the reference group and the patient's probability of accessing treatment.

Panel B presents the same regressions in which we pooled colorectal, breast and prostate cancers; this strategy overcomes the issue of sample size, but prevents the use of cancer-specific covariates in the measurement of physicians' treatment propensity and the drivers of access to treatment; we replace these covariates with cancer site fixed effects. In these specifications, we use the same set of fixed effects employed in the baseline specification (at the year, two-digit zip code, and physician level). These additional placebo tests provide further evidence of the effectiveness of our identification strategy. The OLS estimates indicate a very small but statistically significant relationship between the the treatment rate of the reference group and the patient's probability of accessing treatment; this result could be an indication of the endogeneity issue that even a rich set set of fixed effects does not completely absorb. However, the IV results show no statistically significant relationship, which further supports the validity of our instrumentation strategy. The first stage is also strong for the pooled sample of colorectal, breast, and prostate patients, showing that a physician's treatment propensity matters for all cancer types.

3.3 Mechanisms

Smoking behavior We provide insights into the mechanisms generating our social effect results. We start by looking at the role of smoking behavior in the decision to take up treatment, comparing active smokers to non-smokers. Stigma and biased beliefs are inherently related to smoking, as the emphasis placed on cancer prevention messages may have negative

consequences on smokers, with the result that they feel "undeserving" of medical care.

We have information on the smoking status of patients diagnosed after 2014, thanks to the introduction of a smoking cessation program, where all newly diagnosed cancer patients are surveyed about their smoking habits. For patients with a cancer diagnosis after 2014, we observe whether the patient self-reported as being a current smoker or indicated they had smoked within the past six months. The Appendix reports summary statistics on patients affected by the most frequently occurring cancers: lung, colorectal, breast, and prostate. Table B.14 compares smokers versus non-smokers, Table B.15 compares smokers affected by lung cancer versus smokers affected by colorectal, breast, and prostate cancer, and Table B.16 compares smokers affected by lung cancer versus non-smokers affected by lung cancer. The most notable features are that: (i) the socio-demographic characteristics of all smokers (lung, colorectal, breast, and prostate) are similar; (ii) treatment rates for smokers with colorectal, breast, and prostate cancer are comparable to those for non-smokers; (iii) treatment rates for smokers with lung cancer are significantly lower than those for non-smokers; and (iv) smokers affected by lung cancer are significantly younger than non-smoker lung cancer patients and healthier beyond cancer. In sum, the summary statistics suggest that smokers affected by lung cancer face a higher barrier to accessing treatment than smokers affected by other cancers.

Since we observe the smoking status for a subsample of patients, we can directly test the hypothesis that smokers may more intensely suffer negative stereotypes regarding lung cancer. We perform the regression on the sample of lung cancer patients reporting to be active smokers. Columns 1 and 2 of Table 6 show that both the OLS and the IV coefficients on the share of untreated neighbors are larger (-0.35 and -0.88) and statistically significant. In our view, the results confirm that we are identifying a social discrimination effect, which smokers feel more strongly.

The literature also documents that, in general, smokers tend to exhibit lower adherence to medical guidelines, lower use of healthcare, and higher discount rates with respect to non-smokers: see Cutler et al. (2000), Arcidiacono et al. (2007), Harrison et al. (2010), Darden

and Kaestner (2022). Table B.14 shows that these features are also present in our data.[16] To test whether social effects are driven by smoker-specific attributes rather than negative stereotypes linked to lung cancer, we consider smokers affected by other cancer types as a placebo, with the expectation that the choice of the reference group would not affect the patient's probability of accessing treatment if we were estimating social discrimination specific to lung cancer. Columns 3 and 4 of Table 6 shows that the OLS and IV coefficients on the share of untreated neighbors are not statistically different from zero: our placebo test suggests that alternative explanations of our results related to the general attitude of smokers towards treatment and medical guidance do not seem to hold.

The impact of social effects on the timing and severity of the diagnosis The medical literature documents that feelings of stigmatization and psychological distress may delay seeking medical help: see Leveälahti et al. (2007), Carter-Harris (2015). At the same time, the majority of lung cancers are discovered at an advanced stage simply because the diagnosis of the disease is difficult: importantly, lung cancer is asymptomatic in its early stages with symptoms developing later that may be mistaken for an infection or the long-term effects of smoking. Screening programs are limited and, where present, often target specific populations. In Ontario, no screening program existed during the sample period, and stage IV diagnoses represent half of the diagnoses. This share is stable over time and exhibits limited geographic variation. However, when an individual has symptoms consistent with lung cancer but waits to seek medical attention, the disease can advance exponentially.

First, we test whether social effects impact the stage of the disease at diagnosis. The question addresses the issue of selection in the sample of patients. We regress the stage at diagnosis on all baseline attributes related to the patient (health and socio-demographics), the disease, and the neighborhood. We use two definitions of advanced stage: the first includes both stage III and stage IV (around two-thirds of all diagnoses in our data); and the second considers only metastatic patients (stage IV), versus all the other stages. Regardless of the definition, we show that the stage at diagnosis is mainly determined by the health

[16] Ziebarth (2018) documents a downward bias in risk perceptions about the probability of developing smoking-related cancers and their mortality rates.

and tumor characteristics of individual patients, as patients in poorer health tend to be diagnosed at an earlier stage as opposed to healthier patients. This result is in line with the so-called "waiting time paradox", as documented in the medical literature, a phenomenon whereby patients in poorer health are diagnosed at an earlier stage because the healthcare system more promptly instigates investigations of sicker patients: see Tørring et al. (2013). Notably, socio-economic variables at the patient and the neighborhood level do not impact the disease's discovery stage; columns 5 and 6 of Table 6 show that the share of untreated patients living in the same neighborhood has no effect either. We cannot use the same identification strategy to instrument for the endogeneity of our variable of interest; indeed, matching the physician for early stages would prove impossible as multiple physicians and treatment options are available. However, we can safely infer that social effects are unlikely to drive the stage at diagnosis, as well as all other non-health characteristics at the patient or neighborhood level.

Second, conditional on the stage at diagnosis, we investigate whether social effects are associated with delays in seeking medical care. In all our specifications we control for the symptoms that the patient presents at diagnosis. Under the supervision of a clinician, we categorize the symptoms according to a severity scale of 1-3 and based on whether the diagnosis occurs at the emergency department. We regress our measure of the severity of symptoms at diagnosis against the covariates at the patient and neighborhood level. Column 7 of Table 6 shows that the estimated effect of the share of patients left untreated in the neighborhood on the severity of diagnosis is positive and statistically significant; the coefficient indicates that a one percentage point increase in social effects leads to a 0.26 increase in the severity score. The literature qualitatively documents how social stigma is a barrier to seeking medical help through surveys: we provide a quantification of that effect.

3.4 Survey evidence

As social stigma is not directly observed in the data, we provide complementary evidence suggesting that the estimated social effects can be explained by the role of biased beliefs and social stigma associated with lung cancer. We conduct a survey of a representative

sample of Ontarians to elicit direct measures of attitudes towards lung cancer. Specifically, as part of a larger telephone survey administered across Canada by a specialized survey center, we design five closed-ended questions about perceptions and attitudes toward smoking and lung cancer, which are asked to a representative sample of 402 adults across Ontario. The questions cover: attitudes towards smokers, sympathy towards lung cancer patients, perceptions of the effectiveness of treatment, and support for research funding. Appendix Table D.19 reports the survey questions and a summary of the responses.

Survey responses suggest that around 23 percent of Ontarians report that people around them feel less sympathy for lung cancer patients than for patients affected by other tumors, 20 percent personally feel less sympathetic, 14 percent feel that treating lung cancer is not worthwhile, while 13 percent would prefer supporting research on different cancer types over lung cancer. These three measures of attitude towards lung cancer (sympathy, beliefs on the effectiveness of treatment, and support for research) are strongly correlated with each other. Therefore, we interpret our estimates as including both the stigma and a sense of a hopelessness that the disease is not worth treating. These results are in line with a 2010 survey by the Global Lung Cancer Coalition, in which 22 percent of Canadians admit feeling less sympathy for lung cancer patients: see Ipsos MORI (2010). Survey responses further indicate that male and older respondents are more likely to hold a negative attitude toward lung cancer.

We examine how the elicited variation in the degree of stigma correlates with the measure that we construct in our data. As we do not have a sufficient number of survey respondents by neighborhood, we check the degree of correlation between the quintiles of the untreated share of patients in the data and the average degree of stigma from the survey, calculated for each quintile. The two measures are positively correlated, with a correlation coefficient equal to 0.52.

Table 3: Social effects in access to treatment: baseline results

	Baseline			'Shrunk' share untreated		
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV	First stage	OLS	IV	First stage
Share untreated	-0.072	-0.167		-0.033	-0.379	
	(0.033)	(0.073)		(0.051)	(0.133)	
Physician treatment propensity			-0.150			-0.130
			(0.010)			(0.011)
Share adenocarcinoma			-0.035			-0.051
			(0.024)			(0.023)
Share additional malignancies			-0.216			0.129
			(0.148)			(0.152)
Controls:						
Patient health	Yes	Yes	Yes	Yes	Yes	Yes
Patient socio-demo	Yes	Yes	Yes	Yes	Yes	Yes
3-digit zip code	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects:						
Physician	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Two-digit zip code	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,882	7,882	7,882	7,882	7,882	7,882
F-statistic			98.27			89.57

The dependent variable in each specification is whether the patient is treated (0/1) for lung cancer. An observation is a patient-diagnosis year. The "share untreated" refers to the cumulative share of untreated patients diagnosed in the three previous years in the same three-digit zip code. Columns 1 and 4 present OLS social effects results. Colums 2 and 5 present IV social effects results, instrumenting for "share untreated" using the average treatment propensity of physicians treating the reference group, the percentage of patients in the reference group affected by adenocarcinomas and synchronous multiple lung cancers. Clustered standard errors at the two-digit zip code are in parentheses (45 clusters). The $F-$statistic on the excluded instrument refers to the Wald version of the Kleibergen and Paap (2006) rk-statistic on the excluded instrumental variables for non-i.i.d. errors.

Table 4: Social effects in access to treatment: robustness checks

	Panel A		
	(1) Baseline	(2) Controls for metastases	(3) Control for survival past patients
Share untreated	-0.167	-0.163	-0.179
	(0.073)	(0.074)	(0.095)
Controls:			
Patient health	Yes	Yes	Yes
Patient socio-demo	Yes	Yes	Yes
3-digit zip code	Yes	Yes	Yes
Fixes effects:			
Physician	Yes	Yes	Yes
Year	Yes	Yes	Yes
2-digit zip code	Yes	Yes	Yes
Observations	7,882	6,245	7,882
	Panel B		
	(1) 3-digit zip code	(2) High social connectedness	(3) Low social connectedness
Share untreated	-0.193	-0.344	0.160
	(0.078)	(0.105)	(0.132)
Controls:			
Patient health	Yes	Yes	Yes
Patient socio-demo	Yes	Yes	Yes
3-digit zip code	Yes	Yes	Yes
Past patient characteristics	Yes	Yes	Yes
Fixes effects:			
Physician	Yes	Yes	Yes
Year	Yes	Yes	Yes
Three-digit zip code	Yes	Yes	Yes
Observations	7,882	6,850	4,874

The dependent variable in each specification is whether the patient is treated (0/1) for lung cancer. An observation is a patient-diagnosis year. The "share untreated" refers to the cumulative share of untreated patients diagnosed in the three previous years living in the same three-digit zip code. All specifications present social effects instrumenting for "share untreated" using the average treatment propensity of physicians treating the reference group, the percentage of patients in the reference group affected by adenocarcinomas and synchronous multiple lung cancers. Clustered standard errors at the two-digit zip code are in parentheses (45 clusters).

Table 5: Social effects in access to treatment: placebo tests

	(1)	(2)	(3)	(4)
	Panel A			
	Lung		Colon	
	OLS	IV	OLS	IV
Share untreated	-0.086	-0.239	0.022	0.346
	(0.043)	(0.088)	(0.081)	(0.246)
Controls:				
Patient health	Yes	Yes	Yes	Yes
Patient socio-demo	Yes	Yes	Yes	Yes
Physician characteristics	Yes	Yes	Yes	Yes
3-digit zip code	Yes	Yes	Yes	Yes
Fixed effects:				
Physician	No	No	No	No
Year	Yes	Yes	Yes	Yes
2-digit zip code	Yes	Yes	Yes	Yes
Hospital	Yes	Yes	Yes	Yes
Observations	7,882	7,882	1,490	1,493
	Panel B			
	Lung		Pooled	
	OLS	IV	OLS	IV
Share untreated	-0.072	-0.167	-0.0386	0.291
	(0.033)	(0.073)	(0.018)	(0.181)
Controls:				
Patient health	Yes	Yes	Yes	Yes
Patient socio-demo	Yes	Yes	Yes	Yes
Physician charact	Yes	Yes	No	No
3-digit zip code	Yes	Yes	Yes	Yes
Fixed effects:				
Physician	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
2-digit zip code	Yes	Yes	Yes	Yes
Hospital	No	No	No	No
Observations	7,882	7,882	9,148	9,176

The dependent variable in each specification is whether the patient is treated (0/1) for lung cancer (columns 1 and 2); colorectal cancer (columns 3 and 4 of Panel A); colorectal, breast, and prostate (columns 3 and 4 of Panel B). An observation is a patient-diagnosis year. The "share untreated" refers to the cumulative share of untreated patients diagnosed in the three previous years living in the same three-digit zip code. IV specifications present social effects instrumenting for "share untreated" using the average treatment propensity of physicians treating the reference group, the percentage of patients in the reference group affected by adenocarcinomas and synchronous multiple lung cancers. Clustered standard errors at the two-digit zip code are in parentheses (45 clusters).

Table 6: The impact of social effects on the timing and severity of the diagnosis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Lung smokers		Placebo smokers		All lung cancer patients		Stage IV lung cancer patients
	Treatment 0/1				Stage III and IV 0/1	Stage IV 0/1	Degree of severity 1-3
	OLS	IV	OLS	IV	OLS	OLS	IV
Share untreated	-0.347	-0.876	0.104	-0.591	-0.0007	0.0017	0.263
	(0.107)	(0.284)	(0.066)	(0.506)	(0.011)	(0.013)	(0.154)
Controls:							
Patient health	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Patient socio-demo	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3-digit zip code	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Physician charact	Yes	Yes	Yes	Yes	No	No	No
Fixes effects:							
Physician	No	No	No	No	No	No	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2-digit zip code	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	924	924	499	499	35,648	35,648	7,882

The dependent variable in columns 1 to 4 is whether the patient is treated (0/1) for lung cancer. The dependent variable in column 5 is a dummy identifying advanced stage (stages III and IV) versus non-advanced stage (stages I and II). The dependent variable in column 6 is stage IV versus other stages (0/1). The dependent variable in column 7 is the severity of symptoms at diagnosis (scale 1 to 3). An observation is a patient-diagnosis year. The "share untreated" refers to the cumulative share of untreated patients diagnosed in the three previous years living in the same three-digit zip code. Columns 2, 4, and 7 present social effects instrumenting for "share untreated" using the average treatment propensity of physicians treating the reference group, the percentage of patients in the reference group affected by adenocarcinomas and synchronous multiple lung cancers. Clustered standard errors at the two-digit zip code are in parentheses (45 clusters).

4 A structural model of treatment choice

We now develop a model of specific treatment choice for metastatic lung cancer, focusing on the first treatment choice at the time the disease is diagnosed (first-line therapy).

Following the notation adopted above, let there be $i = 1, ..., I$ patients with stage IV lung cancer diagnosed in each year t. For each patient i, the choice is between treating or not treating the disease: $g = 0, 1$. Conditional on treatment, there are four treatment options: $j = 1, ..., 4$: (i) cisplatin-based chemotherapy; (ii) carboplatin-based chemotherapy;

(iii) single agent chemotherapy; and (iv) innovative therapy (targeted and immunotherapy). The first three options fall under the category of the standard of care but differ in the drugs used and their toxicity profile. Cisplatin doublets (a combination of cisplatin and another chemotherapeutic agent) are considered more effective than carboplatin doublets but are more toxic and less tolerated and hence not recommended for older or sicker patients. Single-agent regimens are used for patients who cannot tolerate platinum-based therapy (cisplatin and carboplatin).[17]

The indirect utility of each patient i from pursuing treatment j is assumed to be additively separable into a component that varies across alternatives j within the treatment nest (V_{ijt}), and a component (W_{igt}) that varies across nests g:

$$u_{ijt} = V_{ijt} + W_{igt} + \varepsilon_{ijt}. \tag{3}$$

The random component of utility follows the distributional assumptions of a two-level nested logit model (McFadden (1978)), which allows valuations to be correlated across alternatives in the same nest. At the top level, there are two nests (the choice is binary): the "treatment" nest $g = 1$, which includes the treatment options, and the "no-treatment" nest $g = 0$, which is a degenerate nest with only alternative $j = 0$. Individual i's utility for the no-treatment option is:

$$u_{i0t} = W_{i0t} + \varepsilon_{i0t}.$$

At the bottom level, the treatment nest consists of the J treatment options. The distribution of ε_{ijt} contains the nesting parameter λ, with $0 < \lambda \leq 1$. The parameter proxies for the degree of dissimilarity of treatment options belonging to the "treatment" nest. As λ tends to

[17] An extension of the present model would be to consider the decision to refer or not a patient to a cancer center by the primary care physicians; variation in referral could contribute to practice variation, as discrimination issues and therapeutic nihilism may impact the referral decision as well. We match patients' records with physicians' claim records to identify the referring physician at the time of diagnosis. The most common specialties of referring doctors are internist, respirologist, and family physician. Of the 15,761 patients diagnosed with metastatic lung cancer, over 80% were referred to a medical oncologist. The most critical drivers of lack of referral are the diagnosis at arrival, health status, and age. Social effects do not appear to be a determinant of referral. We conclude that adding referral to the sequence of decisions that we model would not alter the conclusions of our study.

one, the distribution of the error terms ε_{ijt} approaches an i.i.d. extreme value distribution, so correlation in the error between treatment options is weak. As it tends to zero, the error terms become perfectly correlated and patients/physicians choose the alternative with the highest observable utility. The nested logit results in simple expressions for the choice probabilities. Following Train (2009), we characterize the nested choice as two logit equations. The probability of selecting treatment option j is the product of the conditional probability that treatment option j is chosen in the "treatment" nest (the bottom-level logit) and the marginal probability that patient i chooses to be treated (the top-level logit):

$$s_{ijt} = s_{ijt|g} \cdot s_{igt}.$$

Choice between treatment options The bottom-level choice probabilities are:

$$s_{ijt|g} = \frac{\exp\left(V_{ijt}/\lambda\right)}{\sum_{l \in J} \exp\left(V_{ilt}/\lambda\right)}.$$

We define the inclusive value term I_{i1t} as a measure of the expected aggregate utility that patient i receives from the choice among the alternatives in the nest "treatment" ($g = 1$):

$$I_{i1t} = \log\left[\sum_{j \in J} \exp\left(V_{ijt}/\lambda\right)\right].$$

Choice of whether to pursue treatment The top-level choice probability that a patient chooses to pursue treatment ($g = 1$) is:

$$s_{i1t} = \frac{\exp\left(W_{i1t} + \lambda I_{i1t}\right)}{\exp(W_{i0t}) + \exp\left(W_{i1t} + \lambda I_{i1t}\right)}.$$

At the top level, all patients' and treatments' characteristics included at the bottom level indirectly enter the decision of accessing treatment through the inclusive value term I_{it}.

The probability that patient i chooses the no-treatment option s_{i0t} is:

$$s_{i0t} = 1 - s_{i1t}.$$

We now specify the two deterministic components of utility $(V_{ijt} + W_{igt})$. The first component, V_{ijt}, which depends on variables that describe each treatment option, is specified as follows:

$$V_{ijt} = \alpha_{j1} + x'_{it}\alpha_{j2},$$

where x_{it} is a vector of attributes related to the health of the patient and the disease at the time of diagnosis. At the bottom level, we include physician's attributes, as the limited sample size does not allow the use of physician fixed effects: physician's sex, age, tenure, and two measures of workload to proxy for experience (annual caseload related to lung cancer patients, and total yearly consultations). Finally, all treatment-specific characteristics are absorbed by the constant α_{j1}.[18]

At the top level, as the choice is binary, only relative levels of determinants to access to treatment matter. The second component, W_{igt}, which depends on variables describing the "treatment" against the "no-treatment" nest, is specified similarly to equation (1) and depends on:

1. the outcome of the reference group (social effects): $\overline{d}_{it}$;
2. patient attributes (x_{it}) and patient-specific socio-demographics (z_{it});
3. reference group and neighborhood-specific characteristics, summarized by the vector η_{rt};
4. physician fixed effects to account for the supply side: $\eta_{p(i)}$.

The deterministic component of utility related to the choice of accessing treatment can then be written as:

[18] We do not include the price of each regimen: from the patient's point of view, all drugs included in the regimens are publicly funded. Physicians are on alternative funding plans, and the choice of therapy has no impact on their compensation, as well as their choice of whether to treat the patient or not.

$$W_{igt} = \beta_1 \bar{d}_{it} + x_{it}\beta_2 + z_{it}\beta_3 + \eta_{rt} + \eta_{p(i)}. \tag{4}$$

We define the outcome of the reference group as in Section 3 and follow the same identification strategy to identify the social effects.

4.1 Nested logit specification: results

We present the estimated coefficients of the discrete choice model described by equation (3). We use sequential maximum likelihood methods to estimate the nested logit model. At the upper level, we have a binary choice specification with an endogenous variable, the share of untreated neighbors to proxy for the social effects. To identify social effects in treatment choices, we use a control-function approach: see Heckman (1978), Blundell and Powell (2004). We derive a proxy variable that conditions on the part of the social effects that depends on the unobservable drivers in the treatment decision; that is, the remaining variation in social effects becomes independent of the errors. In practice, we estimate the model in two steps. In the first step, we regress the endogenous share of untreated patients on a set of instruments. In the second step, we derive the errors from the first stage as an additional regressor in the main specification. To estimate the first step, we use variables that explain the share of untreated patients in a neighborhood: the average treatment propensity of physicians treating patients in the reference group, the percentage of patients in the reference group affected by adenocarcinomas and synchronous multiple lung cancers (as instruments), health and socio-demographic attributes related to the neighborhood, and fixed effects at the two-digit zip code, year, and physician level.[19]

We first discuss the determinants of the choice of a specific regimen (bottom level). Table 7 reports the bottom-level results; the base treatment option is cisplatin, which is part of the standard of care and tends to be relatively aggressive compared to other options. Age and health condition at diagnosis (a higher value of the Charlson index indicates worse health)

[19] Given the presence of physician fixed effects at the upper level, we account for the bias arising from the inclusion of individual fixed effects in a non-linear model. We find that, in our setting, the correction has a minimal impact on the parameter estimates.

are the most important drivers of the decision on the type of treatment. Consistent with clinical guidelines, sicker patients are more likely to receive single-agent therapy. Those with squamous cancer are unlikely to receive innovative regimens; this result aligns with the indications of those drugs.

Table 8 reports the maximum likelihood estimates of the top level, the determinants of participation in treatment. The coefficient of the main variable of interest (the share of untreated patients) is negative and precisely estimated. Its marginal effect is similar to the linear specification: an increase of one percentage point in the share of untreated patients is associated with a decrease in the probability of accessing treatment equal to 0.19 percentage points. Intuitively, the patient's age, tumor, and health attributes at diagnosis are the most important drivers of treatment participation. Patients' socio-demographic characteristics also affect treatment participation: higher-income patients and those from wealthier areas are more likely to access treatment. The coefficient of the inclusive value, λ, is in the range of zero to one, and we can reject the logit value of $\lambda = 1$.

As a placebo test, we place our proxy of social effects, the share of untreated patients, at the bottom level, where we study the choice of a specific regimen. This placebo test is helpful to rule out that social effects could impact the probability of accessing each treatment type differently, possibly depending on their side effects and their visibility. However, we do not expect to find statistically significant results; only informed patients would be aware of the side effects for each treatment type, and we expect that those patients would also understand the effectiveness of the treatment. Table D.20 in the Appendix verifies that social effects have no statistically significant relationship with the choice of a specific treatment, even though we are not using any patient or neighborhood-specific socioeconomic attributes.

Table 7: Regimen/therapy choices: bottom level of a nested logit model

	(1) Carboplatin therapy	(2) Single-agent therapy	(3) Innovative therapy
Surgery	-0.966	-1.626	-0.811
(0/1)	(0.258)	(0.657)	(0.309)
Adenocarcinoma	0.508	0.0682	0.732
(0/1)	(0.258)	(0.562)	(0.307)
Squamous cell	0.308	0.058	-0.980
(0/1)	(0.274)	(0.591)	(0.354)
Charlson index	0.0982	0.274	-0.158
medium	(0.104)	(0.204)	(0.121)
Charlson index	0.431	0.709	-0.148
high	(0.130)	(0.236)	(0.157)
Controls:			
Patient health	Yes	Yes	Yes
Patient socio-demo	Yes	Yes	Yes
3-digit zip code	No	No	No
Physician characteristics	Yes	Yes	Yes
Fixed effects:			
Physician	No	No	No
Year	Yes	Yes	Yes
Hospital	Yes	Yes	Yes
Observations		14,592	

The table reports the parameter estimates and standard errors of selected variables for the bottom level of a nested logit model of therapy choice: cisplatin, carboplatin, single-agent therapy, and innovative therapy. The excluded base alternative is cisplatin. The excluded health status category is the lowest Charlson (most healthy individual). The model controls for a constant for each therapy alternative. Standard errors are in parentheses.

Table 8: Treatment participation - A disaggregate nested logit model

	Logit
Share untreated	-1.194
	(0.606)
Inclusive value	0.256
	(0.189)
Controls:	
Patient health	Yes
Patient socio-demo	Yes
3-digit zip code	Yes
Past patient characteristics	Yes
Fixed effects:	
Physician	Yes
Year	Yes
FS2	Yes
Observations	7,127

The table reports the parameter estimates and standard errors for the upper level of the nested logit model where the choice is whether to pursue treatment (0/1). The "share untreated" refers to the cumulative share of untreated patients diagnosed in the three previous years in the same three-digit zip code. Control-function correction is used to address the endogeneity of "share untreated". Clustered standard errors at the two-digit zip code are in parentheses (45 clusters).

5 Counterfactual simulations

Mitigation of stigma and the cost of systemic therapy We now consider what would happen to lung cancer treatment rates, particularly to the adoption of innovative therapies, if patients lived in areas where treatment rates are higher. Table 9 shows the effect of placing patients in an area of low social discrimination, the risk-adjusted 10^{th} percentile of the variable share untreated, which corresponds to a share of untreated patients of 45 percent, similar to colorectal cancer. Intuitively, the percentage of untreated patients decreases by 4 percent, with an increase of 3 percent in the number of patients pursuing innovative treatment.

For each patient, we calculate the total expenditure on systemic therapy drugs, as we have information on the patient's survival, the prices of regimens, including accessory costs[20], and

[20]For each regimen, the costs include: the number of chemotherapy suite visits, the number of ambu-

average dose and frequency of administration. Finally, we use the estimator developed by Zhao and Tian (2001) to estimate the mean healthcare costs accounting for right censoring and the patients' cost history. The calculated costs by regimen align with the estimates from the literature (de Oliveira et al., 2013) and pCODR, the Canadian review board for the approval of oncological drugs.

Following a cost-effectiveness approach that typically guides policy decisions when evaluating a given therapy, we compare these treatment costs with the value for a quality-adjusted life year (QALY). Moving patients to an area of low social discrimination would imply an additional overall cost of CAD 1.3 million for innovative treatment, which is much higher than the increase in costs if those patients were treated with the standard of care. However, the gain in survival is also higher, which justifies the use of innovative therapies with respect to the current "no treatment" scenario: the additional annual cost amounts to CAD 22,913 (USD 17,000) per patient, which is much lower than the gain of CAD 65,000 (USD 50,000) per year of quality life. This has been the de facto standard used by the Canadian medical agency to determine whether to cover drugs or medical procedures.

If the incremental patients are treated instead with cisplatin-based chemotherapy (the standard of care type with the longest survival), we would obtain a cost equal to CAD 7,364 per patient but a loss in terms of survival equal to 160 days, or CAD 28,493 QALY. Cost-benefit is roughly aligned in the scenario "cisplatin" versus "innovative" when looking exclusively at the costs of systemic therapy. Below we consider the overall costs of patients under each scenario.

Total costs We now compare the total costs of treating the additional patients when placing the patients in an area of low social discrimination. We compute individual-level cost data using a macro-based costing methodology that combines information from all datasets presented in Appendix Table A.1. In addition to the cost of administering the therapy discussed above, we also consider a detailed breakdown of costs that we aggregate

latory clinic visits during treatment, nursing and pharmacy workload time to prepare and administer the specific regimen, drugs not included in the New Drug Funding Program and supportive drugs, manager and clerical time for managing and scheduling in the cancer center, and other supplies and costs, including medical/surgical supplies.

into six categories: inpatient hospitalization, outpatient services, emergency department visits, prescription drugs, rehabilitation and long-term care, and physician services. Table 10 reports costs estimated based on Zhao and Tian (2001) accounting for right censoring and the patients' cost history.

While untreated patients have the lowest costs because of their lower survival, they still use significant resources. Our estimates of elevated end-of-life spending, especially driven by inpatient admissions, align with the literature: see Zeltzer et al. (2021). Patients treated with innovative therapy generate the highest costs, but those costs are driven by the high price of the treatment itself, since most of these drugs are still under patent protection. For several other cost categories, these patients are comparable to those treated with the standard of care. In particular, comparing patients treated with innovative therapy to those treated with cisplatin-based chemotherapy shows that their costs are lower for some categories, such as outpatient and emergency visits. Indeed, cisplatin-based therapy tends to be quite aggressive: it can be administered only to healthy patients at the hospital, it implies a lower quality of life, and more frequent use of emergency/urgent care facilities. Our data show that patients treated with cisplatin-based chemotherapy are 36 percent more likely to use emergency care than those treated with innovative therapy, resulting in additional costs for the health system.

To make these costs more comparable across therapies and to account for the different survival, we compute the incremental cost-effectiveness ratio (ICER) per life-year. We find that innovative therapies are between CAD 63,000 and 68,000 more expensive than alternative options per additional year of life. These values should be compared to the value of statistical life: if we use the commonly applied (conservative) estimate of CAD 100,000 per year, lowering social stigma and negative stereotypes would not only benefit patients but also be cost-effective.

Table 9: The effect of mitigating stigma

	Untreated	Cisplatin	Carboplatin	Single-agent	Innovative
Nb. patients - Base	3,630	936	1,396	206	956
Nb. patients - CF	3,487	973	1,462	216	986
Δ patients	-143	37	66	10	30
Estimated cost of treatment (drugs only)					
Estimated survival (dd)	142	522	438	355	682
Avg. cost per patient	-	7,364	5,562	3,211	42,835
Δ cost (100,000$)	-	2.72	3.67	0.32	12.85

The table reports the change in the number of patients and related costs implied by placing all patients in the 10^{th} percentile of the share of untreated patients. The estimates are based on the parameter estimates reported in Table 8 and Table 7. The cost and survival estimates are based on Zhao and Tian (2001); the annual discount rate for the costs and survival time is fixed at 3%.

Table 10: Total costs from diagnosis to death or last contact

	Untreated	Cisplatin	Carboplatin	Single-agent	Innovative
Inpatient	22,138	25,598	23,116	25,536	25,601
Outpatient	6,820	43,258	34,105	27,310	36,964
Emergency	1,133	2,006	1,938	1,941	1,917
Drugs	1,645	23,394	20,168	11,301	54,499
Long term care	6,473	9,042	8,982	8,486	10,225
Physician	7,554	18,160	15,180	13,757	19,898
Total	45,763	121,459	103,489	88,331	149,104
Estimated survival	142	522	438	355	682

The table reports the average health costs by treatment type broken down into six categories: inpatient hospitalization, outpatient services, emergency department visits, prescription drugs, rehabilitation services and long-term care, and physician services. The cost and survival estimates are based on Zhao and Tian (2001); the annual discount rate for the costs and survival time is fixed at 3%.

6 Implications for R&D investment

We have documented that biased beliefs and stigma significantly deter treatment for lung cancer. In this section, we explore the implications of the lower number of treated patients on R&D investments. To quantify the relationship between market size (number of treated patients) and R&D spending, we match two publicly available datasets from the US. Our measure of innovation comes from the National Cancer Institute, which reports publicly funded R&D investment in cancer therapy. We collect the information for the period 2004-2018. Our measure of market size comes from the National Cancer Database, a nationwide oncology database that captures over 70% of all newly diagnosed cancers for 12 cancer sites in the US every year from more than 1,500 affiliated facilities. The database covers the period 2009-2018: it includes the number of cancer patients by year, cancer site, and therapy type, and records the first course of treatment, defined as the method of treatment administered to the patient before disease progression or recurrence. We match these two datasets and follow the American Society of Clinical Oncology guidelines to define which patients are treated for each cancer site and stage (stage I to stage IV). Summary statistics are reported in Table D.21 in the Appendix; R&D spending averages 0.16 million per cancer site/year and increases over time, from $1.9 million in 2003 to 2,2 million in 2018. In parallel, the total number of diagnosed patients also increases in the period 2009-2018, from 1.01 in 2009 to 1.19 million in 2018. Treatment rates average around 80%, with significant variation across cancer sites. Most of the variation in our variables comes from the between variation across cancer sites rather than the within cancer site variation over the years. The between standard deviation for the treatment rate is 11; the within standard deviation is 2.[21]

We estimate the following specification to recover the elasticity of R&D intensity with

[21] The overall number of cancer patients is slightly lower than those reported by the American Cancer Society, as the National Cancer Database does not provide universal coverage. The database does not include untreated patients who do not access the facilities affiliated with the clinical oncology database; hence, treatment rates tend to be overestimated. The use of fixed effects at the cancer site and year level partially addresses the issue of measurement error in the data. The presence of measurement error provides an additional argument for using an instrumental variables approach.

respect to market size:

$$\ln R\&D_{ct} = \alpha \ln(treated_{ct+l}) + \delta_t + \eta_c + \varepsilon_{ct}, \tag{5}$$

which relates R&D spending ($R\&D$) in period t for cancer site c to the number of treated patients ($treated$) in period $t + l$ (our measure of market size); the term δ_t is a year fixed effect, η_c a fixed effect specific to each cancer site, and ε_{ct} an unobserved shock to R&D spending. The coefficient α can be interpreted as the elasticity of R&D effort to market size. As firms rationally anticipate increases in market size and invest in R&D before demand materializes, we use both current ($l = 0$) and lead market size ($l = 5$). To deal with reverse causality between innovation and market size, we instrument $\ln(treated_{ct+l})$ using a measure of *potential* market size, the overall number of patients diagnosed in each period and cancer site. The instrument strongly correlates with the number of treated patients. The exclusion restriction requires that R&D effort should not directly cause changes in the overall number of patients diagnosed. It is reasonable to assume that the condition is satisfied as the diagnosis of cancer is solely based on the presence of malignant cells: R&D effort in diagnostic tools may influence the stage at which the diagnosis happens but not the diagnosis per se. Finally, we estimate the model in first differences to difference out η_c; the first difference estimator exploits cross-sectional variation in the data and requires a weaker exogeneity assumption than demeaning: see Cameron and Trivedi (2005).

Table 11 provides the results. All coefficient estimates suggest a positive relationship between pharmaceutical R&D intensity and market size. Column (1) reports the estimation results of Equation (5) by ordinary least squares: estimates are affected by endogeneity issues. Our preferred specifications deal with the possibility of reverse causality between innovation and market size using an instrumental variables approach (columns 2 and 3). The specifications yield a range of elasticities between 3.4 and 5.6 percent, meaning that a 10 percent increase in market size is associated with a 3.4 to 5.6 percent increase in R&D spending. These numbers are remarkably close to the elasticity estimates obtained by

Giaccotto et al. (2005), who also use R&D intensity as the dependent variable.[22]

Putting together the estimated impact of social effects on the number of treated patients and the elasticity of R&D intensity to market size, back-of-the-envelope calculations suggest that social stigma and biased beliefs are responsible for around a 2 percent of the gap in research funding for lung cancer with respect to other common cancers; this amounts to $7 million every year in US public funding alone.

Table 11: Market size and R&D intensity

	(1)	(2)	(3)
		$\ln R\&D_{ct}$	
$\ln treated_{ct}$	0.382	0.559	
	(0.293)	(0.208)	
$\ln treated_{ct+5}$			0.335
			(0.200)
Year FE	Yes	Yes	Yes
Cancer site FE	Yes	Yes	Yes
Observations	108	108	102
Method	OLS	IV	IV
R-squared	0.109	0.106	0.278

The table reports the OLS and IV estimates of log R&D spending on the number of treated patients. All specifications include cancer-site and year fixed effects. Clustered standard errors at the cancer site level are in parentheses.

7 Conclusion

Lung cancer is the most commonly diagnosed cancer worldwide, accounting for 13% of all new cancer cases. With a five-year survival rate that is the lowest among the leading cancers

[22] In Table 11, standard errors are clustered at the cancer-site level and shown in parentheses. Standard errors are panel-robust to permit errors to be correlated over time for a given cancer site and covariances to differ across cancer sites. While we have only 12 cancer sites, our clusters are perfectly balanced, with few observations per cluster (high homogeneity, low leverage, low influence), so conventional inference is reliable: see MacKinnon et al. (2022). A formal test rejects the null of heteroskedastic-robust standard errors against cluster-robust standard errors.

(lung, colorectal, breast, and prostate), it is also the leading cause of cancer-related deaths. Despite the significant potential for targeted and immunotherapies to improve lung cancer treatment, access to these therapies for lung cancer patients remains low. Low access to treatment is partly caused by belief biases and stigma surrounding lung cancer, which are associated with a reluctance to seek treatment and lower research funding for the disease.

Using administrative data on the population of patients diagnosed with advanced lung cancer in Ontario (Canada) over the last decade, we exploit the unique level of geographic detail to incorporate social stigma in a model of a patient's utility of pursuing treatment. We define biased beliefs and stigma as endogenous social effects and measure them as the share of patients within the same neighborhood who were diagnosed in the previous three years but did not receive treatment. To confirm that the share of untreated patients living in the neighborhood is a good proxy of societally biased beliefs and stigma, we conduct a survey of around 400 adults across Ontario to elicit a direct measure of attitudes towards lung cancer. The variation in the degree of stigma across communities in Ontario positively correlates with the measure we construct in our data, with a correlation coefficient of 0.52.

We develop a model of treatment participation and therapy choice in which patients base their own decisions on the decisions of the reference group. Identification rests on exogenous variation in the treatment propensity of physicians. Biased beliefs and stigma deter access to treatment. By placing all patients in a neighborhood characterized by low stigma, treatment rates increase by 4 percent and the use of innovative therapies by 3 percent. In addition, social effects account for around 2 percent of the gap in research funding for lung cancer, which amounts to $7 million every year in US public funding alone.

Our empirical results inform the policy debate on considering lung cancer stigma and improving societal understanding of lung cancer. We also offer strong evidence showing that patients face accessibility problems linked to stigma, which then slow the adoption of innovative treatments and lower the incentives to invest in R&D. We explore and quantify the link between social discrimination, adoption of innovation, and R&D investments. Recent works have investigated the role of social stigma in learning and reporting the status of stigmatized diseases such as HIV or mental health: see Thornton (2008), Yu (2019), Bharadwaj et al.

(2017), and Cronin et al. (2020). Future research on stigmatized diseases, for which scientific knowledge has produced significant therapeutic advances, will be helpful in understanding to what extent societal biases hinder the diffusion of innovation and, in turn, discourage further R&D investments.

Acknowledgement

This study contracted ICES Data & Analytic Services (DAS) and used de-identified data from the ICES Data Repository, which is managed by ICES with support from its funders and partners: Canada's Strategy for Patient-Oriented Research (SPOR), the Ontario SPOR Support Unit, the Canadian Institutes of Health Research and the Government of Ontario. The opinions, results and conclusions reported are those of the authors. No endorsement by ICES or any of its funders or partners is intended or should be inferred.

Parts of this material are based on data and information compiled and provided by CIHI. However, the analyses, conclusions, opinions, and statements expressed herein are those of the author, and not necessarily those of CIHI.

Parts of this material are based on data and information provided by Ontario Health (OH). The opinions, results, view, and conclusions reported in this paper are those of the authors and do not necessarily reflect those of OH. No endorsement by OH is intended or should be inferred.

References

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

Agarwal, Ruchir and Patrick Gaulé, "What Drives Innovation? Lessons from COVID-19 R&D," IMF Working Papers 2021/048, International Monetary Fund February 2021.

Agha, Leila and David Molitor, "The Local Influence of Pioneer Investigators on Technology Adoption: Evidence from New Cancer Drugs," *The Review of Economics and Statistics*, 2018, *100* (1), 29–44.

Aizer, Anna and Janet Currie, "Networks or neighborhoods? Correlations in the use of publicly-funded maternity care in California," *Journal of Public Economics*, 2004, *88* (12), 2573–2585.

Angrist, Joshua D., "The perils of peer effects," *Labour Economics*, 2014, *30*, 98–108.

Arcidiacono, Peter, Holger Sieg, and Frank Sloan, "Living Rationally under the Volcano? An Empirical Analysis of Heavy Drinking and Smoking," *International Economic Review*, 2007, *48* (1), 37–65.

Arenberg, Douglas, "Lung Cancer," in "Clinical Respiratory Medicine," Elsevier, 2012, pp. 801–809.

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong, "Social Connectedness: Measurement, Determinants, and Effects," *Journal of Economic Perspectives*, aug 2018, *32* (3), 259–280.

Bandiera, Oriana and Imran Rasul, "Social Networks and Technology Adoption in Northern Mozambique," *The Economic Journal*, 2006, *116* (514), 869–902.

Baranov, Victoria, Daniel Bennett, and Hans-Peter Kohler, "The indirect impact of antiretroviral therapy: Mortality risk, mental health, and HIV-negative labor supply," *Journal of Health Economics*, 2015, *44*, 195–211.

Bayer, Patrick, Stephen L. Ross, and Giorgio Topa, "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes," *Journal of Political Economy*, 2008, *116* (6), 1150–1196.

Beaman, Lori, Ariel BenYishay, Jeremy Magruder, and Ahmed Mushfiq Mobarak, "Can Network Theory-based Targeting Increase Technology Adoption?," Technical Report 2020.

Belloni, Alexandre and Victor Chernozhukov, "Least squares after model selection in high-dimensional sparse models," *Bernoulli*, 2013, *19* (2).

Bertrand, Marianne, Erzo F. P. Luttmer, and Sendhil Mullainathan, "Network Effects and Welfare Cultures," *Quarterly Journal of Economics*, 2000, *115* (3), 1019–1055.

Bharadwaj, Prashant, Mallesh M. Pai, and Agne Suziedelyte, "Mental health stigma," *Economics Letters*, 2017, *159*, 57–60.

Blom, Erik F., Kevin ten Haaf, Douglas A. Arenberg, and Harry J. de Koning, "Disparities in Receiving Guideline-Concordant Treatment for Lung Cancer in the United States," *Annals of the American Thoracic Society*, 2020, *17* (2), 186–194.

Blundell, Richard and James L. Powell, "Endogeneity in Semiparametric Binary Response Models," *Review of Economic Studies*, 2004, *71* (3), 655–679.

Brock, William A. and Steven N. Durlauf, "Interactions-Based Models," in "Handbook of Econometrics," Elsevier, 2001, pp. 3297–3380.

Budish, Eric, Benjamin N. Roin, and Heidi Williams, "Do Firms Underinvest in Long-Term Research? Evidence from Cancer Clinical Trials," *American Economic Review*, jul 2015, *105* (7), 2044–2085.

Burke, Mary A., Gary M. Fournier, and Kislaya Prasad, "The Diffusion of a Medical Innovation: Is Success in the Stars?," *Southern Economic Journal*, 2007, *73* (3), 588–603.

Cameron, A. Colin and Pravin K. Trivedi, *Microeconometrics*, Cambridge University Press, may 2005.

Carter-Harris, Lisa, "Lung cancer stigma as a barrier to medical help-seeking behavior: Practice implications," *Journal of the American Association of Nurse Practitioners*, 2015, *27* (5), 240–245.

Cassidy, Richard J., Xinyan Zhang, Jeffrey M. Switchenko, Pretesh R. Patel, Joseph W. Shelton, Sibo Tian, Ronica H. Nanda, Conor E. Steuer, Rathi N. Pillai, Taofeek K. Owonikoko, Suresh S. Ramalingam, Felix G. Fernandez, Seth D. Force, Theresa W. Gillespie, Walter J. Curran, and Kristin A. Higgins, "Health care disparities among octogenarians and nonagenarians with stage III lung cancer," *Cancer*, 2018, *124* (4), 775–784.

Chambers, Suzanne K, Jeffrey Dunn, Stefano Occhipinti, Suzanne Hughes, Peter Baade, Sue Sinclair, Joanne Aitken, Pip Youl, and Dianne L O'Connell, "A systematic review of the impact of stigma and nihilism on lung cancer outcomes," *BMC Cancer*, 2012, *12* (1).

Chan, David C, Matthew Gentzkow, and Chuan Yu, "Selection with Variation in Diagnostic Skill: Evidence from Radiologists," *The Quarterly Journal of Economics*, 2022, *137* (2), 729–783.

Chandra, Amitabh and Douglas O Staiger, "Identifying Sources of Inefficiency in Healthcare," *The Quarterly Journal of Economics*, jan 2020, *135* (2), 785–843.

Chapple, A, S Ziebland, and A McPherson, "Stigma, shame, and blame experienced by patients with lung cancer: qualitative study," *BMJ*, 2004, *328* (7454), 1470.

Chetty, Raj, John N Friedman, and Emmanuel Saez, "Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings," *American Economic Review*, 2013, *103* (7), 2683–2721.

Conley, Timothy G and Christopher R Udry, "Learning about a New Technology: Pineapple in Ghana," *American Economic Review*, 2010, *100* (1), 35–69.

Coscelli, Andrea, "The Importance of Doctors' and Patients' Preferences in the Prescription Decision," *The Journal of Industrial Economics*, 2000, *48* (3), 349–369.

__ **and Matthew Shum**, "An empirical model of learning and patient spillovers in new drug entry," *Journal of Econometrics*, 2004, *122* (2), 213–246.

Crawford, Gregory S. and Matthew Shum, "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica*, 2005, *73* (4), 1137–1173.

Cronin, Christopher J, Matthew P Forsstrom, and Nicholas W Papageorge, "What Good Are Treatment Effects without Treatment? Mental Health and the Reluctance to Use Talk Therapy," Working Paper 27711, National Bureau of Economic Research August 2020.

Currie, Janet M. and W. Bentley MacLeod, "Understanding Doctor Decision Making: The Case of Depression Treatment," *Econometrica*, 2020, *88* (3), 847–878.

Cutler, David, Jonathan S. Skinner, Ariel Dora Stern, and David Wennberg, "Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending," *American Economic Journal: Economic Policy*, 2019, *11* (1), 192–221.

Cutler, David M., Arnold M. Epstein, Richard G. Frank, Raymond Hartman, Charles Harles King, Joseph P. Newhouse, Meredith B. Rosenthal, and Elizabeth Richardson Vigor, "How Good a Deal Was the Tobacco Settlement?: Assessing Payments to Massachusetts," *Journal of Risk and Uncertainty*, 2000, *21* (2/3), 235–261.

Danesh, Noori Akhtar, Gileh Gol Akhtar Danseh, Hsien Yeang Seow, Saad Shakeel, and Christian Finley, "Trends in Survival Based on Treatment Modality in Non-Small Cell Lung Cancer Patients: A Population-Based Study," *Cancer Investigation*, 2019, *37* (8), 355–366.

Darden, Michael and Robert Kaestner, "Smoking, Selection, and Medical Care Expenditures," Technical Report 2022.

Davidoff, Amy J., Mei Tang, Brian Seal, and Martin J. Edelman, "Chemotherapy and Survival Benefit in Elderly Patients With Advanced Non–Small-Cell Lung Cancer," *Journal of Clinical Oncology*, 2010, *28* (13), 2191–2197.

de Castro-Carpeño, Javier, Rosario Perona, and Cristóbal Belda-Iniesta, "Treatment for ALK-mutated non-small-cell lung cancer: a new miracle in the research race," *Clinical and Translational Oncology*, 2011, *13* (11), 774–779.

de Oliveira, C., K. E. Bremner, R. Pataky, N. Gunraj, K. Chan, S. Peacock, and M. D. Krahn, "Understanding the costs of cancer care before and after diagnosis for the 21 most common cancers in Ontario: a population-based descriptive study," *CMAJ Open*, 2013, *1* (1), E1–E8.

Dimick, Justin B., Douglas O. Staiger, and John D. Birkmeyer, "Ranking Hospitals on Surgical Mortality: The Importance of Reliability Adjustment," *Health Services Research*, 2010, *45* (6p1), 1614–1629.

Dubois, Pierre and Margaret Kyle, "The Effects of Pharmaceutical Innovation on Cancer Mortality Rates," TSE Working Papers 16-688, Toulouse School of Economics (TSE) 2016.

__ **and Tuba Tunçel**, "Identifying the effects of scientific information and recommendations on physicians' prescribing behavior," *Journal of Health Economics*, 2021, *78*, 102461.

__ **, Olivier de Mouzon, Fiona Scott-Morton, and Paul Seabright**, "Market size and pharmaceutical innovation," *The RAND Journal of Economics*, 2015, *46* (4), 844–871.

Duflo, Esther and Emmanuel Saez, "Participation And Investment Decisions In A Retirement Plan: The Influence Of Colleagues' Choices," *Journal of Public Economics*, 07 2002, *85*, 121–148.

Dunn, J., G. Garvey, P. C. Valery, D. Ball, K. M. Fong, S. Vinod, D. L. O'Connell, and S. K. Chambers, "Barriers to lung cancer care: health professionals' perspectives," *Supportive Care in Cancer*, 2016, *25* (2), 497–504.

Elliot, Valerie L., Debra Morgan, Julie Kosteniuk, Amanda Froehlich Chow, and Melanie Bayly, "Health-related stigma of noncommunicable neurological disease in rural adult populations: A scoping review," *Health & Social Care in the Community*, 2018, *27* (4).

Ellis, Peter M., Emily .T. Vella, Yee .C. Ung, and the Lung Cancer Disease Site Group, "Systemic Treatment for Patients with Advanced Non-Small Cell Lung Cancer," Technical Report, Cancer Care Ontario 2016.

Finkelstein, Amy, Heidi Williams, and Matthew Gentzkow, "Sources of Geographic Variation in Health Care: Evidence From Patient Migration*," *The Quarterly Journal of Economics*, 07 2016, *131* (4), 1681–1726.

Giaccotto, Carmelo, Rexford E. Santerre, and John A. Vernon, "Drug Prices and Research and Development Investment Behavior in the Pharmaceutical Industry," *The Journal of Law & Economics*, 2005, *48* (1), 195–214.

Gong, Qing, "Physician Learning and Treatment Choices: Evidence from Brain Aneurysms," 2019. University of North Carolina at Chapel Hill.

Granovetter, Mark, "Threshold Models of Collective Behavior," *American Journal of Sociology*, 1978, *83* (6), 1420–1443.

Grossman, Daniel and Umair Khalil, "Neighborhood networks and program participation," *Journal of Health Economics*, 2020, *70*, 102257.

Grubesic, Tony H., "Zip codes and spatial analysis: Problems and prospects," *Socio-Economic Planning Sciences*, 2008, *42* (2), 129–149.

Guiteras, Raymond, James Levinsohn, and Ahmed Mushfiq Mobarak, "Demand Estimation with Strategic Complementarities: Sanitation in Bangladesh," Technical Report, CEPR Discussion Paper No. DP13498 2019.

Hamann, Heidi A., Jamie S. Ostroff, Emily G. Marks, David E. Gerber, Joan H. Schiller, and Simon J. Craddock Lee, "Stigma among patients with lung cancer: a patient-reported measurement model," *Psycho-Oncology*, 2013, *23* (1), 81–92.

Harrison, Glenn W., Morten I. Lau, and E. Elisabet Rutström, "Individual discount rates and smoking: Evidence from a field experiment in Denmark," *Journal of Health Economics*, 2010, *29* (5), 708–717.

Heckman, James J., "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 1978, *46* (4), 931–959.

Hellerstein, Judith K., "The Importance of the Physician in the Generic versus Trade-Name Prescription Decision," *The RAND Journal of Economics*, 1998, *29* (1), 108.

Howlader, Nadia, Gonçalo Forjaz, Meghan J. Mooradian, Rafael Meza, Chung Yin Kong, Kathleen A. Cronin, Angela B. Mariotto, Douglas R. Lowy, and Eric J. Feuer, "The Effect of Advances in Lung-Cancer Treatment on Population Mortality," *New England Journal of Medicine*, 2020, *383* (7), 640–649.

Ipsos MORI, "Perceptions of lung cancer in Canada: an IPSOS Mori report for the Global Lung Cancer Coalition," Technical Report, Global Lung Cancer Coalition 2010.

Kamath, Suneel D., Sheetal M. Kircher, and Al B. Benson, "A Comparison of Cancer Burden and Funding From the National Cancer Institute and Nonprofit Organizations Reveals Disparities in the Distribution of Funding Across Cancer Types," *Journal of the National Comprehensive Cancer Network*, 2019, *17* (3.5), HSR19–094.

Klabunde, Carrie N., Julie M. Legler, Joan L. Warren, Laura-Mae Baldwin, and Deborah Schrag, "A Refined Comorbidity Measurement Algorithm for Claims-Based Studies of Breast, Prostate, Colorectal, and Lung Cancer Patients," *Annals of Epidemiology*, 2007, *17* (8), 584–590.

Kleibergen, Frank and Richard Paap, "Generalized reduced rank tests using the singular value decomposition," *Journal of Econometrics*, jul 2006, *133* (1), 97–126.

Lakdawalla, Darius N., Eric C. Sun, Anupam B. Jena, Carolina M. Reyes, Dana P. Goldman, and Tomas J. Philipson, "An economic evaluation of the war on cancer," *Journal of Health Economics*, 2010, *29* (3), 333–346.

Leveälahti, Helena, Carol Tishelman, and Joakim Öhlén, "Framing the onset of lung cancer biographically: narratives of continuity and disruption," *Psycho-Oncology*, 2007, *16* (5), 466–473.

Lichtenberg, Frank R., "Importation and Innovation," *Economics of Innovation and New Technology*, 2007, *16* (6), 403–417.

__ , "Has medical innovation reduced cancer mortality?," NBER Working Papers 15880, National Bureau of Economic Research 2010.

__ , "The impact of pharmaceutical innovation on premature cancer mortality in Canada, 2000–2011," *International Journal of Health Economics and Management*, 2015, *15* (3), 339–359.

MacKinnon, James G., Morten Ørregaard Nielsen, and Matthew D. Webb, "Cluster-robust inference: A guide to empirical practice," *Journal of Econometrics*, 2022.

Manski, Charles F., "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, 1993, *60* (3), 531.

Matheson, Flora I., James R. Dunn, Katherine L. W. Smith, Rahim Moineddin, and Richard H. Glazier, "Elaboration de l'indice de marginalisation canadien: un nouvel outil d'étude des inégalités," *Canadian Journal of Public Health*, 2012, *103* (S2), S12–S16.

McFadden, Daniel L., "Modelling of Choice of Residential Location," in F. Snickers A. Karlquist, L. Lundquist and J. Weibull, eds., *Spatial Interaction Theory and Residential Location*, 1978.

Moffitt, Robert, "An Economic Model of Welfare Stigma," *American Economic Review*, 1983, *73* (5), 1023–35.

Munshi, Kaivan, "Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution," *Journal of Development Economics*, 2004, *73* (1), 185–213.

Pham, Jonathan, Matthew Conron, Gavin Wright, Paul Mitchell, David Ball, Jennifer Philip, Margaret Brand, John Zalcberg, and Rob G. Stirling, "Excess mortality and undertreatment in elderly lung cancer patients: treatment nihilism in the modern era?," *ERJ Open Research*, 2021, *7* (2), 00393–2020.

Riley, Kristen E., Michael R. Ulrich, Heidi A. Hamann, and Jamie S. Ostroff, "Decreasing Smoking but Increasing Stigma? Anti-tobacco Campaigns, Public Health, and Cancer Care," *AMA Journal of Ethics*, 2017, *19* (5), 475–485.

Sacher, Adrian G., Lisa W. Le, Anthea Lau, Craig C. Earle, and Natasha B. Leighl, "Real-world chemotherapy treatment patterns in metastatic non-small cell lung cancer: Are patients undertreated?," *Cancer*, 2015, *121* (15), 2562–2569.

Shi, Yuankai, Joseph Siu-Kie Au, Sumitra Thongprasert, Sankar Srinivasan, Chun-Ming Tsai, Mai Trong Khoa, Karin Heeroma, Yohji Itoh, Gerardo Cornelio, and Pan-Chyr Yang, "A Prospective, Molecular Epidemiology Study of EGFR Mutations in Asian Patients with Advanced Non–Small-Cell Lung Cancer of Adenocarcinoma Histology (PIONEER)," *Journal of Thoracic Oncology*, feb 2014, *9* (2), 154–162.

Shields, Misty Dawn, Julian A. Marin-Acevedo, and Bruna Pellini, "Immunotherapy for Advanced Non–Small Cell Lung Cancer: A Decade of Progress," *American Society of Clinical Oncology Educational Book*, jun 2021, (41), e105–e127.

Sorensen, Alan T., "Social learning and health plan choice," *The RAND Journal of Economics*, 2006, *37* (4), 929–945.

Stewart, Hannah, John Paul Jameson, and Lisa Curtin, "The relationship between stigma and self-reported willingness to use mental health services among rural and urban older adults.," *Psychological Services*, 2015, *12* (2), 141–148.

Stuber, Jennifer, Kathleen Maloy, Sara Rosenbaum, and Karen Jones, "Beyond Stigma: What Barriers Actually Affect the Decisions of Low-Income Families to Enroll in Medicaid?," *Health Policy and Management Issue Briefs*, 2000.

Suh, James H., Adrienne Johnson, Lee Albacker, Kai Wang, Juliann Chmielecki, Garrett Frampton, Laurie Gay, Julia A. Elvin, Jo-Anne Vergilio, Siraj Ali, Vincent A. Miller, Philip J. Stephens, and Jeffrey S. Ross, "Comprehensive Genomic Profiling Facilitates Implementation of the National Comprehensive Cancer Network Guidelines for Lung Cancer Biomarker Testing and Identifies Patients Who May Benefit From Enrollment in Mechanism-Driven Clinical Trials," *The Oncologist*, 2016, *21* (6), 684–691.

Thornton, Rebecca L, "The Demand for, and Impact of, Learning HIV Status," *American Economic Review*, 2008, *98* (5), 1829–1863.

Topa, Giorgio and Yves Zenou, "Neighborhood and Network Effects," in "Handbook of Regional and Urban Economics," Elsevier, 2015, pp. 561–624.

Tørring, Marie Louise, Morten Frydenberg, Rikke P. Hansen, Frede Olesen, and Peter Vedsted, "Evidence of increasing mortality with longer diagnostic intervals for five common cancers: A cohort study in primary care," *European Journal of Cancer*, 2013, *49* (9), 2187–2198.

Train, Kenneth E., *Discrete Choice Methods with Simulation* number 9780521766555. In 'Cambridge Books.', Cambridge University Press, 2009.

Walter, Julia, Amanda Tufman, Rolf Holle, and Larissa Schwarzkopf, ""Age matters"—German claims data indicate disparities in lung cancer care between elderly and young patients," *PLOS ONE*, 2019, *14* (6), e0217434.

Ward, Michael R. and David Dranove, "The Vertical Chain Of Research And Development In The Pharmaceutical Industry," *Economic Inquiry*, 1995, *33* (1), 70–87.

Wooldridge, Jefrey M., *Econometric Analysis of Cross Section and Panel Data*, The MIT Press, 2010.

Yu, Hang, "Social Stigma as a Barrier to HIV Testing: Evidence from a Randomized Controlled Trial in Mozambique," 2019. Working Paper.

Zeltzer, Dan, Liran Einav, Amy Finkelstein, Tzvi Shir, Salomon M. Stemmer, and Ran D. Balicer, "Why Is End-of-Life Spending So High? Evidence from Cancer Patients," *The Review of Economics and Statistics*, jul 2021, pp. 1–41.

Zhang, Hongming and Jibei Chen, "Current status and future directions of cancer immunotherapy," *Journal of Cancer*, 2018, *9* (10), 1773–1781.

Zhao, Hongwei and Lili Tian, "On Estimating Medical Cost and Incremental Cost-Effectiveness Ratios with Censored Data," *Biometrics*, 2001, *57* (4), 1002–1008.

Ziebarth, Nicolas R., "Lung cancer risk perception biases," *Preventive Medicine*, 2018, *110*, 16–23.

A Appendix A: Dataset construction: lung cancer
For Online Publication

This section details the construction of the main dataset used in our work.

A.1 Data overview

We link multiple datasets using the encrypted patient identifiers. Non-small cell lung cancer cases are identified through the Ontario Cancer Registry, which contains information on cancer site, histology, stage at diagnosis for all patients diagnosed with cancer in Ontario, as well as age, sex, and date of death. The Registered Persons Database contains demographic information and vital statistics on all residents of Ontario who are eligible for universal healthcare coverage in the province. The New Drug Funding Program is a publicly funded drug program in Ontario that covers the costs of novel and expensive intravenous cancer therapies. The database reports all publicly funded intravenous drug therapies administered in hospital and cancer clinics in Ontario. The Activity Level Reporting system contains information on all systemic and radiation therapy services and outpatient oncology clinic visits provided to persons diagnosed with cancer. The Ontario Health Insurance Plan database contains claims for all physician services, including primary care physicians, specialists and other physicians, diagnostic tests and laboratory services. The Ontario Drug Benefits database contains data on all prescription medications dispensed to persons eligible for publicly funded drug coverage, including those aged over 65 years. The Discharge Abstract Database holds data on diagnoses and procedures for all inpatient and outpatient hospital admissions. The National Ambulatory Care Reporting System reports services related to ambulatory care, including same-day surgeries/procedures and emergency department visits. The ICES Physician Database contains information on demographic information on physicians, including their age, sex, specialty, tenure, and location of practice (LHIN). Finally, the Smoking Cessation dataset is part of the Activity Level Reporting and collects information on self-reported smoking status of newly diagnosed patients with cancer after 2014.

Table A.1: Overview of Administrative ICES Databases

Dataset	Data and variables
Ontario Cancer Registry	Cancer site, diagnoses date, stage, tumor histology, collaborative staging (CS)
Registered Person Database	Demographic information, including postal code, income, employment, education, minority
New Drug Funding Program	Record of publicly funded intravenous drugs administered at the hospital (outpatient)
Activity Level Reporting (ALR)	Record of systemic therapy services (date and specific regimens) and radiation
Ontario Health Insurance Plan	Billing and reporting of all physician services, diagnostic tests and visits
Ontario Drug Benefit	Oral systemic therapy and all prescription drugs covered by the Ontario public system (over 65)
Discharge Abstract Database	Inpatient admissions to hospital cancer-related surgeries and other admissions
National Ambulatory Care Reporting	All emergency department visitis in Ontario, including administrative and clinical data
ICES Physician Database	Record of all active physicians, including physician demographics, tenure, specialty
ALR/Smoking cessation	Patient current smoking status

The table reports the list of databases and the main variables contained in the databases available through the Institute for Clinical Evaluative Sciences.

A.2 Treatment: the regimens

To define whether a patient is treated and which therapies are administered between the diagnosis and death or the last recorded follow-up, we combine information from mainly two datasets: the New Drug Funding Program (NDFP) reports the date, time and dose administered to each treated patient of any drug covered by this program, which includes the expensive intravenous chemotherapeutic agents used in outpatient settings; Cancer Activity Level Reporting - Systemic (ALR) details the date, time, and dose of all drugs administered to the patient as part of a regimen (a set of anti-cancer and supportive medications given during an active course of systemic chemotherapy that is named and defined in the Provincial Formulary Regimen List). First, by merging ALR and NDFP using the patient id, we define whether a patient is ever treated: if a patient identifier does not appear on either dataset or if the patient is only administered supportive drugs, we consider the patient as untreated.

Second, for treated patients, we supplement the information on drugs and regimens provided in ALR with the claims from NDFP: this step allows us to verify the accuracy of the regimen codes in ALR, which sometimes display inconsistencies. NDFP claims require standardized reporting with high levels of verification to be processed and reimbursed to hospitals, so they tend to be very accurate.

While some patient identifiers may appear in ALR and not in NDFP, if the regimens they receive are not covered by NDFP, the reverse should not happen. In few cases we have patients identifiers that appear in NDFP but not in

ALR, or patients for which the administration dates do not match precisely. We use the following heuristic process to recover the actual regimen administered: we consider all the regimens that contain the drug reported in NDFP and verify those that are appropriate for the patient, according to the official provincial guidelines, based on cancer histology, intent of systemic therapy, previous treatments, funding rules, and cycle frequency.

Oral targeted drugs are not reimbursed by NDFP, hence they only appear in ALR: we check the accuracy of the reporting using claims from the Ontario Drug Benefit (ODB) database. We remove patients participating to clinical trials only (539 patients), because for those patients we are unable precisely identify which drugs are administered.

Finally, we only keep the first line of treatment. As the ALR variable "line of therapy" is often missing, we reconstruct it following the medical literature: we check for gaps in treatment that are regimen-specific and range between 4 and 8 weeks, depending on whether the regimens administered before and after the gap are the same. For targeted therapy, we use the coverage duration defined by the Exceptional Access Program to identify when a switch happens in the line of therapy.

Table A.2: Overview of Regimens

Regimen Group	Regimen	Drugs	CCO/pCODR	Health Canada	FDA
Cisplatin-based	CISPDOCE	docetaxel; cisplatin	Mar 2003	Aug 2000	Dec 2002
	CISPETOP	etoposide; cisplatin	Apr 1994	Apr 1994	Nov 1983
	CISPGEMC	gemcitabine; cisplatin	Nov 2002	Aug 1999	May 1996
	CISPPEME	pemetrexed; cisplatin	Apr 2014	Feb 2008	Feb 2004
	CISPVINO	vinorelbine; cisplatin	Nov 1997	May 1994	Dec 1994
	CISPVNBL	vinblastine; cisplatin	Apr 1998	Apr 1998	Jan 1982
Carboplatin-based	CRBPDOCE	docetaxel; carboplatin	Mar 2003	Aug 2000	Dec 2002
	CRBPETOP	etoposide; carboplatin	Dec 1981	Dec 1981	Nov 1983
	CRBPGEMC	gemcitabine; carboplatin	Nov 2002	Aug 1999	May 1996
	CRBPPACL	paclitaxel; carboplatin	Mar 2003	Jul 1998	Dec 1992
	CRBPPEME	pemetrexed; carboplatin	Apr 2014	Feb 2008	Feb 2004
	CRBPPEME++PEMB	pemetrexed; carboplatin pembrolizumab;	Apr 2020	Mar 2019	Oct 2016
	CRBPVINO	vinorelbine; carboplatin	Nov 1997	May 1994	Dec 1994
	CRBVNBL	vinblastine; carboplatin	Apr 1998	Apr 1998	Jan 1982
Single agent	DOCE	docetaxel	Aug 2000	Aug 2000	Dec 2002
	GEMC	gemcitabine	Mar 1997	Mar 1997	May 1996
	PACL	paclitaxel	Dec 1993	Dec 1993	Dec 1992
	PEME	pemetrexed	Apr 2014	May 2010	Feb 2004
	VINO	vinorelbine	May 1994	May 1994	Dec 1994
Targeted	AFAT	afatinib	Aug 2014	Nov 2013	Jul 2013
	ALEC	alectinib	Apr 2019	Sep 2018	Dec 2017
	CRIZ	crizotinib	Dec 2015	Nov 2015	Aug 2011
	ERLO	erlotinib	Aug 2012	Jul 2012	Jul 2013
	GEFI	gefitinib	Sep 2011	Dec 2009	Jul 2015
	OSIM	osimertinib	Jan 2020	Jul 2018	Apr 2018
Immuno therapy	PEMB	pembrolizumab	Jan 2018	Jul 2017	Dec 2016

The table reports the list of regimens approved for first-line treatment of stage IV lung cancer classified as standard of care (chemotherapy: CISP, CRBP, SINGLE) and innovative (targeted and immunotherapy). Column 3 reports the drugs contained in each regimen. Column 4-6 report the dates of approval by the Ontario health authority CCO/pCODR (for the regimens), Health Canada, and the FDA (for the drugs).

A.3 Patient attributes

Table A.3 describes all the patient-related variables used in the study, including their definition and source. Table A.4 presents summary statistics for all patient-related variables.

Health-related attributes To control for the patient health status at the time of the diagnosis, which is likely to affect the treatment decision, we extract and construct a number of variables. First, following the medical literature, we use International Classification of Diseases-9 (ICD-9) diagnosis codes to retrieve all claims for each patient's episode of care from the Ontario Health Insurance Plan and calculate the Charlson comorbidity index, adapted for cancer: see Klabunde et al. (2007). The index uses information on the patient's medical history with a look-back period of 2 years to categorize comorbidities and pre-existing medical conditions known to increase the risk of death and, therefore, good predictors of the likelihood of treatment. Second, using hospital discharge data, we identify all cancer-related surgeries performed on the patient, if any: while only less than 3% of lung cancer patients in our sample undergo a surgery, the procedure places a strong physiologic demand on the cardiovascular and respiratory system, so we use it to further proxy for the health status of the patient, complementing the Charlson index. We also retrieve all emergency room visits, all prescription drug claims (aggregated at the ATC2 class level), and the use of preventive care prior to the diagnosis, including all recommended cancer screenings on the basis of the patient's age and sex (breast, cervical, and colorectal). Finally, we include controls for whether the patient required any home care service (including personal homemaking and nursing, among others), which capture the patient's ability to perform daily activities autonomously.

With the introduction of the provincial smoking cessation program in 2014, all newly diagnosed cancer patients are surveyed about their smoking habits and those who may benefit from tobacco cessation advice are referred to an appropriate and available service. For patients with any cancer diagnosis after 2014 we observe whether the patient self-reported as being a current smoker/tobacco user or indicated they had smoked or used tobacco within the past 6 months (see Appendix B for further details).

Cancer attributes Using the SEER ICD-O-3 morphology codes reported in the Ontario Cancer Registry, we classify each patient's non-small cell lung cancer into its histological type, including adenocarcinoma (the most common), squamous cell carcinoma (most frequent among smokers), and other less common histologies, such as large cell carcinoma and bronchiolo-alveolar carcinoma. Using topography codes, which identify the site of origin of the tumor, we control for the presence of multiple neoplasms in the lungs.

The Ontario Cancer Registry reports the collaborative staging (CS) variables, which summarize relevant information on the size and extent of the tumor in the body, based on the specific type of cancer. We select the appropriate variables for lung cancer and construct indices which measure the extent of cancer, if the cancer has spread to the lymph nodes and to distant parts of the body (metastases) and other characteristics that capture the heterogeneity in the disease within the metastatic stage. Unfortunately, since these variables are missing for 25% of the patients (mostly in the very early and very late years of the sample), we only use them in robustness analyses.

Socio-demographics The ICES datasets include some patient-level socio-demographic attributes. We observe their sex, the age in 5-year bins (<45, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, >85), the quintile of income based on the patient's census neighborhood and information on education attainment, employment and minority status.

Table A.3: Overview of patient-related characteristics

Variable	Description	Source
	Health-related attributes at diagnosis	
Charlson index	Charlson comorbidity index adjusted for cancer patients 2 years lookback	authors' calculations
Active smoker	current smoker or smoked in the past 6 months (post 2014)	ICES data
Patient referred	patient was ever referred to smoking cessation program	authors' calculations
Surgery	patient received cancer-related surgery	authors' calculations
Palliative radiotherapy	patient received palliative radiotherapy	authors' calculations
Preventive care	patient underwent required screening for sex-age group: PAP test, mammography, colorectal	authors' calculations
Home care	patient received any home care services before diagnosis	authors' calculations
Homemaking services	patient received personal homemaking services before diag.	authors' calculations
Nursing services	patient received nursing services before diagnosis	authors' calculations
Management services	patient received management services before diagnosis	authors' calculations
Other home care services	patient received other home care services before diagnosis	authors' calculations
ECOG PS	Eastern Cooperative Oncology Group performance status	ICES data
Frequency drug prescriptions (62 variables)	nb. prescription events by ATC2 class before diag.	authors' calculations
	Cancer-related attributes	
Tumor histology		
Adenocarcinoma	cancer morphology: adenocarcinoma	ICES data
Squamous cell carcinoma	cancer morphology: squamous cell carcinoma	ICES data
Large cell carcinoma	cancer morphology: large cell carcinoma	ICES data
Bronchiolo-alveolar carcinoma	cancer morphology: bronchiolo-alveolar carcinoma	ICES data
Multiple tumors in the site	patient has multiple cancers in the lung	authors' calculations
Collaborative staging (CS)		
Tumor extension	localized, extended or very extended tumor	ICES data
Lymphnodes attacked	lymphnodes attacked by tumor	ICES data
Metastases	presence of metastases, regional or distant	ICES data
Specific metastases site	contralateral lung involved, liver, brain, bones	ICES data
Presence of nodules	presence of separate tumor nodules in ipsilateral lung	ICES data
	Socio-demographic characteristics	

Sex	biological sex (male-female)	ICES data
Age	age group (10 5-year bins)	ICES data
Ontario rurality index	Ontario rurality index of the nearest census neighborhood	ICES data
Distance to hospital (km)	distance to the regional cancer center used by the patient	authors' calculations
Income quintile	income quintile based on nearest census neighborhood	ICES data
Education tercile	education tercile based on nearest census neighborhood	ICES data
Employment	employment (above/below median) based on nearest census neighborhood	ICES data
Minority	minority status based on nearest census neighborhood	ICES data
	Health outcomes	
Survival	days between diagnosis and death	authors' calculations
Other		
Diagnosis to consultation	lag in days between diagnosis and consultation	authors' calculations

The table reports an overview of patient-related variables, their definition and source.

Table A.4: Summary statistics of patient-related characteristics: lung cancer

	Cohort	Treatment type			p-value		
		untreated	SOC	innovative			
		(0)	(1)	(2)	(0)=(1)	(0)=(2)	(1)=(2)
Tot. patients	15,761	8,611	5,545	1,605			
		0.55	0.35	0.10			
	Health-related attributes at diagnosis						
Charlson index	1.04	1.18	0.89	0.76	0.00	0.00	0.00
Active smoker (0/1)	0.32	0.36	0.37	0.16	0.40	0.00	0.00
Patient referred to smoking cessation	0.12	0.13	0.14	0.05	0.09	0.00	0.00
Surgery (0/1)	0.03	0.02	0.04	0.03	0.00	0.01	0.02
Palliative radiotherapy (0/1)	0.68	0.64	0.73	0.68	0.00	0.00	0.00
Preventive care	0.48	0.43	0.5	0.6	0.00	0.00	0.00
Home care	0.26	0.34	0.17	0.19	0.00	0.00	0.11
Homemaking services	0.04	0.06	0.01	0.01	0.00	0.00	0.09
Nursing services	0.06	0.08	0.03	0.03	0.00	0.00	0.57
Management services	0.12	0.18	0.06	0.07	0.00	0.00	0.03
Other home care services	0.06	0.09	0.02	0.03	0.00	0.00	0.32
Frequency of drug prescription before diagnosis:							
stomalogical preparation drugs (A01)	0.01	0.02	0.01	0.02	0.13	0.69	0.19
acid related disorders (A02)	2.89	4.07	1.45	1.52	0.00	0.00	0.68
gastrointestinal disorders (A03)	0.22	0.29	0.11	0.21	0.00	0.41	0.22
antiemetics and antinauseants (A04)	0.02	0.02	0.01	0.01	0.27	0.26	0.83
bile and liver theraphy (A05)	0	0.01	0	0	0.11	0.01	0.27
for constipation (A06)	0.96	1.42	0.35	0.6	0.00	0.00	0.05
antidiarrehals (A07)	0.11	0.13	0.1	0.03	0.27	0.00	0.00
digestives (A09)	0.01	0.01	0.01	0.01	0.29	0.41	0.93
diabetes (A10)	2.57	3.28	1.61	2.09	0.00	0.00	0.11
vitamins (A11)	0.11	0.17	0.03	0.06	0.00	0.02	0.36
antithrombotic agents (B01)	1.58	2.26	0.77	0.76	0.00	0.00	0.94
antianemic preparations (B03)	0.24	0.35	0.1	0.16	0.00	0.01	0.38
drugs for cardiac therapy (C01)	0.65	0.98	0.27	0.2	0.00	0.00	0.33
antihypertensives (C02)	0.17	0.26	0.07	0.07	0.00	0.00	0.93
diuretics (C03)	2.36	3.32	1.2	1.22	0.00	0.00	0.91
peripheral vasodilators (C04)	0.08	0.12	0.04	0.01	0.04	0.00	0.08
beta blocking agents (C07)	2.36	3.31	1.24	1.2	0.00	0.00	0.81

calcium channel blockers (C08)	2.45	3.33	1.26	1.86	0.00	0.00	0.00
renin-angiotensin system drugs (C09)	4.25	5.43	2.71	3.19	0.00	0.00	0.04
lipid modifying agents (C10)	5.07	6.65	3.11	3.39	0.00	0.00	0.33
antifungals (D01)	0.12	0.16	0.07	0.11	0.00	0.02	0.03
antipsoriatics (D05)	0.03	0.04	0.03	0.02	0.51	0.33	0.69
antibiotics and chemotheapeutics (D06)	0.14	0.17	0.1	0.11	0.00	0.00	0.81
corticosteroids (D07)	0.39	0.47	0.28	0.37	0.00	0.02	0.02
anti-acne preparations (D10)	0	0	0	0.01	0.72	0.48	0.78
other dermatological preparations (D11)	0.01	0.01	0	0.01	0.04	0.97	0.29
gynecological antiinfectives (G01)	0.01	0.02	0.01	0	0.04	0.00	0.00
sex hormones (G03)	0.15	0.18	0.12	0.1	0.03	0.01	0.31
urologicals (G04)	1.34	1.86	0.65	0.92	0.00	0.00	0.07
pituitary hormones (H01)	0	0.01	0	0	0.14	0.14	.
corticosteroids (H02)	0.4	0.53	0.26	0.19	0.00	0.00	0.04
drugs for tyroid theraphy (H03)	1.13	1.6	0.49	0.77	0.00	0.00	0.02
antibacterials for systemic use (J01)	1.4	1.62	1.14	1.09	0.00	0.00	0.45
antimycotics for systemic use (J02)	0.01	0.01	0	0.01	0.22	0.86	0.56
antimycobacterials (J04)	0.01	0.02	0.01	0	0.27	0.09	0.36
antivirals (J05)	0.04	0.04	0.03	0.02	0.70	0.00	0.03
vaccines (J07)	0.09	0.09	0.08	0.14	0.03	0.00	0.00
antineoplastic agents (L01)	0.13	0.16	0.09	0.07	0.01	0.00	0.41
drugs for endocrine therapy (L02)	0.06	0.09	0.03	0	0.01	0.00	0.00
immunostimulants (L03)	0	0	0	0	0.18	0.18	.
immunosupressants (L04)	0.1	0.14	0.07	0.04	0.03	0.00	0.26
antiinflammatory products (M01)	0.8	0.99	0.6	0.48	0.00	0.00	0.03
muscle relaxants (M03)	0.11	0.15	0.05	0.1	0.01	0.57	0.39
antigout preparation (M04)	0.46	0.71	0.16	0.18	0.00	0.00	0.55
drugs for treatment bone diseases (M05)	1	1.32	0.54	0.83	0.00	0.00	0.01
anesthetics (N01)	0	0	0	0	0.14	0.01	0.08
analgesics (N02)	2.33	3.2	1.3	1.2	0.00	0.00	0.45
antiepileptics (N03)	1.07	1.56	0.42	0.62	0.00	0.00	0.11
anti-Parkinson drugs (N04)	0.28	0.45	0.06	0.13	0.00	0.00	0.28
psycholeptics (N05)	2.11	3.2	0.84	0.61	0.00	0.00	0.13
psychoanaleptics (N06)	2.7	3.94	1.16	1.37	0.00	0.00	0.37
other nervous system drugs (N07)	0.09	0.13	0.04	0.01	0.23	0.08	0.09
antiprotozoals (P01)	0.14	0.19	0.08	0.1	0.00	0.06	0.75
ectoparasiticides (P03)	0	0	0	0	0.03	0.01	0.08

nasal preparations (R01)	0.22	0.25	0.16	0.26	0.00	0.78	0.01
obstructive airway diseases drugs (R03)	3.03	3.94	2.08	1.5	0.00	0.00	0.00
cough and cold preparations (R05)	0.19	0.23	0.14	0.18	0.00	0.05	0.06
antihistamines for systemic use (R06)	0	0	0	0	0.04	0.02	0.16
ophthalmologicals (S01)	1.09	1.35	0.7	1.03	0.00	0.03	0.02
otologicals (S02)	0.01	0.02	0.01	0.02	0.00	0.94	0.09
ophthalmological and otological (S03)	0.01	0.01	0.01	0.01	0.41	0.12	0.42
various (V04)	0	0	0	0	0.10	0.10	.
other drugs	0	0.01	0	0	0.05	0.02	0.37
			Cancer-related attributes				
Tumor histology:							
Adenocarcinoma (0/1)	0.75	0.7	0.77	0.91	0.00	0.00	0.00
Squamous cell carcinoma (0/1)	0.2	0.25	0.18	0.04	0.00	0.00	0.00
Large cell carcinoma (0/1)	0.02	0.02	0.02	0.01	0.59	0.00	0.00
Bronchiolo-alveolar carcinoma (0/1)	0	0	0	0	0.77	0.81	0.95
Multiple tumors in the site (0/1)	0.01	0.01	0.02	0.03	0.00	0.00	0.01
Collaborative staging (0/1):							
Localized tumor	0.42	0.4	0.43	0.47	0.00	0.00	0.08
Extended tumor	0.33	0.34	0.31	0.34	0.01	0.68	0.06
Very extended tumor	0.25	0.26	0.26	0.19	0.51	0.00	0.00
Lymphnodes not attacked	0.2	0.21	0.18	0.22	0.00	0.50	0.01
Regional lymphnodes attacked	0.51	0.51	0.52	0.48	0.47	0.06	0.03
Lymphnodes attacked	0.2	0.19	0.22	0.19	0.00	0.83	0.02
No distant metastases	0	0	0.01	0.01	0.20	0.17	0.45
Distant metastases	0.71	0.71	0.7	0.72	0.34	0.56	0.28
Pleural effusion	0.46	0.47	0.43	0.51	0.00	0.03	0.00
Pericardial effusion	0.21	0.23	0.19	0.26	0.00	0.04	0.00
Contralateral lung involved	0.2	0.19	0.2	0.2	0.31	0.49	0.91
Metastases in the lungs	0.25	0.24	0.25	0.3	0.08	0.00	0.00
Metastases in the bones	0.38	0.38	0.37	0.5	0.76	0.00	0.00
Metastases in the liver	0.17	0.17	0.17	0.17	0.26	0.92	0.48
Metastases in the brain	0.25	0.26	0.2	0.31	0.00	0.01	0.00
No separate tumor nodules ipsilateral lung	0.54	0.55	0.53	0.49	0.02	0.00	0.04
Separate tumor nodules ipsilateral lung	0.3	0.28	0.31	0.36	0.01	0.00	0.00
			Socio-demographic attributes				

Male	0.52	0.54	0.53	0.41	0.20	0.00	0.00
Age < 45	0.01	0.01	0.02	0.04	0.00	0.00	0.00
Age 45-49	0.02	0.01	0.03	0.03	0.00	0.00	0.66
Age 50-54	0.06	0.04	0.08	0.08	0.00	0.00	0.49
Age 55-59	0.10	0.08	0.13	0.12	0.00	0.00	0.22
Age 60-64	0.14	0.11	0.19	0.14	0.00	0.00	0.00
Age 65-69	0.17	0.16	0.20	0.15	0.00	0.64	0.00
Age 70-74	0.17	0.17	0.17	0.16	0.99	0.42	0.44
Age 75-79	0.15	0.18	0.12	0.14	0.00	0.00	0.02
Age 80-84	0.10	0.15	0.04	0.08	0.00	0.00	0.00
Age 85+	0.06	0.10	0.01	0.04	0.00	0.00	0.00
Ontario rurality index	12.04	12.00	12.89	9.26	0.01	0.00	0.00
Distance to hospital (km)	31.26	30.98	33.61	24.65	0.00	0.00	0.00
Income quintile	2.81	2.71	2.92	2.97	0.00	0.00	0.27
Education tercile	1.91	1.87	1.92	2.04	0.00	0.00	0.00
Employment (0/1)	0.48	0.46	0.49	0.52	0.00	0.00	0.05
Minority (0/1)	0.50	0.49	0.48	0.61	0.06	0.00	0.00
			Health outcomes				
1-year survival prob.	0.29	0.12	0.45	0.68	0.00	0.00	0.00
Survival days	336.73	187.58	484.48	626.54	0.00	0.00	0.00

The table reports the summary statistics of all the variables in our sample related to lung cancer patients. The first column includes health-related attributes, tumor attributes, health care utilization measures, and a set of characteristics related to the three-digit zip code of the patient's residence for the whole sample. Columns 2-4 compare those characteristics between (i) untreated patients; (ii) patients treated with the standard of care (SOC or chemotherapy); and (iii) patients treated with innovative therapies. Columns 5-7 report the results of a Welch $t-$test across the subsamples.

A.4 Hospitals and physicians

The matching algorithm consists of the following steps. We match the selected cohort with OHIP, which presents information on the physicians billing their services along with the diagnosis code, the fee code and the service date. First, we select the oncologist that, in every year, tends to have the highest number of visits with the patients. Second, we extract the oncologist(s) billing assessment and consultation services to OHIP related to a patient using the fee codes related to visit, assessment and consultation. Third, we extract the oncologist supervising the chemotherapy using the treatment/service date in ALR and OHIP and the associated LHIN (hospital). Fourth, we keep a window of 30 days around the diagnosis date and select the most frequent physician according to the following specialties (in hierarchical order): medical oncology, radiation oncology, respirology, thoracic surgery, internal medicine, general practice. Fifth, we extract the surgeon performing a surgery to the patient, if applicable. Sixth, we extract the residual physician associated to fee codes related to assessment, consultation, and palliative care. We match the extracted physicians (from one to six) to the patient and select the main treating physician following the presented order. In case a medical oncologist cannot be matched to the patient, we considered the next matched physician in the following hierarchical order of specialty: radiation oncologist, respirologist, surgeon, and general practitioner. We verify that a patient tends to be matched to one main medical oncologist. In the uncommon case of multiple medical oncologists matched with one patient, we select the most frequent one. As a double check on the effectiveness of the matching algorithm, we extract the patients for which a test of the presence of mutations is prescribed, and the associated referring oncologist, when present in the data. The referring oncologist matches with the medical oncologist selected by our algorithm over 80% of the times.

When extracting the information on the referring physician, we keep a window of 5 days around the diagnosis date and select the physician according to the specialty and the diagnosis code, focusing on non-screening physicians. If multiple specialties and diagnosis codes are a possible match, the main referring physician is selected according to his/her specialty in the following order: respirologist, internist or emergency physician, surgeon, and family physician.

Table A.5: List of regional cancer programs and cancer centers

LHIN/Regional Cancer Program	Regional Cancer Center	Host Hospital
Erie St. Clair	Windsor	Windsor Regional Hospital
South West	London	London Health Sciences Centre
Waterloo Wellington	Grand River	Grand River Hospital
Hamilton Niagara	Juravinski	Hamilton Health Sciences
Mississauga Halton Central West	Carlo Fidani	Trillium Health Partners-Credit Valley Site
Toronto Central	Odette	Sunnybrook Health Sciences Centre
Toronto Central	Princess Margaret	University Health Network
Central	Stronach	Southlake Regional Health Centre
Central East	R.S. McLaughlin Durham	Lakeridge Health
South East	Southeastern Ontario	Kingston General Hospital
Champlain	Ottawa Hospital	The Ottawa Hospital
North Simcoe Muskoka	Simcoe Muskoka	Royal Victoria Hospital
North East	Northeast	Health Sciences North/Horizon Santé-Nord
North West	Northwest	Thunder Bay Regional Health Sciences Centre

The table reports the list of 14 regional cancer programs/regions delivering cancer care in Ontario and the associated Regional Cancer Centers. LHIN = Local Health Integrated Network. Mississauga Halton and Central West are two separate LHINs hosting one regional cancer center. The LHIN Toronto Central hosts two regional cancer centers

Table A.6: Overview of physician-related variables

Variable	Description	Source
Sex	biological sex of the doctor (male-female)	ICES data
Age	age in years	ICES data
Specialty	medical oncology, radiation oncologist general practitioner, other	ICES data
Lung cancer patients/year	number of distinct lung cancer patients/year	authors' calculations
Nb. consultations in a year	number of consultations in year	authors' calculations
Date of specialty	physician's career lenght at diagnosis date	ICES data

The table reports an overview of physician-related variables.

A.5 Neighborhood attributes

Table A.7 describes all the neighborhood-related variables used in the study, including their definition and source. To complement the limited socio-economic information on the patients provided by ICES data, we collect rich neighborhood-level statistics for the three-digit zip code (FSA, Forward Sortation Area) of residence of the patient. We use publicly available census data from the 2006, 2011 and 2016 waves, as well as survey responses to the National Health Survey and the Canadian Community Health Survey: jointly, these sources provide information on income level and sources in the neighborhood, education level, employment, ethnicity and immigration, as well as self-reported smoking and drinking habits, food insecurity, incidence of mood disorders, and sense of belonging

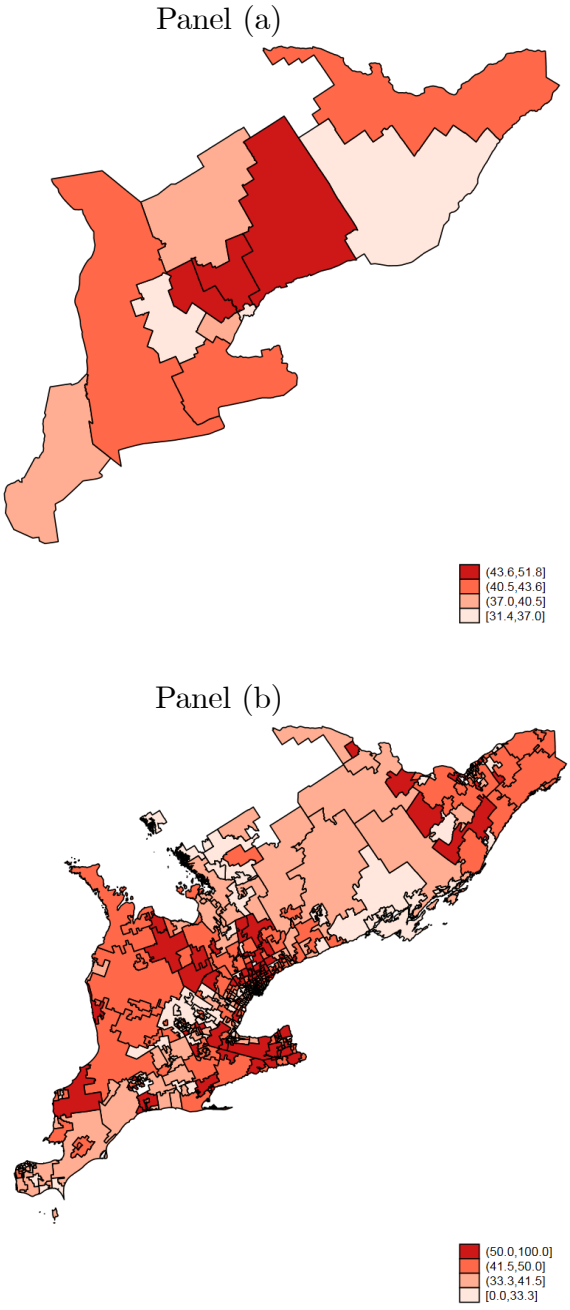
to the local community. We construct measures of area, size, population and density for each FSA. Finally, we collect information on pollution, measured by the particulate matter emissions/releases (<2.5 micrometers, in metric tonnes), derived from the National Pollutant Release Inventory data managed by the Government of Canada: the data reports emissions by company and facility and we aggregate it at the FSA-year level.

Table A.7: Overview of FSA-related variables

Variable	Description	Source
Population	population of the FSA	StatCan HH Survey 2011 & 2016
Population density	population density (inhabitants per km^2)	StatCan HH Survey 2011 & 2016 and authors' calculations
Median income	median household income in the FSA	StatCan HH Survey 2011 & 2016
% income from welfare payments	share of income from welfare payments	StatCan HH Survey 2011 & 2016
Quintiles of marginalization index:	S*hare of the population in the FSA that:*	
instability	experiences high rates of family or housing instability	Public Health Ontario 2016
deprivation	is unable to access and attain basic material needs	Public Health Ontario 2016
dependency	does not have income from employment	Public Health Ontario 2016
ethnic concentration	recent immigrant and/or belonging to a visible minority group (non-Caucasian or non-white in colour)	Public Health Ontario 2016
Share of population:	*Share of the population in the FSA:*	
with no education	with no certificate, diploma or degree	StatCan HH Survey 2011 & 2016
with high school degree	with completed high school degree	StatCan HH Survey 2011 & 2016
with postsecondary degree	with completed postsecondary degree	StatCan HH Survey 2011 & 2016
Unemployment rate	that is unemployed	StatCan HH Survey 2011 & 2016
Participation rate in labor force	that is active in labor force	StatCan HH Survey 2011 & 2016
Average weeks worked	average weeks worked in previous year	StatCan HH Survey 2011 & 2016
Share of population:	*Share of the population in the FSA:*	
aboriginal population	who is of aboriginal identity	StatCan HH Survey 2011 & 2016
immigrant population	that is immigrant	StatCan HH Survey 2011 & 2016
Asian immigrants	that is of Asian origin	StatCan HH Survey 2011 & 2016
South-Eastern Asian immigrants	that is of South-Eastern Asian origin	StatCan HH Survey 2011 & 2016
Smoking rate	that smokes	StatCan HH Survey 2011 & 2016
Share of population:	*Share of the population in the FSA:*	
heavy smokers	that smokes daily	StatCan Health Survey 2007-2019
heavy drinkers	that drinks at least three times per week	StatCan Health Survey 2007-2019
with mood disorder	that has a mood disorder	StatCan Health Survey 2007-2019
food insecure	that is food insecure	StatCan Health Survey 2007-2019
with sense of belonging	that does not feel sense of belonging	StatCan Health Survey 2007-2019
Pollution (pm2.5)	Emissions of particulate matter <2.5 micrometers in metric tonnes	National Pollutant Release Inventory

The table reports an overview of neighborhood-related variables at FSA level (3-digit Canadian zip code), their definition and source. HH = household

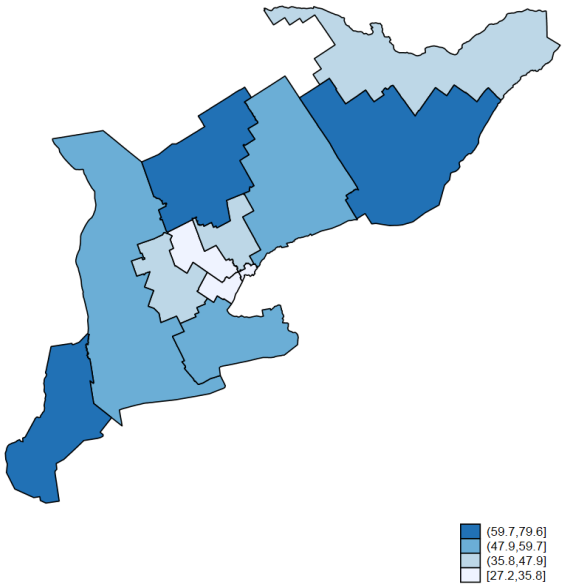
Figure A.1: Treatment Rate of Lung Cancer - LHIN (Panel a) and FSA (Panel b)



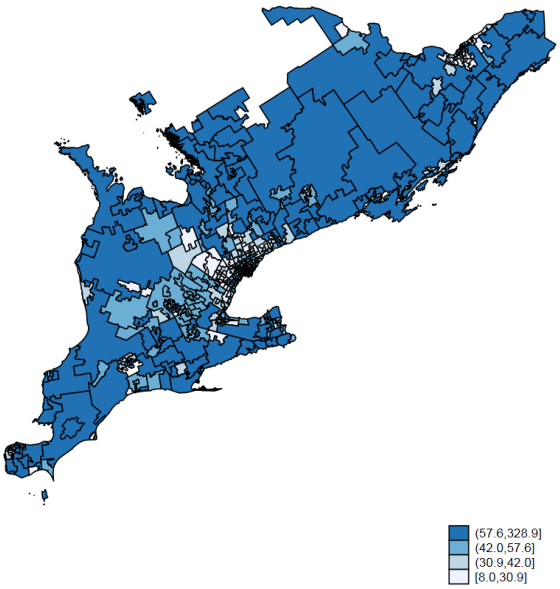
Treatment Rate of Lung Cancer Patients at Local Health Integration Network Area (Panel a) and FSA (three-digit ZIP code) (Panel b). Northern Ontario is excluded. Source: authors' calculations based on ICES data.

Figure A.2: Incidence of Lung Cancer - LHIN

Panel (a)



Panel (b)



Number of lung Cancer Patients per 100,000 inhabitants at Local Health Integration Network Area (Panel a) and FSA (three-digit ZIP code) (Panel b). Northern Ontario is excluded. Source: authors' calculations based on ICES data.

B Appendix B: Other cancers

For Online Publication

B.1 Cohort selection

To select the cohort of colorectal, prostate, and (female) breast cancer, we follow the same procedure used for non-small cell lung cancer.

Selection of the initial cohort is based on site-specific SEER ICD-O-3 topography codes, which identify the site of origin of each neoplasm for each patient. We exclude patients with concurrent tumors in different sites and keep only those with a first diagnosis at the advanced stage of the disease. For all cancer types, we consider only patients initially diagnosed at the metastatic stage (stage IV).

Treatment for advanced colorectal, prostate, and breast cancer is based on systemic therapy. While the protocols are cancer-specific, they all include the administration of chemotherapy, immunotherapy or targeted/hormonal therapy, alone or in combination with radiation, especially for patients with bone metastases. Hence, we consider a patient to be treated if they receive any antineoplastic drug (standard chemotherapy, immunotherapy or targeted/hormonal therapy). This definition allows us to precisely identify treated patients with colorectal and breast cancer, for whom we find treatment rates that are high and in line with reported statistics form other sources. For metastatic prostate cancer, we also include radiotherapy-only as a form of treatment, following the American Society of Clinical Oncology guidelines, that recommend radiotherapy for certain patients with limited metastatic disease.

We extract and create the same variables we use for lung cancer patients described above for patients with colorectal, breast and prostate cancer. Tables B.8, B.9, and B.10 report summary statistics for patient-related attributs of colorectal, breast, and prostate cancer patients. Tables B.11, B.12, B.13 report neighborhood-related attributes for each cancer type.

Table B.8: Summary statistics of patient-related characteristics: colorectal

	(1) Cohort	(2) Treatment type untreated (0)	(3) SOC (1)	(4) p-value (0)=(1)
Tot. patients	8382	2274	6108	
	Health-related attributes at diagnosis			
Charlson index	0.69	1.03	0.57	0.00
Active smoker (0/1)	0.18	0.19	0.18	0.62
Patient referred to smoking cessation (OHIP)	0.05	0.04	0.06	0.00
Surgery (0/1)	0.59	0.47	0.64	0.00
Preventive care	0.44	0.33	0.47	0.00
Home care	0.21	0.38	0.14	0.00
Homemaking services	0.03	0.10	0.01	0.00
Nursing services	0.04	0.09	0.03	0.00
Management services	0.09	0.21	0.05	0.00
Other home care services	0.04	0.11	0.02	0.00
Frequency of drug prescription before diagnosis:				
stomalogical preparation drugs (A01)	0.01	0.02	0.00	0.02
acid related disorders (A02)	1.79	3.96	0.98	0.00
functional gastrointestinal disorders (A03)	0.13	0.30	0.07	0.00
antiemetics and antinauseants (A04)	0.00	0.01	0.00	0.20
drugs for bile and liver theraphy (A05)	0.00	0.00	0.00	0.85
drugs for constipation (A06)	0.77	1.64	0.44	0.00
antidiarrehals (A07)	0.08	0.18	0.04	0.01
digestives (A09)	0.00	0.01	0.00	0.11
drugs for diabetes (A10)	2.02	3.94	1.30	0.00
vitamins (A11)	0.05	0.15	0.02	0.03
antithrombotic agents (B01)	0.94	2.11	0.50	0.00
antianemic preparations (B03)	0.17	0.44	0.08	0.00
drugs for cardiac therapy (C01)	0.42	0.98	0.21	0.00
antihypertensives (C02)	0.12	0.25	0.06	0.01
diuretics (C03)	1.84	4.11	1.00	0.00
peripheral vasodilators (C04)	0.05	0.16	0.00	0.04
beta blocking agents (C07)	1.73	3.76	0.97	0.00

calcium channel blockers (C08)	1.69	3.61	0.98	0.00
renin-angiotensin system drugs (C09)	3.19	6.08	2.11	0.00
lipid modifying agents (C10)	3.21	6.10	2.14	0.00
antifungals (D01)	0.10	0.17	0.07	0.00
antipsoriatics (D05)	0.01	0.02	0.01	0.46
antibiotics and chemotheapeutics (D06)	0.10	0.19	0.06	0.00
corticosteroids (D07)	0.29	0.49	0.21	0.00
anti-acne preparations (D10)	0.01	0.01	0.01	0.85
other dermatological preparations (D11)	0.00	0.01	0.00	0.12
gynecological antiinfectives (G01)	0.01	0.03	0.00	0.05
sex hormones (G03)	0.11	0.20	0.08	0.01
urologicals (G04)	0.82	1.81	0.46	0.00
pituitary hormones (H01)	0.01	0.04	0.00	0.10
corticosteroids (H02)	0.22	0.44	0.13	0.00
drugs for tyroid theraphy (H03)	0.79	1.62	0.48	0.00
antibacterials for systemic use (J01)	0.82	1.45	0.59	0.00
antimycotics for systemic use (J02)	0.00	0.00	0.00	0.99
antimycobacterials (J04)	0.00	0.00	0.00	0.32
antivirals (J05)	0.02	0.04	0.01	0.00
vaccines (J07)	0.06	0.08	0.05	0.00
antineoplastic agents (L01)	0.07	0.16	0.04	0.03
drugs for endocrine therapy (L02)	0.05	0.09	0.04	0.07
immunostimulants (L03)	0.00	0.00	0.00	.
immunosupressants (L04)	0.06	0.12	0.04	0.05
antiinflammatory products (M01)	0.47	0.69	0.38	0.00
muscle relaxants (M03)	0.03	0.07	0.02	0.28
antigout preparation (M04)	0.27	0.57	0.16	0.00
drugs for treatment bone diseases (M05)	0.70	1.64	0.36	0.00
anesthetics (N01)	0.00	0.00	0.00	0.42
analgesics (N02)	1.12	2.33	0.68	0.00
antiepileptics (N03)	0.63	1.42	0.34	0.00
anti-Parkinson drugs (N04)	0.28	0.80	0.09	0.01
psycholeptics (N05)	1.19	2.78	0.60	0.00
psychoanaleptics (N06)	1.71	4.13	0.81	0.00
other nervous system drugs (N07)	0.02	0.05	0.01	0.37
antiprotozoals (P01)	0.08	0.12	0.07	0.02
ectoparasiticides (P03)	0.00	0.00	0.00	0.54

nasal preparations (R01)	0.12	0.20	0.09	0.00
obstructive airway diseases drugs (R03)	1.17	2.23	0.77	0.00
cough and cold preparations (R05)	0.09	0.16	0.06	0.00
antihistamines for systemic use (R06)	0.00	0.00	0.00	0.20
ophthalmologicals (S01)	0.89	1.60	0.62	0.00
otologicals (S02)	0.01	0.02	0.01	0.03
ophthalmological and otological prep before diag (S03)	0.01	0.02	0.00	0.00
various (V04)	0.00	0.00	0.00	.
other drugs	0.00	0.01	0.00	0.14
		Cancer-related attributes		
Cancer histology				
Adenocarcinoma (0/1)	0.91	0.91	0.91	0.70
Mucinous adenocarcinoma (0/1)	0.06	0.06	0.07	0.57
Signet-ring cell carcinoma (0/1)	0.02	0.02	0.02	0.28
Multiple tumors in the site (0/1)	0.07	0.03	0.08	0.00
		Socio-demographics attributes		
Male	0.57	0.54	0.59	0.00
Age group:				
<45	0.06	0.02	0.07	0.00
45-49	0.05	0.01	0.06	0.00
50-54	0.09	0.03	0.10	0.00
55-59	0.11	0.05	0.13	0.00
60-64	0.13	0.07	0.15	0.00
65-69	0.14	0.10	0.15	0.00
70-74	0.14	0.15	0.14	0.96
75-79	0.12	0.17	0.10	0.00
80-84	0.10	0.20	0.06	0.00
85+	0.07	0.20	0.02	0.00
Ontario rurality index	12.02	10.18	12.71	0.00
Distance to hospital (km)	30.35	26.58	31.75	0.00
Income quintile	2.94	2.78	3.00	0.00
Education tercile	1.95	1.91	1.97	0.01
Employment (0/1)	1.50	1.49	1.51	0.05
Minority (0/1)	1.49	1.52	1.48	0.00

		Health outcomes		
1-year survival prob.	0.59	0.19	0.73	0.00
Survival days	672	254	827	0.00

The table reports the summary statistics of all the variables in our sample related to colorectal cancer patients. The first column includes health-related attributes, tumor attributes, health care utilization measures, and a set of characteristics related to the three-digit zip code of the patient's residence for the whole sample. Columns 2 and 3 compare those characteristics between (i) untreated patients; and (ii) patients treated with the standard of care (SOC or chemotherapy). Column 4 reports the results of a Welch $t-$test across the two subsamples.

Table B.9: Summary statistics of patient-related characteristics: female breast

	(1) Cohort	(2) Treatment type untreated (0)	(3) SOC (1)	(4) p-value (0)=(1)
Tot. patients	3773	838	2935	
		Health-related attributes at diagnosis		
Charlson index	0.55	0.77	0.48	0.00
Active smoker (0/1)	0.14	0.10	0.14	0.21
Patient referred to smoking cessation (OHIP)	0.04	0.03	0.04	0.12
Surgery (0/1)	0.24	0.14	0.27	0.00
Preventive care	0.71	0.46	0.76	0.00
Home care	0.24	0.34	0.21	0.00
Homemaking services	0.04	0.09	0.03	0.00
Nursing services	0.05	0.07	0.04	0.00
Management services	0.12	0.19	0.10	0.00
Other home care services	0.05	0.09	0.04	0.00
Frequency of drug prescription before diagnosis:				
stomalogical preparation drugs (A01)	0.01	0.01	0.01	0.75
acid related disorders (A02)	2.23	4.50	1.58	0.00
functional gastrointestinal disorders (A03)	0.25	0.60	0.15	0.05
antiemetics and antinauseants (A04)	0.01	0.00	0.01	0.50
drugs for bile and liver theraphy (A05)	0.00	0.00	0.00	.
drugs for constipation (A06)	0.95	2.46	0.52	0.00
antidiarrehals (A07)	0.08	0.07	0.08	0.72
digestives (A09)	0.00	0.00	0.00	.
drugs for diabetes (A10)	2.12	4.31	1.49	0.00
vitamins (A11)	0.10	0.25	0.06	0.29
antithrombotic agents (B01)	1.37	2.89	0.94	0.00
antianemic preparations (B03)	0.21	0.47	0.13	0.11
drugs for cardiac therapy (C01)	0.54	1.29	0.33	0.01
antihypertensives (C02)	0.16	0.40	0.09	0.10
diuretics (C03)	2.66	5.95	1.73	0.00
peripheral vasodilators (C04)	0.01	0.00	0.01	0.13
beta blocking agents (C07)	2.25	4.42	1.63	0.00

calcium channel blockers (C08)	1.83	3.54	1.34	0.00
renin-angiotensin system drugs (C09)	3.51	6.95	2.53	0.00
lipid modifying agents (C10)	3.45	5.90	2.75	0.00
antifungals (D01)	0.12	0.23	0.09	0.01
antipsoriatics (D05)	0.02	0.04	0.02	0.33
antibiotics and chemotheapeutics (D06)	0.13	0.31	0.07	0.00
corticosteroids (D07)	0.23	0.51	0.15	0.00
anti-acne preparations (D10)	0.00	0.00	0.00	0.45
other dermatological preparations (D11)	0.01	0.02	0.00	0.21
gynecological antiinfectives (G01)	0.01	0.00	0.01	0.30
sex hormones (G03)	0.13	0.16	0.12	0.58
urologicals (G04)	0.45	1.06	0.28	0.02
pituitary hormones (H01)	0.00	0.00	0.00	0.38
corticosteroids (H02)	0.34	0.36	0.33	0.87
drugs for tyroid theraphy (H03)	1.57	3.22	1.10	0.00
antibacterials for systemic use (J01)	0.73	1.13	0.61	0.00
antimycotics for systemic use (J02)	0.00	0.00	0.00	0.67
antimycobacterials (J04)	0.00	0.00	0.00	0.32
antivirals (J05)	0.02	0.04	0.02	0.03
vaccines (J07)	0.05	0.04	0.06	0.02
antineoplastic agents (L01)	0.11	0.23	0.08	0.30
drugs for endocrine therapy (L02)	0.02	0.02	0.01	0.83
immunostimulants (L03)	0.00	0.00	0.00	.
immunosupressants (L04)	0.04	0.05	0.03	0.46
antiinflammatory products (M01)	0.61	1.22	0.44	0.00
muscle relaxants (M03)	0.10	0.09	0.11	0.83
antigout preparation (M04)	0.29	0.61	0.20	0.07
drugs for treatment bone diseases (M05)	1.17	2.34	0.83	0.00
anesthetics (N01)	0.00	0.00	0.00	0.69
analgesics (N02)	1.77	3.66	1.23	0.00
antiepileptics (N03)	0.75	1.62	0.51	0.02
anti-Parkinson drugs (N04)	0.27	0.94	0.08	0.04
psycholeptics (N05)	1.75	4.03	1.10	0.00
psychoanaleptics (N06)	2.70	5.82	1.81	0.00
other nervous system drugs (N07)	0.00	0.00	0.00	0.11
antiprotozoals (P01)	0.09	0.23	0.05	0.22
ectoparasiticides (P03)	0.00	0.00	0.00	0.10

nasal preparations (R01)	0.10	0.15	0.08	0.07
obstructive airway diseases drugs (R03)	0.94	1.86	0.68	0.00
cough and cold preparations (R05)	0.07	0.13	0.05	0.02
antihistamines for systemic use (R06)	0.00	0.00	0.00	0.57
ophthalmologicals (S01)	0.76	1.62	0.51	0.00
otologicals (S02)	0.01	0.01	0.01	0.71
ophthalmological and otological prep (S03)	0.00	0.01	0.00	0.37
various (V04)	0.00	0.00	0.00	.
other drugs	0.00	0.00	0.00	0.40
		Cancer-related attributes		
Cancer histology				
Infiltrating duct carcinoma (0/1)	0.69	0.60	0.72	0.00
Lobular carcinoma (0/1)	0.09	0.08	0.09	0.45
Multiple tumors in the site (0/1)	0.02	0.02	0.03	0.07
		Socio-demographics attributes		
Age group:				
<45	0.10	0.03	0.12	0.00
45-49	0.09	0.04	0.10	0.00
50-54	0.11	0.05	0.12	0.00
55-59	0.11	0.08	0.12	0.00
60-64	0.12	0.09	0.12	0.01
65-69	0.11	0.10	0.12	0.07
70-74	0.10	0.13	0.09	0.01
75-79	0.10	0.15	0.09	0.00
80-84	0.09	0.17	0.07	0.00
85+	0.08	0.18	0.05	0.00
Ontario rurality index	10.02	9.34	10.21	0.17
Distance to hospital (km)	28.45	28.67	28.39	0.88
Income quintile	2.89	2.77	2.92	0.01
Education tercile	1.99	1.93	2.01	0.03
Employment (0/1)	0.49	0.46	0.49	0.06
Minority (0/1)	0.54	0.57	0.53	0.05
		Health outcomes		
1-year survival prob.	0.72	0.34	0.83	0.00

Survival days	938	420	1086	0.00

The table reports the summary statistics of all the variables in our sample related to female breast cancer patients. The first column includes health-related attributes, tumor attributes, health care utilization measures, and a set of characteristics related to the three-digit zip code of the patient's residence for the whole sample. Columns 2 and 3 compare those characteristics between (i) untreated patients; and (ii) treated patients. Column 4 reports the results of a Welch $t-$test across the two subsamples.

Table B.10: Summary statistics of patient-related characteristics: prostate

	(1) Cohort	(2) Treatment type untreated (0)	(3) SOC (1)	(4) p-value (0)=(1)
Tot. patients	6127	1362	4765	
		Health-related attributes at diagnosis		
Charlson index	0.89	1.17	0.81	0.00
Active smoker (0/1)	0.14	0.12	0.14	0.42
Patient referred to smoking cessation (OHIP)	0.06	0.04	0.06	0.02
Surgery (0/1)	0.28	0.33	0.26	0.00
Preventive care	0.32	0.29	0.33	0.13
Home care	0.31	0.41	0.27	0.00
Homemaking services	0.04	0.08	0.03	0.00
Nursing services	0.10	0.15	0.09	0.00
Management services	0.17	0.25	0.15	0.00
Other home care services	0.07	0.12	0.05	0.00
Frequency of drug prescription before diagnosis:				
stomalogical preparation drugs (A01)	0.01	0.01	0.00	0.06
acid related disorders (A02)	2.95	4.95	2.38	0.00
functional gastrointestinal disorders (A03)	0.14	0.26	0.11	0.11
antiemetics and antinauseants (A04)	0.01	0.03	0.01	0.31
drugs for bile and liver theraphy (A05)	0.00	0.00	0.00	0.10
drugs for constipation (A06)	1.05	1.74	0.86	0.00
antidiarrehals (A07)	0.09	0.11	0.08	0.39
digestives (A09)	0.02	0.08	0.00	0.33
drugs for diabetes (A10)	3.28	4.70	2.88	0.00
vitamins (A11)	0.09	0.19	0.07	0.04
antithrombotic agents (B01)	1.96	3.21	1.60	0.00
antianemic preparations (B03)	0.27	0.53	0.19	0.03
drugs for cardiac therapy (C01)	0.94	1.49	0.78	0.01
antihypertensives (C02)	0.14	0.18	0.12	0.19
diuretics (C03)	2.70	4.96	2.05	0.00
peripheral vasodilators (C04)	0.01	0.01	0.01	0.53
beta blocking agents (C07)	2.70	4.08	2.31	0.00

calcium channel blockers (C08)	2.13	2.86	1.92	0.00
renin-angiotensin system drugs (C09)	4.75	6.99	4.11	0.00
lipid modifying agents (C10)	5.25	7.72	4.55	0.00
antifungals (D01)	0.17	0.24	0.15	0.02
antipsoriatics (D05)	0.03	0.07	0.02	0.20
antibiotics and chemotheapeutics (D06)	0.15	0.21	0.14	0.03
corticosteroids (D07)	0.43	0.58	0.39	0.00
anti-acne preparations (D10)	0.00	0.00	0.00	0.53
other dermatological preparations (D11)	0.01	0.04	0.00	0.30
gynecological antiinfectives (G01)	0.00	0.00	0.00	0.22
sex hormones (G03)	0.03	0.05	0.03	0.49
urologicals (G04)	4.04	6.19	3.43	0.00
pituitary hormones (H01)	0.01	0.01	0.01	0.98
corticosteroids (H02)	0.30	0.54	0.23	0.02
drugs for tyroid theraphy (H03)	0.72	1.39	0.52	0.00
antibacterials for systemic use (J01)	1.68	2.13	1.55	0.00
antimycotics for systemic use (J02)	0.00	0.00	0.00	0.81
antimycobacterials (J04)	0.00	0.01	0.00	0.32
antivirals (J05)	0.04	0.04	0.04	0.95
vaccines (J07)	0.11	0.12	0.11	0.17
antineoplastic agents (L01)	0.07	0.12	0.05	0.26
drugs for endocrine therapy (L02)	0.26	0.27	0.26	0.88
immunostimulants (L03)	0.00	0.00	0.00	.
immunosupressants (L04)	0.05	0.05	0.05	0.87
antiinflammatory products (M01)	0.79	0.88	0.76	0.35
muscle relaxants (M03)	0.10	0.05	0.11	0.21
antigout preparation (M04)	0.60	0.85	0.53	0.07
drugs for treatment bone diseases (M05)	0.35	0.56	0.29	0.05
anesthetics (N01)	0.00	0.00	0.00	0.71
analgesics (N02)	1.75	2.39	1.57	0.00
antiepileptics (N03)	0.85	1.40	0.70	0.02
anti-Parkinson drugs (N04)	0.27	0.48	0.20	0.05
psycholeptics (N05)	1.42	2.61	1.08	0.01
psychoanaleptics (N06)	2.28	4.51	1.64	0.00
other nervous system drugs (N07)	0.12	0.27	0.08	0.31
antiprotozoals (P01)	0.03	0.05	0.03	0.24
ectoparasiticides (P03)	0.01	0.01	0.00	0.36

nasal preparations (R01)	0.22	0.26	0.20	0.20
obstructive airway diseases drugs (R03)	1.20	1.96	0.98	0.00
cough and cold preparations (R05)	0.10	0.16	0.09	0.04
antihistamines for systemic use (R06)	0.00	0.00	0.00	0.32
ophthalmologicals (S01)	1.24	1.55	1.15	0.02
otologicals (S02)	0.01	0.02	0.01	0.09
ophthalmological and otological (S03)	0.01	0.01	0.01	0.49
various (V04)	0.00	0.00	0.00	.
other drugs	0.01	0.01	0.01	0.53
		Cancer-related attributes		
Cancer histology				
Adenocarcinoma (0/1)	0.87	0.82	0.88	0.00
Small cell carcinoma (0/1)	0.01	0.00	0.01	0.24
Intraductal carcinoma (0/1)	0.01	0.02	0.01	0.02
		Socio-demographics attributes		
Age group:				
<45	0.00	0.00	0.00	0.83
45-49	0.01	0.00	0.01	0.00
50-54	0.04	0.02	0.05	0.00
55-59	0.08	0.04	0.09	0.00
60-64	0.13	0.08	0.14	0.00
65-69	0.16	0.12	0.17	0.00
70-74	0.16	0.14	0.17	0.06
75-79	0.14	0.14	0.14	0.69
80-84	0.14	0.21	0.12	0.00
85+	0.13	0.23	0.10	0.00
Ontario rurality index	12.90	10.80	13.50	0.00
Distance to hospital (km)	33.60	29.60	34.74	0.00
Income quintile	3.05	2.94	3.08	0.00
Education tercile	2.00	2.00	2.00	0.85
Employment (0/1)	0.50	0.50	0.50	0.92
Minority (0/1)	0.47	0.51	0.46	0.00
		Health outcomes		
1-year survival prob.	0.83	0.70	0.87	0.00

Survival days	1109	916	1164	0.00

The table reports the summary statistics of all the variables in our sample related to prostate cancer patients. The first column includes health-related attributes, tumor attributes, health care utilization measures, and a set of characteristics related to the three-digit zip code of the patient's residence for the whole sample. Columns 2 and 3 compare those characteristics between (i) untreated patients; and (ii) treated patients . Column 4 reports the results of a Welch $t-$test across the two subsamples.

Table B.11: Summary statistics of neighborhood-related characteristics: colorectal

	(1) Cohort	(2) Treatment type untreated (0)	(3) SOC (1)	(4) $p-$value (0)=(1)
Urban (%)	83	87	81	0.00
Population density	2165	2433	2066	0.00
Median income	30983	30652	31107	0.00
% income from welfare payments	22	22	22	0.26
Unemployment rate	8.10	8.31	8.02	0.00
Pollution (pm2.5)	26.56	24.44	27.34	0.26
Quintiles of marginalization index:				
instability	3.00	3.14	2.95	0.00
deprivation	3.20	3.31	3.15	0.00
ethnic concentration	3.00	3.14	2.95	0.00
Share of population:				
with high school degree	0.27	0.27	0.27	0.40
South-Eastern Asian immigrants	0.05	0.05	0.05	0.05
heavy smokers	0.14	0.14	0.14	0.01
heavy drinkers	0.36	0.36	0.37	0.01
Tot. patients	8382	2274	6108	

The table reports the summary statistics of the variables in our sample related to neighborhood characteristics of colorectal cancer patients. Columns 2 and 3 compare those characteristics between (i) untreated patients; and (ii) treated patients. Column 4 reports the results of a Welch $t-$test across the two subsamples.

Table B.12: Summary statistics of neighborhood-related characteristics: female breast

	(1)	(2)	(3)	(4)
	Cohort	Treatment type		$p-$value
		untreated	SOC	
		(0)	(1)	(0)=(1)
Urban (%)	86	88	86	0.19
Population density	2511	2567	2495	0.58
Median income	30978	30296	31173	0.00
% income from welfare payments	21.45	21.91	21.32	0.03
Unemployment rate	8.22	8.42	8.17	0.00
Pollution (pm2.5)	22.21	19.31	23.04	0.24
Quintiles of marginalization index:				
instability	3.01	3.03	3.01	0.62
deprivation	3.22	3.36	3.18	0.00
ethnic concentration	3.17	3.28	3.14	0.02
Share of population:				
with high school degree	0.27	0.27	0.27	0.16
South-Eastern Asian immigrants	0.05	0.06	0.05	0.01
heavy smokers	0.13	0.13	0.13	0.86
heavy drinkers	0.36	0.36	0.36	0.34
Tot. patients	3773	838	2935	

The table reports the summary statistics of the variables in our sample related to neighborhood characteristics of female breast cancer patients. Columns 2 and 3 compare those characteristics between (i) untreated patients; and (ii) treated patients. Column 4 reports the results of a Welch $t-$test across the two subsamples.

Table B.13: Summary statistics of neighborhood-related characteristics: prostate

	(1) Cohort	(2) Treatment type untreated (0)	(3) SOC (1)	(4) $p-$value (0)=(1)
Urban (%)	82	85	81	0.00
Population density	2219	2556	2123	0.00
Median income	31200	30961	31268	0.09
% income from welfare payments	22	22	22	0.41
Unemployment rate	8.12	8.25	8.09	0.01
Pollution (pm 2.5)	31.3	27.79	32.3	0.20
Quintiles of marginalization index:				
instability	3	3.11	2.97	0.00
deprivation	3.18	3.23	3.16	0.10
ethnic concentration	2.97	3.13	2.92	0.00
Share of population:				
with high school degree	0.27	0.27	0.27	0.39
South-Eastern Asian immigrants	0.05	0.05	0.04	0.07
heavy smokers	0.14	0.13	0.14	0.04
heavy drinkers	0.36	0.36	0.36	0.66
Tot. patients	6127	1362	4765	

The table reports the summary statistics of the variables in our sample related to neighborhood characteristics of prostate cancer patients. Columns 2 and 3 compare those characteristics between (i) untreated patients; and (ii) treated patients. Column 4 reports the results of a Welch $t-$test across the two subsamples.

B.2 Smoking status

With the introduction of the provincial smoking cessation program in 2014, all newly diagnosed cancer patients are surveyed about their smoking habits and those who may benefit from tobacco cessation advice are referred to an appropriate and available service. For patients with any cancer diagnosis after 2014 we observe whether the patient self-reported as being a current smoker/tobacco user or indicated they had smoked or used tobacco within the past 6 months. Table B.14 reports the summary statistics for smokers versus non-smokers affected by one of the top four cancers under investigation (lung, colorectal, breast and prostate); Table B.15 reports the summary statistics for smokers affected by lung cancer versus smokers affected by the other three cancers (colorectal, breast and prostate); Table B.16 reports the summary statistics for smokers affected by lung cancer versus non-smokers affected by lung cancer.

The smoking status is recorded for a subset of patients (around 45% in 2015 and 70% in later years). For the top 4 cancers, we observe 9,596 patients with non-missing records, out of the 17,201 diagnoses for 2014-2018. For lung cancer, 2,907 out of 4,269 patients with non-missing records are active smokers, roughly a third: the figure is twice as large as that of the other three cancers (15 percent of smokers). The average smoking rate in the general Ontario population was 18 percent over the same period.

Smokers affected by one the top four cancers look similar along several dimensions. They are significantly younger than non-smokers and, as a consequence, healthier beyond cancer. Smokers also tend to use health care to a lesser extent, as captured by lower take-up of preventive care, home care and fewer drug prescriptions: this may be a combination of younger age and attitude towards lower health care use more generally. Lower use of medical care is also consistent with worse socio-economic status: smokers are poorer and less educated than non-smokers and come from neighborhoods that are more rural, further away from hospitals, with lower median income and employment rates, marginalized along all dimensions and more polluted. Smokers with stage IV lung cancer are significantly less likely to be treated than non-smoker lung cancer patients and their raw survival rates are worse, while treatment rates for smokers with colorectal, breast, and prostate cancer are comparable to those for non-smokers. When treated, smokers with lung cancer are more likely to receive standard of care rather than innovative therapy, consistent with the more common squamous histology of their tumor. The zip codes where they reside display higher incidence of lung cancer and lower treatment rates as well.

Table B.14: Summary statistics of lung, colorectal, breast and prostate cancer patients with available smoking status

	Cohort	Non smokers (0) (0)	Smokers (1) (1)	$p-$value (0)=(1)
Treatment (%)	73	76	65	0
	Health-related attributes at diagnosis			
Charlson index	0.81	0.82	0.75	0.01
Surgery (0-1)	0.22	0.24	0.16	0.00
Preventive care (%)	0.48	0.51	0.41	0.00
Home care use (%)	0.26	0.27	0.24	0.00
Multiple tumors	0.04	0.04	0.03	0.12
	Socio-demographic attributes			
Age:				
<45	0.04	0.04	0.03	0.01
45-49	0.03	0.03	0.03	0.67
50-54	0.07	0.06	0.09	0.00
55-59	0.11	0.09	0.17	0.00
60-64	0.14	0.12	0.22	0.00
65-69	0.16	0.16	0.18	0.01
70-74	0.16	0.17	0.16	0.31
75-79	0.13	0.14	0.08	0.00
80-84	0.09	0.11	0.04	0.00
85+	0.07	0.08	0.01	0.00
Distance to hospital (km)	30.58	29.5	34.28	0.00
Income quintile	2.92	3.00	2.66	0.00
Education terciles	1.96	2.00	1.82	0.00
Employment (0/1)	0.71	0.73	0.63	0.00
Minority (0/1)	0.7	0.72	0.59	0.00
	Health outcomes			
1-year survival prob.	0.62	0.65	0.49	0.00
Survival days	588	619	482	0.00
	Neighborhood characteristics			
Urban	0.82	0.83	0.81	0.02
Median income	31,105	31,253	30,600	0.00
% income from welfare payments	21.96	21.61	23.13	0.00
Unemployment rate	8.12	8.08	8.24	0.00
Pollution (pm25)	27.79	26.1	33.56	0.02
Quintiles of marginalization index:				
instability	2.93	2.88	3.08	0.00
deprivation	3.16	3.11	3.33	0.00
ethnic concentration	2.96	3.03	2.72	0.00
Share of polulation:				
with high school degree	0.27	0.27	0.28	0.00
South-Eastern Asian immigrants	0.05	0.05	0.03	0.00
heavy smokers	0.12	0.12	0.13	0.00
heavy drinkers	0.35	0.35	0.36	0.00
Tot. patients	9,596	7,422	2,174	

The table reports the summary statistics of selected variables in our sample related to lung, colorectal, female breast, and prostate cancer patients for whom we have information about their smoking status (smokers or non-smokers). Columns 2 and 3 compare non-smokers to smokers. Column 4 reports the results of a Welch $t-$test across the two subsamples.

Table B.15: Summary statistics of current smokers affected by lung, colorectal, breast and prostate cancer patients

	Cohort	Lung (0)	Other cancers (1)	
		(0)	(1)	(0)=(1)
Treatment (%)	73	76	65	0
	Health-related attributes at diagnosis			
Charlson index	0.75	0.89	0.53	0.00
Surgery (0-1)	0.16	0.02	0.39	0.00
Preventive care (%)	0.41	0.43	0.37	0.01
Home care use (%)	0.24	0.24	0.23	0.63
Multiple tumors	0.03	0.02	0.06	0.00
	Socio-demographic attributes			
Age:				
<45	0.03	0.01	0.06	0.00
45-49	0.03	0.02	0.05	0.00
50-54	0.09	0.07	0.12	0.00
55-59	0.17	0.16	0.18	0.15
60-64	0.22	0.22	0.21	0.41
65-69	0.18	0.2	0.16	0.01
70-74	0.16	0.17	0.14	0.03
75-79	0.08	0.09	0.05	0.00
80-84	0.04	0.05	0.02	0.00
85+	0.01	0.01	0.01	0.55
Distance to hospital (km)	34.28	33.32	35.89	0.30
Income quintile	2.66	2.65	2.68	0.69
Education terciles	1.82	1.82	1.81	0.76
Employment (0/1)	0.63	0.46	0.89	0.00
Minority (0/1)	0.59	0.43	0.88	0.00
	Health outcomes			
1-year survival prob.	0.49	0.32	0.78	0.00
Survival days	482.24	333.39	731.91	0.00
	Neighborhood characteristics			
Urban	0.81	0.81	0.8	0.81
Median income	30,600	30,559	30,668	0.64
% income from welfare payments	23.13	23.23	22.95	0.37
Unemployment rate	8.24	8.21	8.29	0.44
Pollution (pm25)	33.56	34.04	32.77	0.83
Quintiles of marginalization index:				
instability	3.08	3.08	3.08	0.91
deprivation	3.33	3.33	3.32	0.81
ethnic concentration	2.72	2.70	2.76	0.32
Share of polulation:				
with high school degree	0.28	0.28	0.28	0.71
South-Eastern Asian immigrants	0.03	0.03	0.03	0.49
heavy smokers	0.13	0.13	0.13	0.66
heavy drinkers	0.36	0.36	0.36	0.63
Tot. patients	2174	1362	812	

The table reports the summary statistics of selected variables in our sample related to lung, colorectal, female breast, and prostate cancer patients who are all current smokers. Columns 2 and 3 compare lung cancer smokers to colorectal, female breast, and prostate cancer smokers. Column 4 reports the results of a Welch $t-$test across the two subsamples.

Table B.16: Summary statistics of lung cancer patients with available smoking status

	Cohort	Non smokers (0)	Smokers (1)	
		(0)	(1)	(0)=(1)
Treatment (%)	73	76	65	0
	Health-related attributes at diagnosis			
Charlson index	0.97	1.02	0.89	0.0
Surgery (0-1)	0.02	0.02	0.02	0.9
Preventive care (%)	0.51	0.55	0.43	0.0
Home care use (%)	0.27	0.28	0.24	0.0
Adenocarcinoma	0.79	0.81	0.75	0.0
Squamous cell carcinoma	0.17	0.16	0.22	0.0
Multiple tumors	0.02	0.02	0.02	0.3
	Socio-demographic attributes			
Male (%)	0.52	0.51	0.53	0.2
Age:				
<45	0.02	0.02	0.01	0.0
45-49	0.02	0.02	0.02	0.4
50-54	0.05	0.05	0.07	0.0
55-59	0.11	0.09	0.16	0.0
60-64	0.15	0.12	0.22	0.0
65-69	0.17	0.16	0.2	0.0
70-74	0.18	0.19	0.17	0.2
75-79	0.15	0.17	0.09	0.0
80-84	0.09	0.11	0.05	0.0
85+	0.05	0.07	0.01	0.0
Distance to hospital (km)	30.43	29.08	33.32	0.0
Income quintile	2.83	2.91	2.65	0.0
Education terciles	1.91	1.95	1.82	0.0
Employment (0/1)	0.49	0.5	0.46	0.0
Minority (0/1)	0.49	0.51	0.43	0.0
	Health outcomes			
1-year survival prob.	0.39	0.43	0.32	0.0
Survival days	383.35	406.76	333.39	0.0
	Neighborhood characteristics			
Urban	0.83	0.84	0.81	0.0
Median income	30,802	30,916	30,559	0.0
% income from welfare payments	22.19	21.7	23.23	0.0
Unemployment rate	8.19	8.17	8.21	0.5
Pollution (pm25)	27.52	24.47	34.04	0.0
Quintiles of marginalization index:				
instability	2.94	2.87	3.08	0.0
deprivation	3.21	3.16	3.33	0.0
ethnic concentration	2.97	3.10	2.70	0.0
Share of polulation:				
with high school degree	0.27	0.27	0.28	0.0
South-Eastern Asian immigrants	0.05	0.06	0.03	0.0
heavy smokers	0.12	0.12	0.13	0.0
heavy drinkers	0.35	0.35	0.36	0.0
Tot. patients	4,269	2,907	1,362	

The table reports the summary statistics of selected variables in our sample related to lung cancer patients that are currently smokers and lung cancer patients that are not current smokers. Columns 2 and 3 compare non-smokers to smokers (both affected by lung cancer). Column 4 reports the results of a Welch $t-$test across the two subsamples.

B.3 Treatment toxicities: a comparison

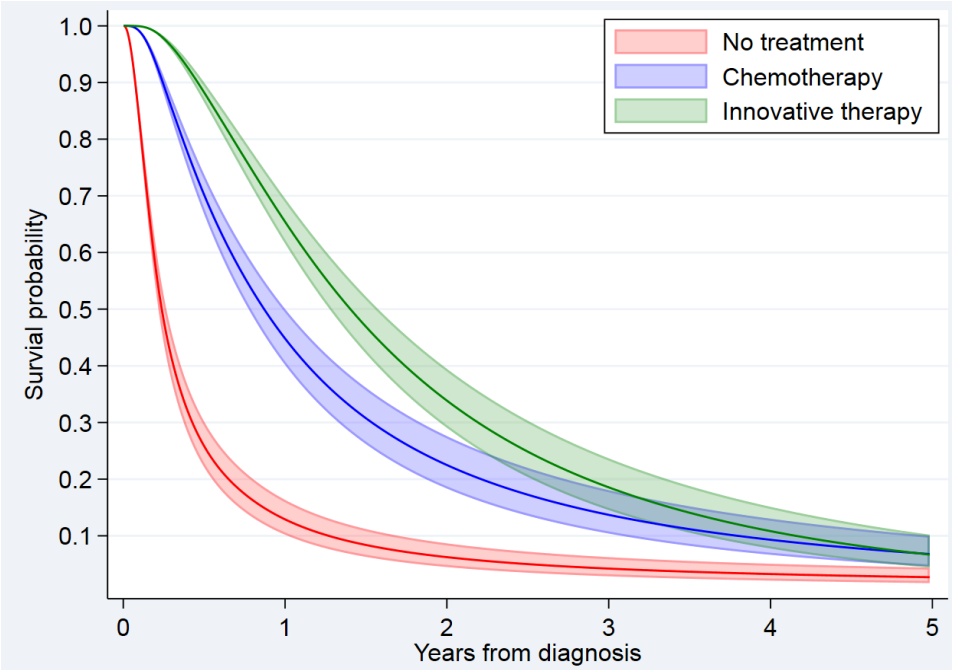
Table B.17: A qualitative comparison of treatment toxicities: lung vs. colorectal cancer

	Lung cancer				Colorectal cancer	
	chemotherapy		innovative therapy		chemotherapy	
Side effects	frequent	severe	frequent	severe	frequent	severe
Myelosuppression	✓	✓			✓	✓
Neurotoxicity	✓	✓	✓		✓	✓
Nausea, vomiting	✓✓		✓		✓✓	✓
Metabolic disorders	✓		✓		✓✓	
Fatigue	✓✓		✓		✓✓	
Rash, alopecia	✓✓	✓	✓	✓	✓	

The table presents a qualitative comparison between lung cancer and colorectal cancer in terms of treatment toxicity.

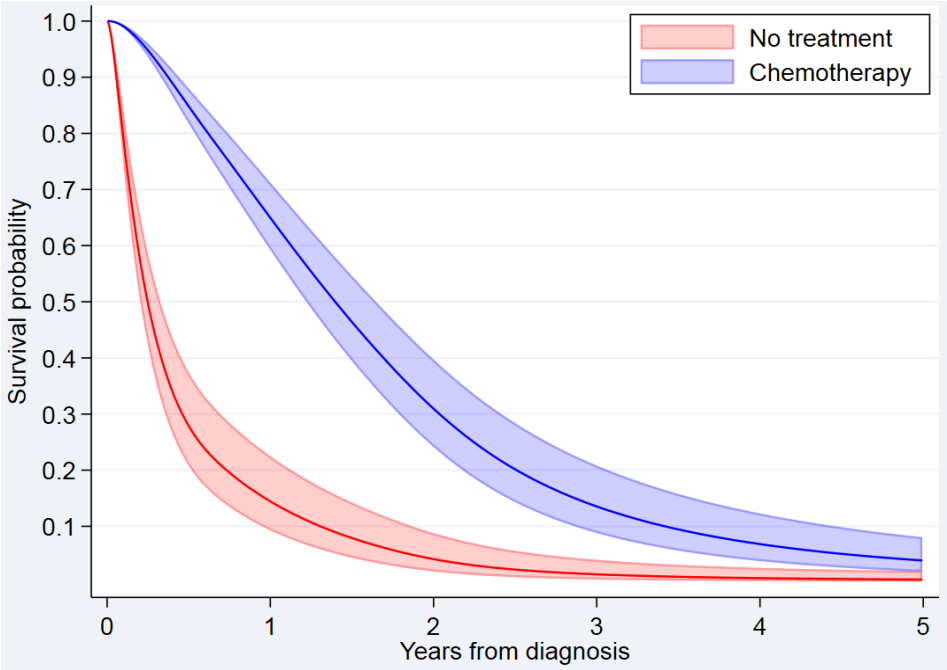
B.4 Survival analysis

Figure B.3: Survival curves by treatment type: lung cancer



Adjusted Kaplan–Meier survival curves based on the treatment classification we use in our work: no treatment, chemotherapy (standard of care), and innovative therapy. This graph is based on the estimates of a flexible parametric survival model which includes sex, age group, treatment modality, histology of tumor, Charlson index, surgery dummy, the use of palliative radiology, and year of diagnosis. Following Danesh et al. (2019), the model also includes interaction terms between age group and histology, treatment modality and year of diagnosis. In addition, age group, treatment modality, and year of diagnosis are included as time-dependent variables. The curves all refer to an hypothetical female patient, receiving palliative radiotherapy, no surgery, histology adenocarcinoma, age between 65-69, low Charlson index (healthy), diagnosed in year 2018 and treated at Toronto Central, treated according to the three treatment modes.

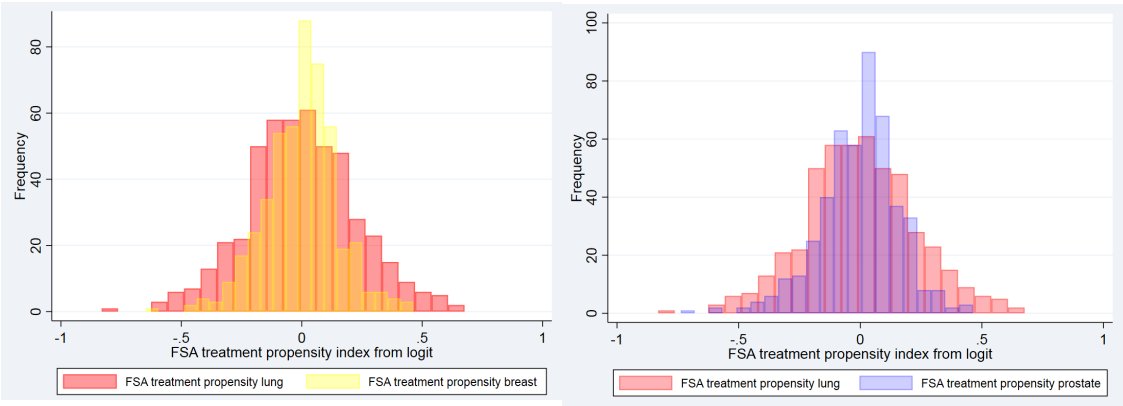
Figure B.4: Survival Curves by treatment type: colorectal cancer



Adjusted Kaplan–Meier survival curves for colorectal cancer patients based on whether they are treated or not. This graph is based on the estimates of a flexible parametric survival model which includes sex, age group, treatment modality, histology of tumor, Charlson index, surgery dummy, the use of palliative radiology, and year of diagnosis. Following Danesh et al. (2019), the model also includes interaction terms between age group and histology, treatment modality and year of diagnosis. In addition, age group, treatment modality, and year of diagnosis are included as time-dependent variables. The curves all refer to an hypothetical female patient, receiving palliative radiotherapy, no surgery, histology adenocarcinoma, age between 65-69, low Charlson index (healthy), diagnosed in year 2018 at Toronto Central Central.

B.5 Treatment rates

Figure B.5: Geographic variation in treatment rates: lung vs breast and lung vs prostate



Panel A

Panel B

Panel A and B show the risk-adjusted treatment rate at the FSA level; the rate is an empirical Bayes estimate of the FSA-level intercept from a random effect logit model of whether a patient receives treatment regressed on patient and tumor characteristics and a FSA-level random intercept. Panel A overlays the risk-adjusted treatment rate of lung cancer and breast cancer; Panel B overlays the risk-adjusted treatment rate of lung cancer and prostate cancer.

C Appendix C: Innovation in lung cancer treatment
For Online Publication

The treatment of lung cancer experienced major innovations in the past two decades. In the 1990s, several chemotherapeutic agents were discovered (paclitaxel, docetaxel, vinorelbine, gemcitabine, pemetrexed) and used in patients with advanced disease either as single therapy, or combined with platinum compounds (cisplatin and carboplatin). The use of platinum doublets led to increases in median survival to 9 months (1-year survival of 30%-35%), up from median survival of 3-4 months for untreated patients (1-year survival of approximately 15%, see Danesh et al. (2019), Sacher et al. (2015)). In the 2000s, improved understanding of the molecular basis of cancer and cheaper genetic sequencing led to treatments exploiting specific molecular abnormalities (targeted therapy). Treatment has become more complex over time, in part because of recognition of tumor-specific and patient-specific traits that predict a greater likelihood of success, or lack of success, with specific drugs. Though epidermal growth factor receptor (EGFR) mutations are only present in nearly 15% of lung cancer patients, they are strong predictors of the efficacy of specific inhibitors of EGFR such as erlotinib or gefitinib. Patients with EGFR-mutated tumors can achieve response rates higher than 70% and, most importantly, can achieve an overall survival longer than two years (de Castro-Carpeño et al. (2011)). Following a similar research path, discovery of fused proteins based on anaplastic lymphoma kinase rearrangements has opened up the possibility of blockage by specific inhibitors such as crizotinib. All of these targeted agents improve survival to up to 2 years in metastatic patients with relevant mutations. At the same time, they present a side effect profile that is milder and more manageable than standard platinum-based chemotherapy, making them good candidate treatments even for older patients with comorbidities. CCO guidelines recommend targeted agents even for patients with poor performance status, a measure of cancer patients' ability to tolerate therapy. Targeted therapy is allowed even for patients who are capable of only limited self-care and confined to bed for up to 50% of their time (Ellis et al. (2016)).

For patients without a targetable oncogene, new developments since the early 2000s stemmed from the use of immunotherapy. Immunotherapy, also called biological therapy, acts on the immune system to strengthen or restore its ability to fight cancer. Immunotherapy agents used to treat lung cancer are checkpoint inhibitors: they block the functioning of specific proteins called checkpoints (mostly PD-1 and PD-L1), which prevent the immune system from attacking cancer cells. Monoclonal antibodies atezolizumab, nivolumab, and pembrolizumab are the most commonly used immune checkpoint inhibitors for patients with non-small cell lung cancer. They were first introduced as second-line treatment of advanced NSCLC, where they showed substantial improvements compared to standard chemotherapy. Use in first-line settings for patients without mutations, alone or in combination with chemotherapy, showed gains in overall survival comparable to targeted therapy. They cause frequent but non severe immune-related adverse events and are generally better tolerated than classic cytotoxic chemotherapeutic agents. For this reason, they are broadly approved as first-line treatment for patients with advanced NSCLC who do not have contraindications to immunotherapy and whose tumors do not harbor actionable driver mutations: Shields et al. (2021).

D Appendix D: Additional Figures and Tables

Table D.18: Test of quasi-random assignment of physician to the neighborhood

	Physician treatment propensity
Share heavy smokers	-2.160
	(1.159)
Share heavy drinkers	-0.521
	(0.635)
Pollution (pm 2.5)	6.34e-05
Quintiles of marginalization index:	(0.000260)
2. instability	0.0257
	(0.0618)
3. instability	0.0170
	(0.103)
4. instability	0.0192
	(0.119)
5. instability	0.00444
	(0.134)
2. deprivation	-0.0265
	(0.100)
3. deprivation	-0.154
	(0.146)
4. deprivation	-0.0611
	(0.141)
5. deprivation	-0.0558
	(0.181)
2. ethnic concentration	-0.00912
	(0.126)
3. ethnic concentration	-0.100
	(0.148)
4. ethnic concentration	-0.0649
	(0.154)
5. ethnic concentration	-0.147
	(0.163)
Share with high school degree	0.960
	(1.847)
Unemployment rate	-0.000790
	(0.0284)
Median income	-2.12e-07
	(1.31e-05)
% income from welfare payments	-0.0140
	(0.0131)
Population density	-1.43e-05
	(1.65e-05)
Share of South-Eastern Asian immigrants	1.241
	(0.451)
Urban (%)	0.0452
	(0.109)
Observations	15,761
R-squared	0.097
Year FE	Yes
Joint $p-$value	0.103

The table reports the OLS estimates of a regression of physician treatment propensity on neighborhood characteristics. The regression is estimated on the sample of patients described in Table 1. Physician treatment propensity is an empirical Bayes estimate of the physician-level intercept from a random effect logit model of whether a patient receives treatment regressed on patient and tumor characteristics and a physician-level random intercept. The $p-$value reported at the bottom of the column is from an $F-$test of the joint significance of the variables listed in the rows. Robust standard errors, two-way clustered at the physician and neighborhood level, are reported in parentheses.

Table D.19: Survey: attitude toward lung cancer patients

	(1) Weighted means	(2) Regression: Reference group low sympathy toward lung cancer patients (0/1)
1. Reference group's negative attitude towards smokers		0.12
Most people you know look down on smokers. Do you...?		(0.05)
1: Strongly agree	0.37	
2: Somewhat agree	0.35	
3: Neither agree nor disagree	0.17	
4: Somewhat disagree	0.07	
5: Strongly disagree	0.04	
2. Reference group's perception of lung cancer as a hopeless disease		0.39
Most people you know think that treating metastatic lung cancer patients is not worthwhile as it takes away from the resources available to treat other patients and the quality of life when receiving treatment for lung cancer is poor anyway		(0.06)
1: Strongly agree	0.06	
2: Somewhat agree	0.08	
3: Neither agree nor disagree	0.12	
4: Somewhat disagree	0.2	
5: Strongly disagree	0.54	
3. Reference group's no support for lung cancer research		0.38
Most people you know would not support lung cancer research aimed at finding better treatments. Instead, they would prefer supporting research on other types of cancer		(0.06)
1: Strongly agree	0.04	
2: Somewhat agree	0.09	
3: Neither agree nor disagree	0.11	
4: Somewhat disagree	0.23	
5: Strongly disagree	0.52	
4. Reference group's sympathy for lung cancer patients		n/a
Most people you know have less sympathy toward people with lung cancer than people with other types of cancer.		
1: Strongly agree	0.09	
2: Somewhat agree	0.14	
3: Neither agree nor disagree	0.11	
4: Somewhat disagree	0.17	
5: Strongly disagree	0.51	
5. Shared opinion with reference group		0.06
Overall, do you share the opinions of most people you know regarding lung cancer patients?		(0.02)
1. Yes	0.75	
2. No	0.25	
6. Own degree of sympathy toward lung cancer patients (0/1)		0.63
1. Sympathy above low	0.18	(0.07)
2. Sympathy equal or below low	0.82	

The table summarizes responses to questions from the survey described in Section 3.4. Column 1 reports the weighted averages for each indicated variable. Column 2 reports the coefficient and the standard error (in parentheses) of the regression: $y = \beta \cdot$ reference group low sympathy toward lung cancer patients (0/1) $+ \varepsilon$, where y identifies the survey variable.

Table D.20: Regimen/therapy choices: bottom level of a nested logit model and social effects

	(1) Carboplatin therapy	(2) Single-agent therapy	(3) Innovative therapy
Share untreated	0.463	-0.242	-0.309
	(0.312)	(0.601)	(0.357)
Surgery (0/1)	-0.962	-1.617	-0.819
	(0.257)	(0.656)	(0.310)
Adenocarcinoma	0.508	0.0707	0.732
(0/1)	(0.258)	(0.562)	(0.307)
Squamous cell	0.307	0.0565	-0.973
(0/1)	(0.274)	(0.591)	(0.355)
Charlson index	0.0999	0.272	-0.158
(medium)	(0.104)	(0.204)	(0.121)
Charlson index	0.434	0.702	-0.149
(high)	(0.130)	(0.236)	(0.157)
Controls:			
Patient health	Yes	Yes	Yes
Patient socio-demo	Yes	Yes	Yes
3-digit zip code	No	No	No
Physician characteristics	Yes	Yes	Yes
Fixed effects:			
Physician	No	No	No
Year	Yes	Yes	Yes
Hospital	Yes	Yes	Yes
Observations		14,592	

The table reports the parameter estimates and standard errors of selected variables for the bottom level of a nested logit model of therapy choice: cisplatin, carboplatin, single-agent therapy, and innovative therapy. The excluded base alternative is cisplatin. The "share untreated" refers to the cumulative share of untreated patients diagnosed in the three previous years in the same three-digit zip code. The excluded health status category is the lowest Charlson (most healthy individual). The model controls for a constant for each therapy alternative. Standard errors are in parentheses.

Table D.21: Summary statistics: innovation and market size

Variable		Mean	Std. dev.
R&D spending	overall	159,133	156,326
in $'000	between		160,338
	within		27,208
Treated patients	overall	66.47	53.54
in '000	between		55.37
	within		5.66
Diagnosed patients	overall	82.72	64.20
in '000	between		66.18
	within		8.53
Treatment rate	overall	79.49	10.63
	between		10.87
	within		1.97

The table reports unweighted averages by cancer site and year, within standard deviation (variation over years for a given cancer site) and between standard deviation (variation across cancer sites). The number of observations is 180 (12 cancer sites $\times$ 15 years) for the variable R&D spending, 120 (12 cancer sites $\times$ 10 years) for all the other variables.