NORTHWESTERN UNIVERSITY

Frictions to treatment decisions at the end of life

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Economics

By

Aleksandra Joanna Paluszynska

EVANSTON, ILLINOIS

June 2023

Abstract

Hazy decisions: The effect of dementia on medical decision-making

I estimate the causal effect of having dementia on the course of treatment for unrelated diseases by leveraging differences in the relative time of onset of dementia and the other condition in a difference-in-differences event-study framework. To demonstrate this approach I look at heart attacks and show that after accounting for individual and calendar-time fixed effects the “dementia first” and “heart attack first” groups exhibit parallel trends in health care utilization before the heart attack. I find that health care cost is persistently lower following a heart attack for those who had dementia already and that this effect is driven by decreases in frequency of care. Dementia patients who experience a heart attack reduce the number of interactions with the health care system, while those who do not yet have dementia markedly increase their frequency of care in response to the health shock. I establish that this finding holds universally for a variety of other acute and chronic conditions and is not driven by higher mortality among dementia patients. Long-term reductions in medical care received are driven by care over which patients have the most discretion such as diagnostic tests and prescription drugs, which points to reduced follow-up care and treatment adherence

as possible mechanisms behind my estimates.

Do procedure codes matter? A case study of Advance Care Planning

This project aims at understanding the way in which new procedure codes affect provision and cost of medical care. Primarily used for reimbursement purposes, new codes can generate additional revenue for existing services or lead to modifications in their insurance coverage that in turn affect physicians' financial incentives and can lead to changes in both quantity and kind of services provided. I study the case of Advance Care Planning (ACP), a 30 minute counseling service that became reimbursable by Medicare in 2016, to show that the new reimbursement lead to a 16% increase in the provision of all counseling services among physicians who used the new procedure code. Importantly, only 3pp of this increase can be attributed to the provision of ACP itself, and the remaining 13pp reflects an increase in the use of other services, such as optional screening for depression and alcohol abuse. Finally, I show that this increase is driven by recipients of ACP, suggesting that a more generous reimbursement of one service can substantially affect the provision of other services, which should be incorporated into projections of the potential cost of new reimbursement policies.

Acknowledgment

I thank David Dranove, Matthew Notowidigdo and Molly Schnell for their invaluable guidance. I also thank seminar participants at Northwestern University for helpful comments. All errors are my own.

Contents

List of Tables

List of Figures

Chapter 1

Hazy decisions: The effect of dementia on medical decision-making

1.1 Introduction

People at old age face a variety of high-stakes decisions concerning insurance, medical treatment, end-of-life care and living arrangements. Although of great consequence, these decisions often come about at a time when the individual is not fully capable of making them due to age-related cognitive decline. At the extreme end, dementia is a loss of routine memory, skills and knowledge that interferes with independent functioning and can thus hinder decision-making. The total cost of medical care of the 6.2 million dementia patients in the US in 2021 was $355 billion (Alzheimer's Association, 2021). In fact, the National Institutes of Health spend twice as much on dementia-related research as on research devoted to heart disease (National Institutes of Health, 2022). Dementia is progressive, incurable, and increasingly prevalent as our society ages, so understanding its effect on decisions concerning medical care is crucial to policy.

One important question to consider is how dementia affects medical decisions concerning *other* diseases. Chronic conditions, that is diseases that last at least one year, are a major force behind the recent rise in health care expenditures in the United States (Thorpe, Ogden, and Galactionova, 2010). People suffering from at least one chronic condition accounted for 86% of health care spending in 2010, whereas those with more than one condition accounted for nearly three quarters of it (Gerteis, Izrael, Deitz, LeRoy, Ricciardi, Miller, and Basu, 2014). While a major policy concern, chronic conditions are well-defined in medical records and their onset constitutes a decision-making setting whenever multiple courses of treatment are available, which often vary in cost and physical burden to the patient. Thus, onset of chronic conditions constitutes a meaningful setting for studying the effect of dementia on medical decision-making. On one hand, having dementia could exacerbate the agency problem between the patient and physician as well as concerns for supplier-induced demand, leading to more invasive treatment and higher cost. On the other hand, memory issues of dementia patients could lead to trouble with treatment adherence, and complicate the management of invasive treatment options.

To date, no causal estimates of the effect of dementia on health care utilization exist. A variety of work in the medical literature documents the positive correlation between dementia and health care cost, but such estimates are usually based on a comparison between patients with and without dementia that suffers from selection bias. In particular, dementia patients are more sick overall and might seek more health care in the first place, as interacting with the health care system is a prerequisite for a formal dementia diagnosis. Crucially, it is impossible to study the causal effects of dementia through experimentation, so estimates can only be obtained through observational studies, common in economics, but not yet undertaken in the context of dementia.

In this paper I estimate the effect of dementia diagnosis on medical care of unrelated diseases using a dynamic difference-in-differences design. Specifically, I use the differential timing of dementia diagnosis and onset of another health condition such as heart attack to identify the causal effect of dementia on health care cost attributable to the other condition. Using the onset of the unrelated disease as the event, I estimate the cost trajectory around it with a standard two-way fixed effects event-study regression. Assuming that the condition's onset coincides with the timing of decisions concerning medical treatment of the condition, I compare the estimated cost trajectory between those who had dementia at the time to those who had it diagnosed afterwards. Restricting the comparison to people diagnosed with dementia mitigates the aforementioned concern of selection bias. The difference in health care utilization after condition onset between the "dementia first" and "condition first" people captures the effect that having dementia at the time of another health shock has on the subsequent treatment trajectory.

I find that medical utilization following a heart attack is lower for those who had dementia at the time of the heart attack as compared to those who had it diagnosed afterwards. In other words, dementia patients receive less treatment associated with the heart attack. Moreover, this difference is driven by frequency of care: dementia patients return to their pre-heart attack frequency of care within six months, whereas non-dementia patients experience a persistent 50% increase in frequency of care following the heart attack. I find that this is driven by types of care over which patients have the most discretion, such as outpatient surgeries, pathology and lab or prescription drugs. Similar patterns also arise if, instead of heart attacks, I consider a different chronic or acute condition, such as cancer or pneumonia.

To identify the causal effect of dementia on treatment of another disease, I assume that the relative time of onset of dementia and the other disease is as good as randomly assigned. For this to hold, the disease in question should be biologically unrelated to dementia in the sense that having it does not affect the propensity of having dementia and vice versa. Moreover, as I use formal diagnoses on one's medical claims as a proxy for actual disease presence, identification requires that these diagnoses be independent. Thus, a disease that is biologically independent from dementia could violate this assumption if it affects the probability of dementia diagnosis, conditional on one's underlying dementia severity. In the case of heart attack, it is usually severe enough that it can hardly go undetected, alleviating concerns that dementia could affect the likelihood of a formal heart attack diagnosis. However, it is possible that having a heart attack increases health monitoring, which in turn leads to higher probability of dementia diagnosis and possibly biasing my estimates. To address this concern, I show that my conclusions hold across a variety of other health shocks, regardless of the empirical relationship between their diagnoses and those of dementia.

I use claims data on 8 million Medicare Advantage beneficiaries from the Health Care Cost Institute's (HCCI) database that encompasses a substantial fraction of privately insured people over 65 years old in the United States. These data contain records of all interactions of an enrollee with the health care system including inpatient, outpatient and professional services as well as purchases of prescription drugs. I measure the time of onset of a disease as the date when a corresponding diagnosis code appears on a person's record, following the guidelines of the Chronic Conditions Warehouse, whereas care frequency refers to the number of days per quarter in which the patient had a specific kind of claim. The HCCI database has been used in the economic literature (Curto, Einav, Finkelstein, Levin, and Bhattacharya, 2019; Einav, Finkelstein, Oostrom, Ostriker, and Williams, 2020) and comes with the caveat that beneficiaries drop out of the panel if they switch insurers.

To compare the trajectories of care frequency between the "dementia first" and "heart attack first" groups I run a two-way fixed effects event-study regression where heart attack is the event, and time-to-event dummies are interacted with an indicator for being in the treated group (i.e., the "dementia first" group). To be able to separately identify individual, calendar-time and event-time fixed effects I also include a pure control group of beneficiaries who had dementia but not a heart attack. The interacted time-to-event dummies trace out the difference in care frequency between the "dementia first" and "heart attack first" groups around the time of heart attack. I provide evidence corroborating the parallel trends assumption by showing that the difference in utilization between these two groups is not statistically significant before the event.

A major concern when comparing health care utilization of "dementia first" and "heart attack first" patients after their heart attack is that this is when the latter group gets diagnosed with dementia, so part of the observed difference is driven by dementia treatment. As dementia is a continuously evolving disease, I interpret dementia diagnosis as a designation that it is severe enough to be noticed and recorded by the physician. Empirically, this happens at a time of more health care utilization (Deb, Thornton, Sambamoorthi, and Innes, 2017), which could be because worsening health accelerates dementia progression or because it brings about interactions with the health care system that in turn lead to a formal diagnosis of conditions undetected before. By examining heterogeneity of the effect by the time of dementia onset I show that care frequency for "heart attack first" patients remains significantly elevated with respect to "dementia first" if we restrict attention to the period before their dementia diagnosis. Finally, the magnitude of the estimated effect is too large to be explained by care associated with dementia alone, as "dementia first" interact with the health care system more frequently before the heart attack but less frequently afterwards, suggesting that their heart attacks are indeed treated differently.

I show that the long-term decrease in frequency of care for dementia patients following another health shock is not specific to heart attacks. In particular, I consider 21 major chronic conditions and 6 acute conditions (e.g., pneumonia) to show that their onset is *never* followed by a long-term increase of frequency of care among "dementia first" patients. At the same time, both "health shock first" and "no dementia" patients (those who had the health shock but not dementia) exhibit significant long term increases in care frequency following a health shock. Importantly, as true independence between diagnoses of dementia and other conditions is not only hard to show but also unlikely, the fact that my main finding holds across conditions that are correlated with dementia in various ways is reassuring. For example, heart attacks are much more common after dementia, presumably because they tend to appear at older ages, while the opposite is true for heart failure; nevertheless both lead to a similar conclusion when it comes to dementia decreasing care frequency following another health shock. Moreover, the result is not driven by differences in mortality as restricting attention to survivors has a minimal effect on the estimates.

This paper makes a key contribution to the existing literature on the medical cost of dementia by providing causal estimates. A notable strand of this literature relies on the Health and Retirement Study which includes cognitive ability measures that allow researchers to abstract from formal dementia diagnoses to show that predicted dementia status is strongly correlated with self-reported medical costs (Hurd, Martorell, Delavande, Mullen, and Langa, 2013; Kelley, McGarry, Gorges, and Skinner, 2015). A related strand of literature relies on claims data and yields positive correlation between a formal dementia diagnosis and realized health care utilization with marked cost increases around the time of diagnosis (Lin, Zhong, Fillit, Chen, and Neumann, 2016; Hoffman, Maust, Harris, Ha, and Davis, 2022). However, as noted by Deb et al. (2017) in their review of this literature, a third of dementia diagnoses occurs during hospitalizations for another serious condition, which casts doubt on the causal interpretation of the aforementioned estimates of the cost of dementia. The bulk of this cost that can be attributed to home care and nursing home care is indeed plausibly driven by dementia, which impairs one's ability to function independently. However, the mechanism behind the estimated cost increase for other kinds of care remains unexplained, especially that no treatment of dementia exists, which points to the possibility of omitted variable bias. This paper leverages the timing of onset of unrelated conditions to estimate a negative effect of dementia on medical spending attributed to the condition of interest that allows for a causal interpretation.

Second, this paper contributes to the literature studying the relationship between cognition and decision-making. A growing body of literature documents sub-optimal choices made by the elderly population in the US when it comes to insurance and retirement planning (Abaluck and Gruber, 2011; Keane and Thorp, 2016). These papers postulate choice complexity as the main driver of consumers deviating from rationality and it is not clear to what extent this is exacerbated by cognitive decline. When it comes to financial well-being, cognitive ability is correlated with portfolio choice (Christelis, Jappelli, and Padula, 2008) and credit-management mistakes (Agarwal, Gabaix, Driscoll, and Laibson, 2009). Moreover, people tend to see age-related cognitive decline as a risk to sound financial decisions (Ameriks, Caplin, Lee, Shapiro, and Tonetti, 2022) and dementia specifically has been linked to missed credit card payments and subprime credit scores (Nicholas, Langa, Bynum, and Hsu, 2020). In this paper, I discuss possible ways in which cognitive decline introduces frictions into medical decision-making that can be thought of as an agency problem with the physician making decisions on behalf of the patient. I focus on dementia, which is a type of cognitive decline that is increasingly prevalent, especially among people who receive a lot of medical care. Moreover, by purging my estimates from the direct effect of dementia I document decreases in frequency of interactions with the health care system that can plausibly be driven by a decline in one's cognitive ability.

Third, this paper is related to a growing literature on regional variation in health care utilization. In the light of health care spending continuing to grow rapidly in the United States, stark differences in spending between locations provoke attempts to identify sources of inefficiencies in high-cost regions (Fisher, Bynum, and Skinner, 2009). Finkelstein, Gentzkow, and Williams (2016) show that approximately 50% of cost variation can be attributed to the demand side, i.e., patient health and preferences, while Cutler, Skinner, Stern, and Wennberg (2019) show that in the case of heart attack treatments patient preferences do not seem to play a role, unlike physician beliefs. Furthermore, Fisher, Wennberg, Stukel, Gottlieb, Lucas, and Pinder (2003) argue Medicare enrollees in high-spending regions do not have better outcomes and Song, Skinner, Bynum, Sutherland, Wennberg, and Fisher (2010) document regional variation in diagnostic practices which are likely to affect spending. This paper studies dementia as a demand-side driver of health care utilization and finds that it can explain some differences in cost between people suffering from a given chronic condition. Furthermore, my empirical strategy can be used to study interactions between other chronic conditions, particularly if they have the capacity to affect decision-making (e.g., depression).

Last but not least, dementia is a growing concern for policy-makers that until now has hardly been studied by economists despite being uniquely suited for the use of economic methods (Chandra, Coile, and Mommaerts, 2020). Apart from dementia's exorbitant cost, those affected by it are a particularly vulnerable population, shown to receive less pain-relief (Morrison and Siu, 2000), more invasive medical interventions (Goldfeld, Hamel, and Mitchell, 2013) and more burdensome care at the end of life (Mitchell, Teno, Kiely, Shaffer, Jones, Prigerson, Volicer, Givens, and Hamel, 2009) than the corresponding survival benefits would justify. This paper documents an empirical pattern that suggests that having dementia at the onset of another disease leads to less frequent interactions with the health care system, perhaps due to lack of follow-up care or neglect. Although further research is required to shed light on the exact mechanisms that underlie this finding, the paper points to long-term treatment consistency as a possible source of care improvement.

The remainder of the paper is structured as follows. Section 1.2 describes the dataset, population of interest and empirical strategy. Section 1.3 shows estimates of differences in health care utilization following a heart attack and establishes the main drivers of these estimates. Section 1.4 summarizes the most important results for a variety of other chronic and acute conditions and relates them to the heart attack case. Section 1.5 concludes.

1.2 Data and methods

This section sets the stage for my empirical analysis. I start by describing the dataset and population of interest in section 1.2.1. Then, in sections 1.2.2 and 1.2.3 I provide background information on dementia and other health conditions considered, respectively. In section 1.2.4 I discuss empirical differences between the treated and control groups as well as the possibility of inverse propensity score weighting to account for different covariate distributions. Section 1.2.5 outlines my empirical design and the corresponding identifying assumptions. Finally, section 1.2.6 considers potential mechanisms through which having dementia at the onset of another condition could affect subsequent health care utilization.

1.2.1 Medicare Advantage claims data

This paper uses data from Health Care Cost Institute's (HCCI) commercial claims database that includes claim-level information from three large insurers in the United States: CVS Health/Aetna, Humana, and Blue Cross Blue Shield (HCCI, 2020). As the analysis concerns dementia, I restrict attention to Medicare Advantage (MA) plans that are offered by these insurers as an alternative to traditional fee-for-service Medicare coverage. The HCCI data span years 2012 to 2020 in which the three participating insurers accounted for around 40% of the MA market, depending on year (KFF, 2022). Importantly, overall enrollment in MA plans grew steadily over the years from 13 million people in 2012 to 24 million in 2020, which corresponded to 27% and 43% of the Medicare population, respectively. Although Medicare eligibility is primarily based on age, as it covers those aged 65 and older, some younger people become eligible through disability status and are therefore excluded from my analysis to limit the possibility of selection along that dimension.

In principle, MA and Traditional Medicare (TM) serve the same purpose, but they achieve it in distinct ways. TM uses a fee-for-service model that sets prices for medical services to then partially reimburse enrollees for their utilization, while MA operates on a managed care basis. Thus, insurers receive a risk-adjusted capitated monthly payment for each enrollee in exchange for providing medical care to them, which must encompass all services covered by TM but can also include additional benefits to attract consumers as they are often faced with a wide selection of MA plans to choose from (39 on average, KFF, 2022). From the patient's perspective, MA plans often restrict access to some providers while offering better cost-sharing, for example by introducing annual out-of-pocket maximums. Unsurprisingly, differences between TM and MA plans lead to substantial selection, as MA enrollees are younger and healthier on average. Moreover, MA plans are characterized by 9-30% lower health care utilization than TM, depending on the method used for accounting for differences in patient populations (Curto et al., 2019).

For the relevant population of MA enrollees, the HCCI database contains all of their insurance claims covered by Medicare, which are split into the following four categories. Inpatient claims come from providers designated as hospitals as well as physicians whenever they bill for services provided in an inpatient hospital setting. Outpatient claims are submitted by institutional outpatient providers, e.g., specific departments within a hospital, and include emergency room care preceding a hospital admission, home health care and durable medical equipment rentals. Physician claims correspond to services provided by physicians in non-institutional settings which covers the bulk of routine care, whereas pharmacy claims record purchases of prescription drugs. Importantly, nursing home claims appear in this database only if they are connected to a specific injury and preceded by a hospital admission, as otherwise they are not covered by Medicare. Finally, all claims come with detailed codes on the service in question as well as the final amount paid to the provider (i.e., allowed amount) with the accompanying split between the patient and insurer. I use total spending on all of the aforementioned kinds of medical care excluding prescription drugs as a baseline measure of health care utilization and look into more specific categories of spending to shed light on potential mechanisms behind my results.

The HCCI data come with three important caveats. First, for confidentiality reasons it is stripped of all identifying information when it comes to insurers and insurance plans, which precludes from accounting for plan characteristics in regression analysis[1]. Second, to protect enrollee privacy these data are not research-identifiable, so in addition to encrypted enrollee identifiers it contains detailed geographic information (i.e., 5-digit zip code) but not age nor

---

[1] As noted by Curto et al. (2019) who used a previous version of the HCCI database, it omits highly capitated plans, Special Needs Plans and plans with certain data issues. However, no documentation on the exclusion criteria exists, so I follow previous papers in assuming that the coverage pertains to the most salient types of plans in the MA market.

any information concerning death. As the bulk of my analysis will rely on within-individual variation, this is not a major concern, but precludes me from investigating mortality as an outcome as well as speaking to the role of age in driving differences in health care utilization between various groups or treatment effect heterogeneity. Finally, when an enrollee dies or switches insurers, they disappear from the panel and no information on the underlying reason is available[2]. Abaluck, Bravo, Hull, and Starc (2021) report that on average 4.7% of MA enrollees die and 11.4% of them switch insurers every year, which introduces potential selection issues into my research design.

1.2.2 Background on dementia

The term *dementia* refers to a group of symptoms including loss of memory, language and problem-solving skills that interfere with a person's ability to perform everyday activities. It can be caused by a variety of underlying health conditions, most notably by Alzheimer's disease that accounts for 60-80% of all dementia cases (Alzheimer's Association, 2021). Importantly, although typical age-related changes in cognition can lead to forgetfulness or confusion, dementia is distinct due to its severity and interference with basic daily tasks (e.g., a forgetful person might have trouble finding the right word but would not call common things by the wrong one). Furthermore, many conditions such as depression, Lyme disease and certain vitamin deficiencies can exhibit dementia-like symptoms that are transitory or reversible with treatment, setting them aside from dementia, which is inherently progressive.

For the purposes of my analysis I identify dementia using diagnosis codes. Specifically, I

---

[2] When an enrollee switches insurers but remains within the scope of the database their identifier changes, which results in being treated as a different person in my analysis.

define one's dementia onset as the first time that a relevant diagnosis code[3] appears on her claims. As dementia is progressive, I interpret such diagnosis as an indicator for its symptoms becoming severe enough for the health professional in question to record it on one's claim. Of course, conditional on the actual severity of dementia, the probability of a diagnosis can depend on a variety of factors, such as the sensitivity of a health professional to perceiving dementia symptoms. For the sake of simplicity, I will abstract from such considerations and interpret one's dementia onset to mean that the symptom severity crossed some average threshold; naturally I will only focus on the first time that this happens. What matters for my analysis is that I expect dementia severity to be on average higher among those who had dementia diagnosed a year ago as compared to those who have it diagnosed today.

In Table 1.1 I show summary statistics concerning dementia in the HCCI data. Note that in my sample of people aged 65 or older the prevalence of dementia is 3.7%, markedly lower than the national average for this age group, 11.3% (Alzheimer's Association, 2021) and consistent with the MA population being healthier on average. By comparing columns 1 and 2 we can see that dementia patients have higher levels of health care utilization on average. In panel A we see that dementia patients spend 60% more on inpatient, outpatient and physician care (i.e., total cost excluding drugs) but there is substantial variation in this difference across spending categories. For example, dementia patients spend almost 15 times more on nursing home care but only 10% more on diagnostic imaging, which is often pointed to as a source of wasteful care (e.g., back imaging in the case of nonspecific low back pain, Schwartz, Landon, Elshaug, Chernew, and McWilliams, 2014).

---

[3] I follow the Chronic Conditions Warehouse's list of diagnosis codes for "Alzheimer's Disease and Related Disorders or Senile Dementia".

Table 1.1: Spending patterns of dementia patients

	Dementia?		Dementia	
	No (1)	Yes (2)	Pre (3)	Post (4)
A. Average annual utilization ($)				
Inpatient, outpatient, physician care	7260	11637	8418	19496
Drug cost	1050	1081	1120	1058
Nursing home cost	56	832	386	1539
Diagnostic imaging cost	325	357	308	494
B. Probability of health shocks (%)				
Fall	5	33	13	20
Hip fracture	1	10	4	7
Pneumonia	6	33	12	20
C. Other characteristics				
Number of chronic conditions	3	9		
Number of beneficiaries	7.59M	288K		

Notes. Based on data for years 2012-2019. Only includes care covered by Medicare (e.g., only a fraction of nursing home care is). Column 2 uses people for whom a dementia diagnosis code is observed, column 1 uses everyone else. Columns 3 and 4 use claims of dementia patients before and after their dementia diagnosis, respectively. I use CCWs list of 25 primary chronic conditions for Panel C.

Overall, dementia patients have higher health care spending even before their diagnosis (column 3), consistent with selection bias, as health care use is a prerequisite for any formal diagnosis. At the same time, they are at a much higher risk of suffering from various injuries (Panel B) both before and after dementia diagnosis, so higher utilization could also be explained by the progression of dementia symptoms throughout one's life, regardless of formal diagnosis. The diagnosis is clearly most meaningful when we look at nursing home claims as it is associated with a quadrupling of the nursing home spending captured by the HCCI data. Crucially, this accounts for less than 10% of the overall health care spending of dementia patients in my sample which is in stark contrast with the 75% found in other papers due to the aforementioned limited coverage of nursing home services by MA plans.

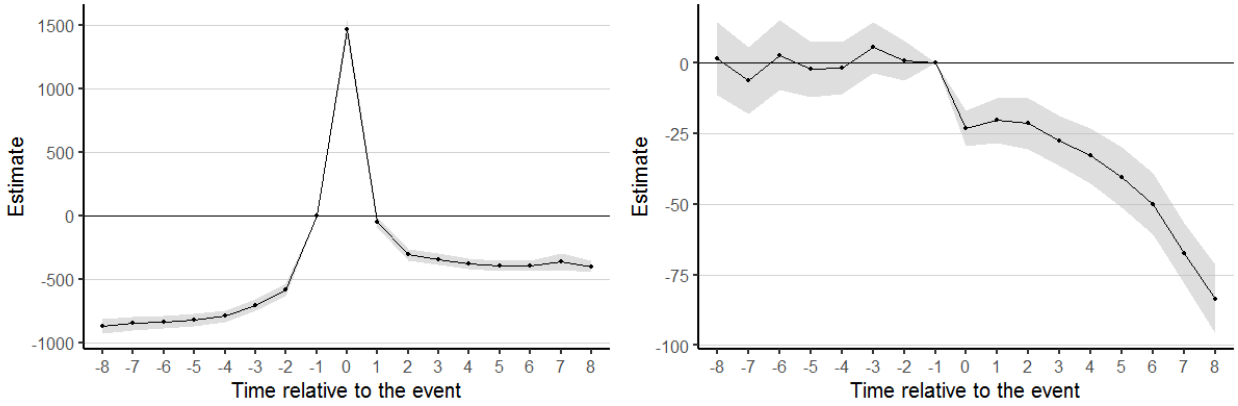
When it comes to prescription drugs, differences between the two groups are negligible, and comparing columns 3 and 4 reveals that drug utilization among dementia patients falls after the formal diagnosis. Importantly, as of 2020 several medications that alleviate dementia symptoms were available, even though none of them slow or stop the disease's progression[4] (Alzheimer's Association, 2021). Moreover, dementia patients are more likely to suffer from almost any other chronic condition – in Panel C we see that on average they are diagnosed with 3 times as many chronic conditions as non-dementia patients. Together these facts make dementia patients' prescription drug utilization seem surprisingly low, pointing to the possibility of issues with drug adherence. In Figure 1.1b I plot the trajectory of prescription drug utilization for dementia patients to document that a marked and persistent drop in drug use happens at the same time as a formal dementia diagnosis. Due to these distinct patterns of drug use, for the remainder of the paper I will refer to inpatient, outpatient and physician services captured by the HCCI data as "medical services" and analyze prescription drug spending separately.

In Figure 1.1a I show the trajectory of utilization of medical services around the time of dementia diagnosis. The increase in their use appears well before the diagnosis, which is consistent with prior literature that found that dementia is often diagnosed during a hospitalization related to a different health shock. Moreover, the sharp increase in spending around event time is temporary and within a year spending stabilizes at around $500 more compared to the level from before the diagnosis. This empirical pattern, although not causal, points to the possibility of dementia causing a trend break in health care utilization, which

---

[4] The first drug that could potentially reverse the damage to brain cells caused by Alzheimer's disease, aducanumab, has been approved by the FDA in 2021 (Woloshin and Kesselheim, 2022).

Figure 1.1: Trajectory of health care utilization around dementia diagnosis



(a) Inpatient, outpatient and physician services

(b) Prescription drugs

Notes. Time is in quarters. The plots show the average trajectory of cost of medical services (i.e., inpatient, outpatient and professional services) and prescription drug spending around the time of dementia diagnosis after accounting for common time trends and idiosyncratic differences between individuals. The estimates correspond to relative event-time dummies from a two-way fixed effects event-study regression with dementia diagnosis as the event.

would violate the parallel trends assumption central to my main analysis. Moreover, the spike in spending is completely temporary if we restrict attention to a balanced panel, so the long-term increase in Figure 1.1a could be driven by elevated spending around the time of death rather than a trend break at the time of dementia.

1.2.3 Other medical conditions

Chronic conditions are defined broadly as diseases that can limit the capacity to perform daily activities and require ongoing medical attention for at least 1 year. In order to study them in my sample, I follow Chronic Condition Data Warehouse's (CCW) algorithms to identify the time of onset of 27 chronic conditions, including heart attack, anemia and various cancers (CCW, 2020). The algorithms define a set of diagnosis and procedure codes[5]

---

[5] Diagnosis codes in HCCI data follow the International Classification of Diseases (ICD) coding system that classifies diseases, injuries, health encounters and inpatient procedures. This classification system has

as well as the number and type of medical claims that should feature them in order to indicate the presence of a particular condition. Moreover, CCW defines a reference period for each condition that determines the length for which it is thought to last after it is first identified. Gorina and Kramarow (2011) investigate the accuracy of these algorithms for a subset of conditions as compared to disease designations based on data on outcomes instead of claims (i.e., questionnaires, medical examinations and death certificates). Unsurprisingly, they found CCW to be more accurate for diseases that require frequent health care utilization (e.g., diabetes but not arthritis) and showed that accuracy can be increased as the reference period expands. Such measurement error is one of the inherent caveats of using medical claims data and calls for caution when it comes to interpreting the results in terms of consequences of disease presence instead of formal disease diagnosis.

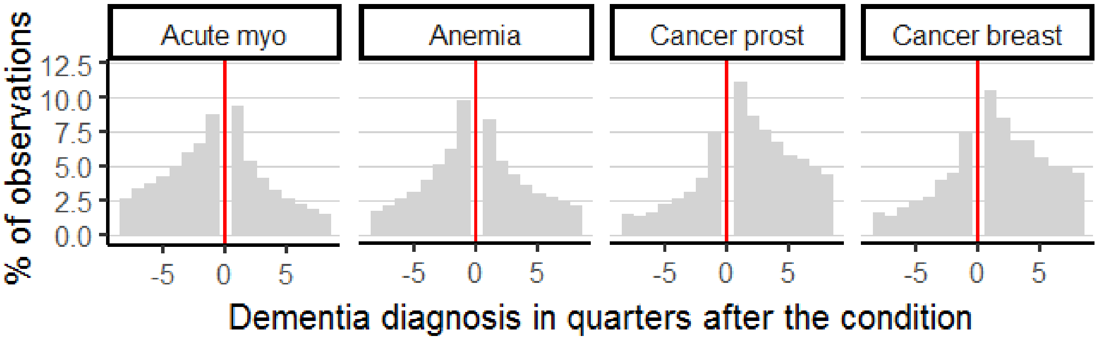
I focus on onset of a chronic condition as a meaningful event for one's health care utilization. Chronic conditions are often cited as the underlying reason of the bulk of health care spending in the United States. For example, Yu, Ravelo, Wagner, Phibbs, Bhandari, Chen, and Barnett (2003) show that veterans who suffer from at least one chronic condition (72%) account for 96% of medical spending of the Veterans Administration. Although one cannot claim that this spending is caused by chronic conditions, the population suffering from them is certainly of interest to policymakers (Parekh, Kronick, and Tavenner, 2014). In fact, more than two thirds of Medicare beneficiaries suffer from more than one chronic condition (Lochner and Cox, 2013), which makes the question of how multiple conditions interact with each other even more pressing. To analyze this, I restrict attention to the first time each

---

been revised in 2015 from ICD-9 to ICD-10, so CCW algorithms rely on either set of codes depending on the date of the claim.

condition is diagnosed for a beneficiary and compute the distance in quarters between this diagnosis and that of dementia. As multiple chronic conditions can be diagnosed during the same health shock (e.g., during one hospitalization) I exclude diagnoses that happen in the same quarter.

Figure 1.2: Distribution of relative timing of condition onset and dementia diagnosis



Notes. Each bar indicates the percentage of people that had dementia diagnosed a specific number of quarters after the condition in question. Diagnoses in the same quarter (0 on horizontal axis) are omitted. Negative (positive) values indicate dementia diagnoses before (after) the condition.

I plot the distribution of relative diagnosis times for a few conditions in Figure 1.2. Clearly, these distributions are not symmetric: in the case of prostate cancer an overwhelming fraction of the mass points to positive differences, i.e. people for whom dementia was diagnosed after prostate cancer. This could be explained both by the biological relationship between the two diseases and by differences in ways of diagnosing them. In particular, dementia tends to evolve over many years and becomes most severe at old age, whereas cancers commonly arise earlier in the course of life. This mechanism alone would translate into the height of histogram bars increasing from left to right. On the other hand, the empirical pattern could also reflect the fact that cancers are often detected early due to routine screening procedures or that they are severe enough to substantially increase one's frequency of interactions with the health care system, which could lead to more subsequent diagnoses. In fact, increased health monitoring during a health shock is the likely explanation for a substantial fraction of diagnoses happening within one quarter of each other (i.e., the first bar on either side of the red line tends to be much bigger than the one further out). Thus, even the presence of strong medical evidence of *biological* independence of some condition and dementia would not allow me to rely on independence of their diagnoses. Therefore, I replicate my analysis for a variety of chronic conditions and investigate whether they depend on the sign of the observed relationship with dementia diagnosis.

In my main analysis I focus on acute myocardial infraction (AMI), commonly known as heart attack, as the primary condition of interest. It has been extensively studied in the medical literature due to its high cost and prevalence (800 thousand cases per year in the US), whereas Cutler et al. (2019) document substantial regional variation in medical expenditures attributable to its treatment. More importantly, AMIs acute nature makes it hard to remain unnoticed, while studies show that only 1-2% of cases are misdiagnosed, mostly due to incorrect interpretations of diagnostic tests (Kwok, Bennett, Azam, Welsh, Potluri, Loke, and Mallen, 2021), which alleviates the concern of its diagnosis depending on one's dementia status. Nevertheless, the histogram in Figure 1.2 corresponding to AMI shows that it is diagnosed much more often after dementia. Reassuringly, there is no discontinuous change in the probability of AMI diagnosis around the time of dementia onset (i.e., AMI diagnoses one quarter before and one quarter after dementia are similarly likely) and the asymmetric distribution of diagnosis time could be explained by heart attack risk increasing with age. When it comes to the biological relationship between these two diseases, unlike other chronic conditions (e.g., hypertension), dementia is not considered a risk factor for AMI (Roe, Messenger, Weintraub, Cannon, Fonarow, Dai, Chen, Klein, Masoudi, McKay, Hewitt, Brindis, Peterson, and Rumsfeld, 2010) while no consistent conclusion emerges from the scarce literature investigating the effect of dementia on risk of heart attack (Ng, Turek, and Hakim, 2013).

In addition to the aforementioned chronic conditions, I define six acute ones based on the ICD code directory: leg fracture, hip fracture, fall, traffic accident, pneumonia and influenza. The primary reason for investigating the effect of dementia on spending around the time of such health shocks is the fact that they are not, except for hip fracture, considered chronic and therefore are unlikely to affect reimbursement for physician services in a systematic way. Moreover, the consequences of such health shocks on one's health are likely transitory and presumably allow for less variation in terms of treatment decisions, so comparing the estimates for them to those for chronic conditions has the potential to shed light on the mechanisms at play. Finally, such injuries are an important concern in the management of dementia patients – e.g., fall and trauma are the primary reason for 26% of their hospitalizations (Alzheimer's Association, 2021). In Figure 1.A.1 in the Appendix, I show that all of these health shocks (except for traffic accident) are indeed more common after a dementia diagnosis. In this case dementia is likely *causing* these injuries, so the corresponding estimates should be interpreted with caution. It is also worth bearing in mind that such injuries are likely to happen multiple times over one's lifetime, whereas I only consider their first onset for each individual.

1.2.4 “Dementia first” and “condition first” groups

My empirical design features three groups: “dementia first”, “condition first” and “pure control”, which includes those who had dementia diagnosed but not the other condition (I refer to people from each of these groups using the respective group names). Therefore, before proceeding to my identification strategy I characterize the way these groups differ from each other when it comes to health care utilization and prevalence of chronic conditions. I focus on AMI as the other condition, so for “dementia first” and “AMI first” I investigate summary statistics constrained to quarters preceding the heart attack by at least two quarters in order to allow for it to affect utilization one quarter before it is officially diagnosed. Importantly, I expect “dementia first” to be on average older than “AMI first” due to dementia risk increasing sharply with age that is not observed in my data. Moreover, the most extreme cases of heart attack will naturally not appear in the “AMI first” group if the person in question dies within two quarters of AMI, as I omit people who had the two diagnoses within two quarters of each other.

Table 1.2 shows selected summary statistics by group assignment. In Panel A we see that “dementia first” patients have more frequent claims but spend less per day with claim, while spending significantly less on prescription drugs than “AMI first”. When it comes to medical services, “no AMI” people fall in between “AMI first” and “dementia first” in terms of cost, but this is driven by care intensity (i.e., higher cost per day and lower number of days with claims). These patterns are consistent with “dementia first” patients being the oldest and sickest and therefore requiring more frequent but not necessarily more invasive medical

Table 1.2: Summary statistics by relative time of dementia and AMI onset

	No AMI (1)	AMI first (2)	Dem. first (3)	Diff. (4)	
A. Overall health care use in the reference period					
Medical services: $	1980.4	1700.2	2048.3	348.1	*
Medical services: $/day	201.5	186.4	162.4	−24.0	*
Medical services: days	6.8	6.9	8.9	2.0	***
Prescription drugs: $	294.1	414.0	321.2	−92.8	**
Prescription drugs: $/day	170.7	195.4	161.8	−33.6	**
Prescription drugs: days	1.8	2.1	2.1	−0.1	
B. Average quarterly health care use in pre-period					
Medical services: $	1982.4	1435.1	1875.5	440.4	***
Medical services: $/day	213.9	184.3	173.0	−11.3	
Medical services: days	6.8	6.4	8.2	1.8	***
Prescription drugs: $	294.3	376.0	332.8	−43.3	
Prescription drugs: $/day	165.3	178.0	162.8	−15.2	
Prescription drugs: days	1.8	2.2	2.1	0.0	
C. Health care cost by category in the reference period ($)					
Ambulance transportation	49.3	34.3	68.0	33.7	***
Evaluation and management	242.6	211.8	254.6	42.8	***
Inpatient claims	888.7	542.2	840.5	298.3	**
Nursing home claims	139.5	61.1	223.5	162.4	***
Outpatient surgery	149.5	191.6	117.3	−74.3	**
Pathology and lab	52.7	51.9	47.5	−4.4	
Radiology	110.6	75.7	94.4	18.7	
D. Pre-existing heart conditions diagnosed (%)					
Atrial fibrillation	17.9	16.5	24.8	8.3	***
Heart failure	21.7	22.3	40.5	18.1	***
Ischemic heart disease	34.4	48.2	56.2	8.0	***
Hypertension	74.1	79.1	89.3	10.2	***
Stroke/transient ischemic attack	0.9	0.4	2.2	1.8	***
E. Other characteristics (%)					
Female?	57.8	46.8	51.2	4.4	***
Insurance plan: HMO	4.1	4.6	3.4	−1.3	***
Insurance plan: IND	56.1	61.6	60.0	−1.6	
Insurance plan: POS	15.7	11.7	14.9	3.3	***
Insurance plan: PPO	23.8	21.9	21.5	−0.5	

Notes. Reference period is two quarters before AMI diagnosis. Cost per day ($/day) is calculated only among those who had non-zero cost. Column 4 is the difference between columns 3 and 2. For days and cost per day measures by spending category as well as a full set of pre-existing conditions see Appendix Table 1.A.1.

interventions on average[6]. Panel B shows that some of these patterns lose significance if we average health care use over the entire pre-period: “dementia first” spend more than “AMI first” on medical services due to higher care frequency, but are not significantly different in terms of prescription drug use. Panel C demonstrates that these differences are driven primarily by invasive care such as inpatient, ambulance and nursing home claims. On the other hand, “dementia first” spend less on outpatient surgeries than “AMI first” and do not differ in terms of spending on diagnostic care (i.e., pathology and lab, radiology).

When it comes to chronic conditions, “dementia first” patients are more likely to suffer from all kinds of heart-related pre-existing conditions[7] (Panel D). This could be due to age, as they had more time to accumulate diagnoses, as well as higher health monitoring following a dementia diagnosis. However, Ng et al. (2013) point out that, unlike heart attack, the conditions from Panel D have been consistently linked to the risk of dementia, a link that is thought to be causal. Moreover, both “dementia first” and “AMI first” people are more likely to suffer from heart failure, ischemic heart disease and hypertension than the “no AMI” group, which points to the possibility of selection bias as those who experience a heart attack are likely more predisposed to having heart issues in general. Finally, Panel E shows that “dementia first” patients enroll in somewhat different insurance plans and are more likely to be female. In particular, they are less likely to enroll in health maintenance organization (HMO) plans, but more likely to enroll in point of service (POS) plans, both of

---

[6] In Appendix Table 1.A.2 I show that higher frequency of care for “dementia first” is somewhat smaller but highly significant for other heath shocks as well, while differences in cost per day of care are consistently insignificant. Curiously, the “no condition” group is often on average more costly than either “dementia first” or “condition first” groups.

[7] They are, in fact, significantly more likely to suffer from any kind of chronic condition, as demonstrated by Appendix Table 1.A.1.

which are generally more restrictive when it comes to provider choice than the other plans. Although statistically significant, these differences are small in magnitude as HMO and POS plans account for less than 20% of enrollment among dementia patients.

The aforementioned possibility of unobservable differences in health between the "dementia first" and "AMI first" groups, corroborated by the observed differences in health care utilization before the heart attack casts doubt on the unconfoundedness assumption that underlies my empirical strategy. As discussed by Imbens (2015), depending on the functional form of conditional expectations, different covariate distributions for the treated ("dementia first") and control ("AMI first") groups can bias least squares estimates by violating unconfoundedness. This issue can be alleviated by inverse propensity score weighting, that consists of two steps: estimating the probability of being in the treated group based on observable characteristics and using the inverse of each observations' probability of belonging to its group as weights in the regression (Abadie and Cattaneo, 2018). I use patients' health histories to estimate their propensity to have dementia before heart attack. This leads to a dimensionality problem as the number of possible procedures and diagnoses that one could have had in her claims is larger than the number of observations, and even restricting attention to the most common ones would lead to a substantial loss of degrees of freedom if I were to account for interactions between them. Therefore, I use tree-based ensemble models for estimating the propensity score[8] (see section 1.B of the Appendix for details).

In Table 1.3 I show selected summary statistics depending on the model used for esti-

---

[8] Lee, Lessler, and Stuart (2010) show that in certain settings machine learning models substantially outperform traditional methods like logistic regression when it comes to estimating the propensity score. Moreover, this is fundamentally a prediction exercise, which is what machine learning models are uniquely suitable for (Mullainathan and Spiess, 2017).

Table 1.3: Summary statistics by relative time of dementia and propensity score weights

	No weights			Random forest			Gradient boosting		
	AF	DF	Diff.	AF	DF	Diff.	AF	DF	Diff.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. Utilization of medical services									
$	1700	2048	348 *	1713	2046	333 *	1713	2046	333 *
$/day	186	162	−24 *	187	162	−25 *	187	162	−25 *
Days	6.9	8.9	2.0 ***	7.0	8.9	2.0 ***	7.0	8.9	2.0 ***
B. Utilization of prescription drugs									
$	414	321	−93 **	411	320	−91 **	411	320	−91 **
$/day	195	162	−34 **	196	161	−35 **	196	161	−35 **
Days	2.1	2.1	−0.1	2.1	2.1	0.0	2.1	2.1	0.0
C. Pre-existing heart conditions diagnosed (%)									
Atrial fibrillation	16.5	24.8	8.3 ***	16.5	24.9	8.4 ***	16.5	24.9	8.4 ***
Heart failure	22.3	40.5	18.1 ***	22.4	40.4	18.0 ***	22.4	40.4	18.0 ***
Hypertension	79.1	89.3	10.2 ***	79.0	89.3	10.3 ***	79.0	89.3	10.3 ***
Ischemic heart disease	48.2	56.2	8.0 ***	48.3	56.1	7.8 ***	48.3	56.1	7.8 ***
Stroke	0.4	2.2	1.8 ***	0.4	2.2	1.8 ***	0.4	2.2	1.8 ***

Notes. Columns correspond to different propensity score weights: no weights for columns 1-3, random forest weights for columns 4-6 and gradient boosting weights for columns 7-9. AF is "AMI first", DF is "Dementia first". Columns 3, 6 and 9 show the difference between the DF and AF groups together with indicators for statistical significance.

mating the propensity score. Clearly, differences between the "AMI first" and "Dementia first" groups change slightly depending on the weighting scheme used, but their significance remains the same[9]. For example, in panel A we can see that the difference in average spending on medical services in the reference quarter is smaller in either of the weighted samples and that this change is driven by higher weights assigned to patients from the "AMI first" group who have higher spending. However, most of the other summary statistics do not change markedly upon weighting – for example differences in chronic condition prevalence shown in panel C shrink by 0.2 percentage points at most. Importantly, inverse propensity

---

[9] As noted by Imbens (2015), it is not significance but magnitude of the differences in covariate distributions that matters for the unconfoundedness assumption.

score weighting is meant to bridge the gap between the distribution of claims information that is predictive of one's relative dementia timing, and not of all health-related characteristics. In particular, the algorithm relies on detailed claims information[10] instead of aggregate utilization or chronic condition designation, so eliminating differences when it comes to the distribution of these characteristics is not necessary. Finally, I will show that my main results are robust to using either of the weighting schemes considered, and will include individual and time fixed effects in order to account for the inherent differences between the "AMI first" and "dementia first" groups.

1.2.5 Empirical strategy

In order to identify the causal effect of having dementia at the onset of a heart attack I rely on a two-way fixed effects difference-in-differences design. More specifically, I account for time-invariant individual characteristics and time trends common to all individuals by including individual and calendar-time fixed effects, respectively. In addition, I include relative-time-to-event dummies with heart attack as the event, which reflect changes in the outcome variable around the time of event with respect to the reference period. Finally, I also include the set of relative-time dummies interacted with an indicator for having dementia at the time of event that capture the treatment effects of interest, i.e., differences in the trajectory of changes between the "dementia first" and "AMI first" groups. This leads to

---

[10] Using a machine learning algorithm with detailed diagnosis and procedure codes allows the predictive model to infer complicated patterns in the data that can potentially capture unobservables such as age (e.g., mammograms become routine at a certain age and their use could therefore be predictive of one's relative time between dementia and heart attack).

the following regression equation[11]:

$$\text{Health care use}_{it} = \alpha_i + \tau_t + \sum_{r \neq -2} (\beta_r + \gamma_r \times \text{Dementia first}_i) + \varepsilon_{it}, \tag{1.1}$$

where i indexes individuals, t calendar time (year-quarters) and r corresponds to the relative time to event with $r = -2$ as the reference period. Thus, I allow for one period of anticipatory behavior, so that $r < -2$ corresponds to the pre-period and $r > -2$ to the treatment period. Importantly, in order to be able to identify both calendar-time and relative-time fixed effects, I also include people who were diagnosed with dementia but not heart attack, which I refer to as "pure" control group. For all of the chronic conditions considered the pure control group is overwhelmingly larger than any of the other groups, so to ease computations I use a 20% sample[12]. Moreover, as time-to-event is not defined for the pure control group, I assign them to the reference period $r = -2$. Thus, their utilization patterns contribute to the estimation of calendar-time but not relative-time fixed effects[13], capturing health care utilization time trends common to people with dementia.

Depending on the context, the dependent variable measures health care cost of a particular category of services, the number of days with claims from this category or the ratio of the two (i.e., cost per day). Standard errors are always clustered at the individual level,

---

[11] Estimates reported in the paper come from separate regressions for "condition first" and "dementia first" groups instead of a regression with an interaction term as shown here. This leads to minor differences in the estimated time fixed effects that are negligible due to the fact that the pure control group dominates these estimates anyway. Upon re-accessing the data enclave for final revisions I will update the reported results to come from the regression as presented.

[12] The 20% sample of the pure control group is still substantially larger than either the "dementia first" or "condition first" group, regardless of the condition considered. For the heart attack specifically, it is more than three times larger.

[13] Naturally, individual fixed effects for the pure control group are also estimated, but this does not affect the estimates of relative-time coefficients which rely on within-individual variation.

following other papers in the literature that use the two-way fixed effects event-study design in the context of individual health care utilization (e.g., Dobkin, Finkelstein, Kluender, and Notowidigdo, 2018). Importantly, as individuals die, switch insurers or get diagnosed with heart attack later in my time horizon, the panel is unbalanced. In order to shed light on the extent to which my results are driven by changes in the composition of my sample, I also report results restricted to the balanced sample as well as estimates of attrition by using an indicator for being present in the data as the outcome variable. Another crucial robustness check is meant to address concerns about whether the pure control group offers a valid comparison for people suffering from a heart attack, as they could differ in health care utilization trends due to superior cardiovascular health. Thus, I also consider a version of equation (1.1) without the pure control group, in which I include 5-digit zip code instead of individual fixed effects. I compare these estimates to the respective ones from the main specification to determine the extent to which my results are driven by using the pure control group to estimate time fixed effects.

My design follows Sun and Abraham (2020) in relying on the following three identifying assumptions: parallel trends in baseline outcomes, no anticipatory behavior before the reference period and treatment effect homogeneity across different heart attack onset times. Crucially, unlike in the standard two-way fixed effects event study design, my parameters of interest are γ_r instead of β_r, as I focus on the difference in event study trajectories depending on one's dementia status at the time of heart attack onset. Thus, in this case the three standard assumptions are formulated in terms of differences between the "dementia first" and "condition first" groups, and it is the relative timing of dementia with respect to the other condition that should be as good as randomly assigned, not the time of condition onset. In particular, I assume that the difference in health care utilization between "dementia first" and "condition first" people would have been constant under the counterfactual scenario of "dementia first" being diagnosed with dementia after the condition. Moreover, I assume that the relative time of dementia diagnosis is not correlated with one's anticipatory behavior, so that if there is any anticipation beyond relative time $r = -1$ that I account for explicitly, it is the same for both "dementia first" and "condition first" groups. Finally, although my design explicitly allows for differences in event-study trajectories between the "dementia first" and "condition first" groups, I assume that these trajectories are homogenous within group and do not depend on the calendar time of heart attack onset.

Naturally, none of these assumptions can be verified directly, but I provide empirical evidence to support them: I test parallel trends and lack of anticipation in observed outcomes, and investigate treatment effect heterogeneity. Conceptually, the parallel trends assumption could be violated if dementia diagnosis changes one's health care utilization trend. This could be the case if this diagnosis has a persistent effect on health care use, for example due to increased health monitoring or through higher reimbursements for services provided to those with a history of dementia. Empirically, as shown in section 1.2.2, I find some evidence of a trend break when it comes to the use of medical services that seems driven by deaths, which is why restricting attention to survivors via the use of a balanced panel is an important robustness check in my analysis. Moreover, as dementia is a continuously evolving disease, it is likely constantly affecting health care utilization rather than introducing a trend break at the time of diagnosis. When it comes to anticipation, a major concern is that the possible increased monitoring of the health of "dementia first" patients allows them to address deteriorating heart health sooner, but it is unlikely to extend beyond the two quarters of anticipation that I allow for. Also, heart attacks are inherently unpredictable, which is one of the reasons for focusing on them; the possibility of violating this assumption is a bigger concern for other conditions and leads to my caution in the interpretation of their respective estimates.

A final caveat of my analysis stems from the use of a two-way fixed effects design that has been shown in recent literature to produce biased estimates. Among others, Borusyak, Jaravel, and Spiess (2022) show that event-study estimates in such settings consist of weighted averages of treatment effects with possibly negative weights and some of the underlying treatment effects coming from so-called "forbidden comparisons". This problem arises because units that have already been treated (i.e., after their heart attack) are essentially used as a control group for the not-yet-treated, which is an undesirable comparison unless a particularly strong version of the treatment effect homogeneity assumption holds. Crucially, the authors note that a relatively large number of never-treated units alleviates this issue, as they dominate the effective control group. In fact, alternative estimators proposed in the literature that address this explicitly remove these forbidden comparisons when calculating treatment effects (Callaway and Sant'Anna, 2021). In this paper the size of my pure control group dwarfs the size of both "dementia first" and "condition first" for any condition considered, which eases the concern of a possible bias of the estimator, as the fraction of comparisons that is "forbidden" is small, although non-zero.

1.2.6 Mechanisms

The change in health care use due to having dementia at the onset of another disease could arise due to five potential mechanisms: patient bargaining power, physician financial incentives, treatment adherence, interference of other decision-makers or change in preferences for care intensity. Below, I discuss each of these mechanisms in turn to show that, depending on their relative importance, the combined effect of having dementia at the onset of another condition on subsequent treatment intensity is ambiguous.

Decisions concerning medical treatment could be thought of as an agency problem in which the physician decides on behalf of her patient, in order to maximize a combination of patient utility and her own revenue function, which depends on the kind of treatment prescribed. Consider the following stylized example: the patient is diagnosed with prostate cancer which could be treated by surgery or put under active surveillance until further symptoms develop, as such cancers often grow slowly and might result in no further harm if left untreated. In this case, surgery is an aggressive form of treatment that brings revenue to the physician[14], while active surveillance is non-invasive and cheap. From the perspective of patient utility the health outcome is of primary importance, but for the same health benefit she presumably prefers less aggressive treatment due to its lower price, limited side-effects and lack of burdensome surgery. In the end, while deciding on which treatment to recommend the physician will consider potential health gains, revenue and patient utility associated with either of the treatment options.

---

[14] Of course, the physician who diagnoses the disease does not necessarily perform the surgery, so her financial stake in choosing the aggressive treatment is a major simplification. It is indeed possible, that she benefits financially from active surveillance that she would perform at follow-up visits, while referring the patient to a surgeon brings her no benefit.

The revenue consideration incentivises physicians to choose treatments that are more aggressive than what patients would have chosen for themselves[15]. The extent to which physicians heed their patients' preferences (e.g., inquiries for cheaper treatment due to budget constraints) could be lower for dementia patients as they lose the ability to advocate for themselves due to memory issues and confusion. If this is the case, and patients prefer less treatment than physicians on average, then we would expect higher treatment intensity among those who already have dementia. Moreover, dementia diagnosis could itself affect physicians' financial incentives as patients with pre-existing conditions often qualify for higher reimbursement rates. On the other hand, such additional pay might be offset by higher effort required to manage a patient with dementia. In fact, physicians could prescribe less invasive treatment to dementia patients in order to avoid this effort, leading to a negative effect on health care utilization.

Dementia can also have various consequences for patient utility as a function of prescribed treatment. For example, it can hinder patients' ability to follow doctor's orders, as forgetful patients skip medications or follow-up visits[16], resulting in lower health care use. However, if physicians anticipate issues with treatment adherence among patients with dementia they might prescribe treatment that requires less systematic care, which could be more invasive and costly such as the surgery for prostate cancer. At the same time, dementia might affect

---

[15] In practice, such tendency for excessive treatment could also be caused by fear of litigation, which is referred to as "defensive medicine" in the literature, following the idea that physicians do as many procedures and tests as possible in order to minimize the risk of overlooking something.

[16] Following one's treatment plan often requires additional decisions, e.g., which prescriptions to fill under a binding budget constraint. Chandra, Flack, and Obermeyer (2021) show that Medicare beneficiaries often end up not filling any of their prescriptions when faced with drug price increases, presumably because they cannot easily rank their prescription importance. Confusion caused by dementia could exacerbate this problem, leading to less medication adherence.

patient preferences for care intensity in either direction: staying at a hospital might be less of an issue for a person that stays at a nursing home anyway, but it could also lead to disproportionate discomfort as the person does not know what and why is happening to her. Thus, any effect of dementia on health care use could be explained by changes in preferences that are not observed, but we expect such changes to be continuous along with dementia progression rather than discrete at the time of its diagnosis.

Finally, other decision-makers such as family members or nursing home staff could participate in this decision-making process, as people with advanced dementia often receive assistance from a full-time caregiver. Such caregivers could steer the patient towards more invasive care as they do not bear the full costs of it (e.g., side effects) and could value more the quantity rather than quality of life of their life (e.g., having the loved one around for longer). On the other hand, care that requires active management such as active surveillance for prostate cancer, could be particularly burdensome for caregivers as it costs them effort and time to arrange. Finally, caregivers might have a financial stake in the management of care of the person they care for. In particular, the prospect of inheritance could incentivise a caregiver to advocate for cheaper treatment options in order to increase the value of their inheritance. Unfortunately, my data does not provide information on caregivers so I cannot speak to this mechanism directly, but based on the above discussion I expect it to lower health care utilization of people with dementia as the cost of effort and financial considerations dominate the remaining effects.

1.3 Estimates for heart attack

In this section I show estimates of the effect of dementia on health care spending following a heart attack. Section 1.3.1 discusses estimates for the cost of medical services and shows that the negative treatment effects are driven by frequency of care (i.e., extensive margin) and are robust to the choice of propensity score model used for weighting. In section 1.3.2 I use heterogeneity by time of dementia diagnosis to show that at most 50% of the estimated treatment effect can be attributed to the treatment of dementia. Further, section 1.3.3 establishes that decreases in care frequency due to dementia are only observed for non-invasive care such as pathology and lab, whereas the treatment effect is in fact positive for inpatient care. I consider the use of prescription drugs separately in section 1.3.4, and document striking decreases that support the hypothesis of treatment adherence as a crucial mechanism behind my results. Finally, in section 1.3.5 I investigate the role of attrition due to death or insurance switches, whereas section 1.3.6 concerns treatment effect heterogeneity by patient characteristics.

1.3.1 Effect on cost of medical services

In Figure 1.3a I plot estimates of the event-study trajectory for “AMI first” (β_r), “dementia first” ($\beta_r + \gamma_r$), and their differences, which are treatment effects of interest γ_r. For both groups we see a substantial spike in spending on medical services at the time of the heart attack, which is $3070 smaller for “dementia first”. Moreover, “AMI first” experience a persistent increase in use of medical services following the heart attack, which is indicative of possible complications or routine follow-up care. Qualitatively, a similar pattern arises for those who never had dementia (see Appendix section 1.C), whereas the treatment effect for "dementia first" converges to zero two quarters after the heart attack. Thus, estimates of γ_r are significant[17] and negative for all $r > -1$, suggesting that having dementia at the onset leads to less invasive treatment of heart attack and little subsequent follow-up care. Of course, "dementia first" patients spend more on medical services in the first place, so it is possible that some of this care is substituted with heart attack-related follow-up care in the post period.

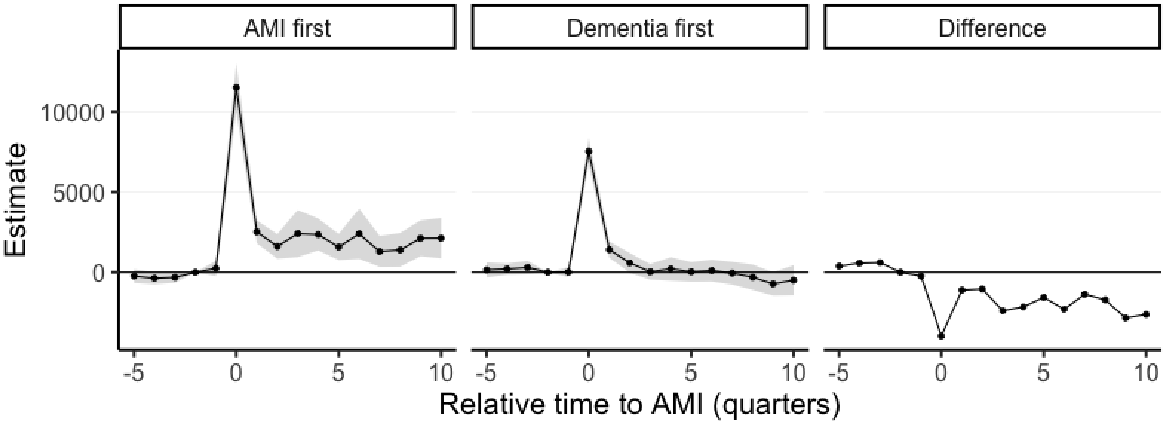
In order to show that the long-term difference in spending is driven by the extensive rather than intensive margin, Figures 1.3b and 1.3c show estimates for days with claims and cost per day with claims, respectively. Clearly, when it comes to immediate treatment of heart attack (i.e. $r = 0$), the increase in both the number of days and cost per day is smaller for "dementia first", with a more substantial difference for cost per day. However, in the long term cost per day is almost the same for both groups, indicating that conditional on receiving care both groups spend a similar amount per day of care. This rules out the possibility of "dementia first" patients undergoing low-intensity care management and "AMI first" patients getting expensive but infrequent medical interventions. Even more surprisingly, the persistent difference in care frequency shown in Figure 1.3b results in a reversal of the average care frequency difference between the two groups. To be more specific, before the heart attack "AMI first" had on average 7 days with claims per quarter, whereas "dementia first" had 2 more, but the treatment effect of -4 leads them to have 2 days less per quarter.

---

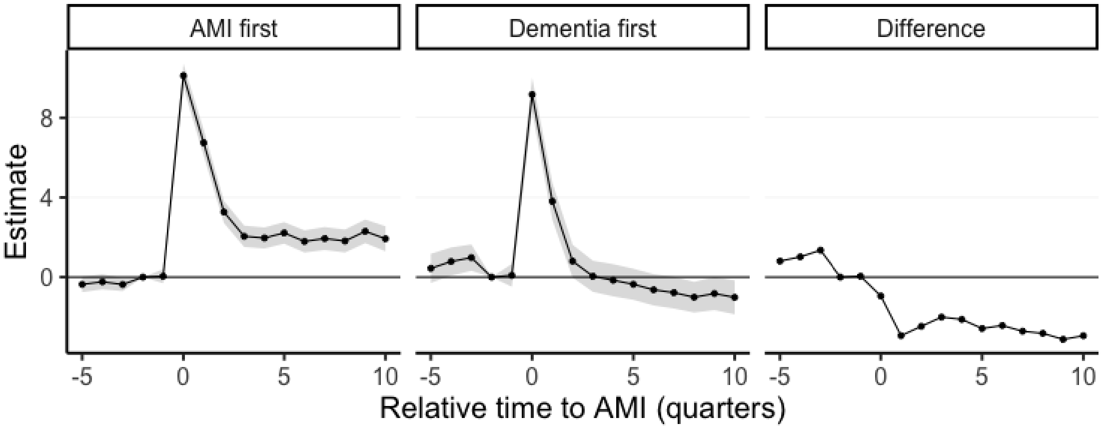
[17] Confidence intervals will be added when I re-access the data enclave for final revisions.

Figure 1.3: Event-study estimates for utilization of medical services

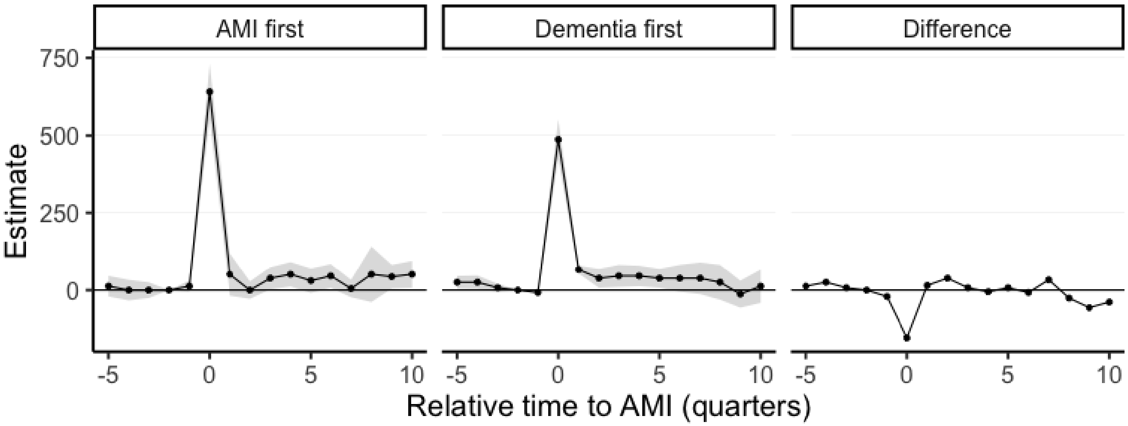
(a) Cost ($)



(b) Days with a claim



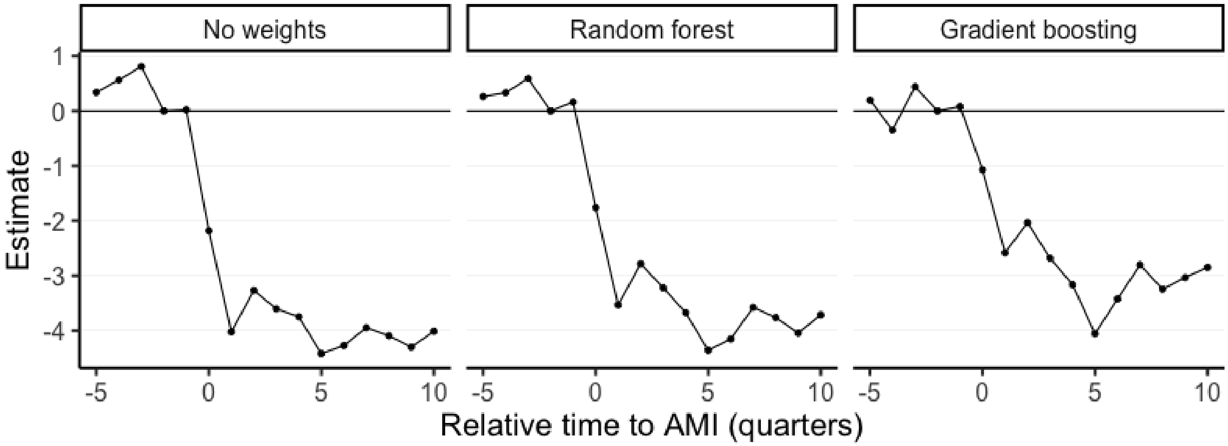
(c) Cost per day with a claim ($)



Notes. Panels “AMI first”, “Dementia first” and “Difference” show estimates of β_r, $\beta_r + \gamma_r$ and γ_r, respectively. Mean of dependent variable at $r = -2$ for “AMI first” and “Dementia first” is (a) 1700 and 2048, (b) 6.9 and 8.9, (c) 186.4 and 162.4, respectively.

This difference emerges a quarter after the heart attack ($r = 1$) and persists till the end of the time horizon of the plot.

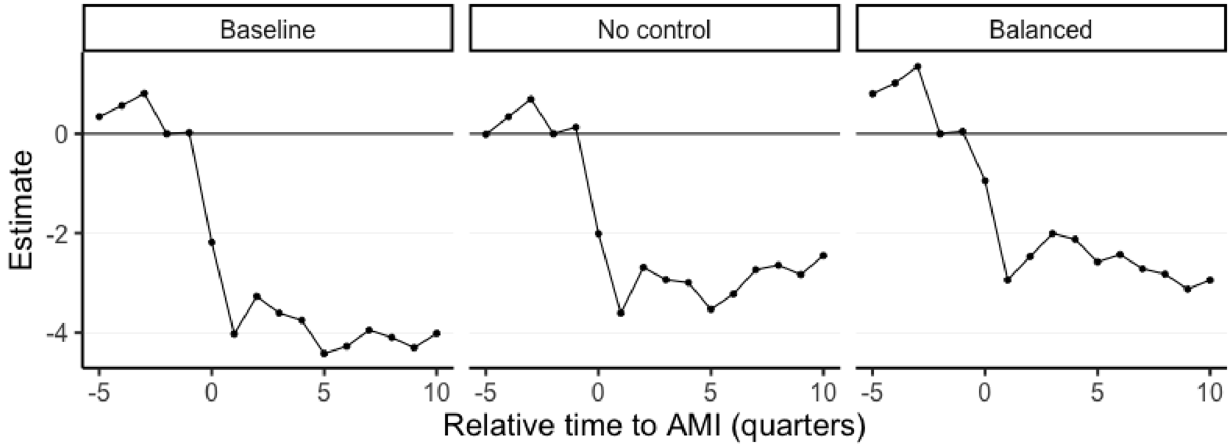
Figure 1.4: Estimated treatment effects for days of care by weighting model



Notes. Each pane shows estimates of γ_r from a regression of days with claims for medical services using a different set of regression weights.

As the effect of dementia on heart attack care is driven by care frequency measured by the number of days with claims per quarter, I will focus on this outcome for the remainder of this section. Figure 1.4 shows how inverse propensity score weighting affects these estimates: the small positive pre-trend from a regression with no weights becomes smaller once random forest weights are used, and disappears completely under a gradient boosting weighting scheme. The last set of weights also halves the estimated treatment effect for $r = 0$ (from -2 to -1 days of care), and reduces the long-term effect from -4 to -3 days of care two years after the heart attack. Thus, differences in covariate distributions between "dementia first" and "AMI first" groups seem to indeed bias my estimates, but correcting for the bias does not alter the resulting conclusions. When it comes to the baseline difference in frequency of care, weighting the sample changes it from 1.98 to 1.97, regardless of the type of weights, corroborating the above claim that heart attack leads to a change of sign of the difference in care frequency between the two groups.

Figure 1.5: Estimated treatment effects for days of care by specification



Notes. Each pane shows estimates of γ_r from an unweighted regression of days with claims for medical services. “No control” means that there is no pure control group and instead of individual fixed effects the regression includes 5-digit zip code fixed effects. “Balanced” means that the sample is restricted to individuals present in the panel throughout the relative time horizon.

In Figure 1.5 I plot the treatment effect trajectory for two alternative regression specifications together with the original one. Clearly, removing the pure control group attenuates my estimates, and it is not possible to tell whether that is caused by removing this group or by altering the regression specification as the “No control” specification includes 5-digit zip code instead of individual fixed effects due to the underidentification issue (Borusyak et al., 2022). In the “Balanced” pane I restrict attention to patients who were present in the panel throughout the relative time horizon shown on the plot, which removes the effect of attrition that can be caused by switching insurers or death. Clearly, this also attenuates the long-term (i.e., $r > 4$) estimate from -4 to -2.5, which means that the treatment effects are larger in absolute terms among those who died or switched insurers. This points to the possibility of reductions in care frequency following a heart attack resulting in increased mortality among dementia patients. Unfortunately, this cannot be investigated directly in the current dataset and thus remains speculative. Moreover, the balanced sample is highly selected, as "dementia first" must survive much longer following dementia than "AMI first" in order to be included in this sample, which naturally restricts attention to milder dementia cases. Finally, the balanced sample raises concerns with violations of the parallel trends assumption, and this is the case regardless of the weighting scheme applied.

1.3.2 Heterogeneity by time of dementia onset

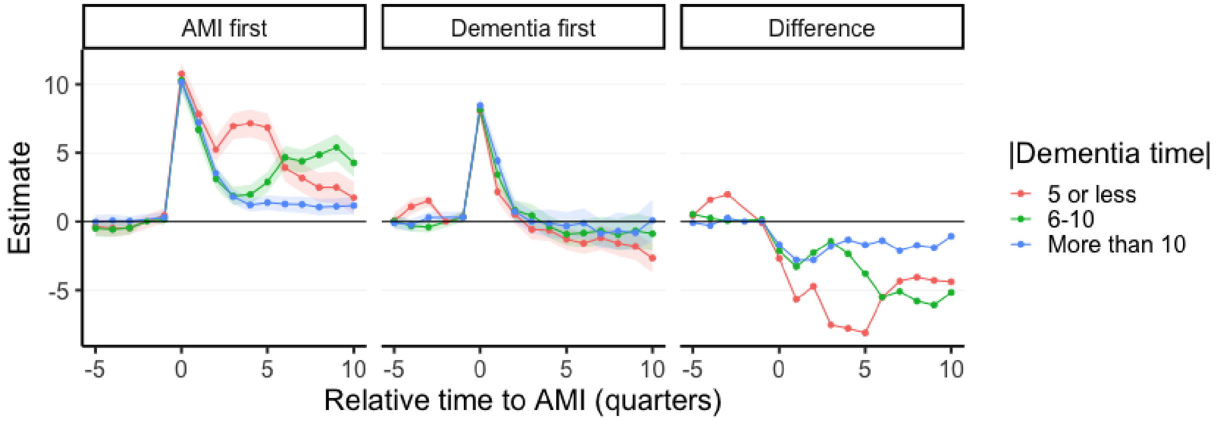
In this section I investigate the concern that some of the treatment effect might be driven by dementia care itself. In particular, by definition, "dementia first" and "AMI first" groups are diagnosed with dementia in the pre- and post-period, respectively, possibly within the time horizon of my event-study plots. Thus, the temporary increase in health care utilization associated with a dementia diagnosis could explain some of the observed differences in spending around the time of heart attack between the two groups. To be more specific, some of the "dementia first" are diagnosed with dementia at $-5 \leq r < -2$[18], which might explain the pre-trends observed in Figure 1.3b. Similarly, some of the health care utilization of "AMI first" for $r > 2$ will coincide with the time of dementia diagnosis. However, it is important to note that the pure control group also experiences such dementia-related increases in spending and is included in the analysis precisely for the purpose of differencing them out. Moreover, as documented in the literature (Deb et al., 2017), dementia is often diagnosed

---

[18] Dementia diagnoses at $|r| \leq 2$ are excluded from the analysis due to measurement error concerns.

during treatment of another condition, so health care spending around its diagnosis does not necessarily correspond to dementia care, and could be part of heart attack treatment (e.g., one could get diagnosed with dementia during a hospitalization due to post-AMI complications). Nevertheless, differences in spending that cannot be explained by utilization increases around dementia diagnosis are more unambiguously attributable to heart attack treatment and are thus crucial to my analysis.

Figure 1.6: Treatment effect trajectory by relative time of dementia diagnosis



Notes. Panes “AMI first”, “Dementia first” and “Difference” show estimates of β_r, $\beta_r + \gamma_r$ and γ_r, respectively. Colors correspond to absolute dementia time in quarters relative to heart attack.

In Figure 1.6 I show event-study estimates for both groups and the resulting differences grouped by the number of quarters between dementia and heart attack diagnosis. The group “5 or less” among “dementia first” had dementia diagnosed at $-5 \leq r < 2$ and, as expected, has positive and significant pre-trends. Importantly, for “dementia first” the “6-10” group’s event-study estimates are not significantly different from those for “more than 10”, which corroborates the assumption that dementia diagnosis does not introduce trend brakes in health care utilization. Moreover, as the trajectory of health care use after a heart attack does not depend on how long ago dementia was diagnosed it seems that dementia severity does not matter beyond the point of diagnosis[19]. A similar conclusion can be drawn for the "AMI first" group – the estimated trajectory differs between the three groups at the time of dementia diagnosis as spending increases temporarily at $3 \leq r \leq 5$ and $6 \leq r \leq 10$ for the "5 or less" and "6-10" groups, respectively. However, no significant differences appear for $r \leq 2$, before either of the three groups had dementia diagnosed. Finally, the difference in spending trajectories for the two "more than 10" groups can be interpreted as a lower bound, in absolute terms, for the treatment effect that is unrelated to dementia treatment, as people in that group receive a dementia diagnosis beyond the scope of the plot. Crucially, this lower bound is significant, negative, and stabilizes at around 50% of the overall treatment effect. Thus, out of the estimated long-term decrease of 4 days of care per quarter caused by having a dementia diagnosis, at least half can be attributed to treatment unrelated to dementia.

1.3.3 Effect on different spending components

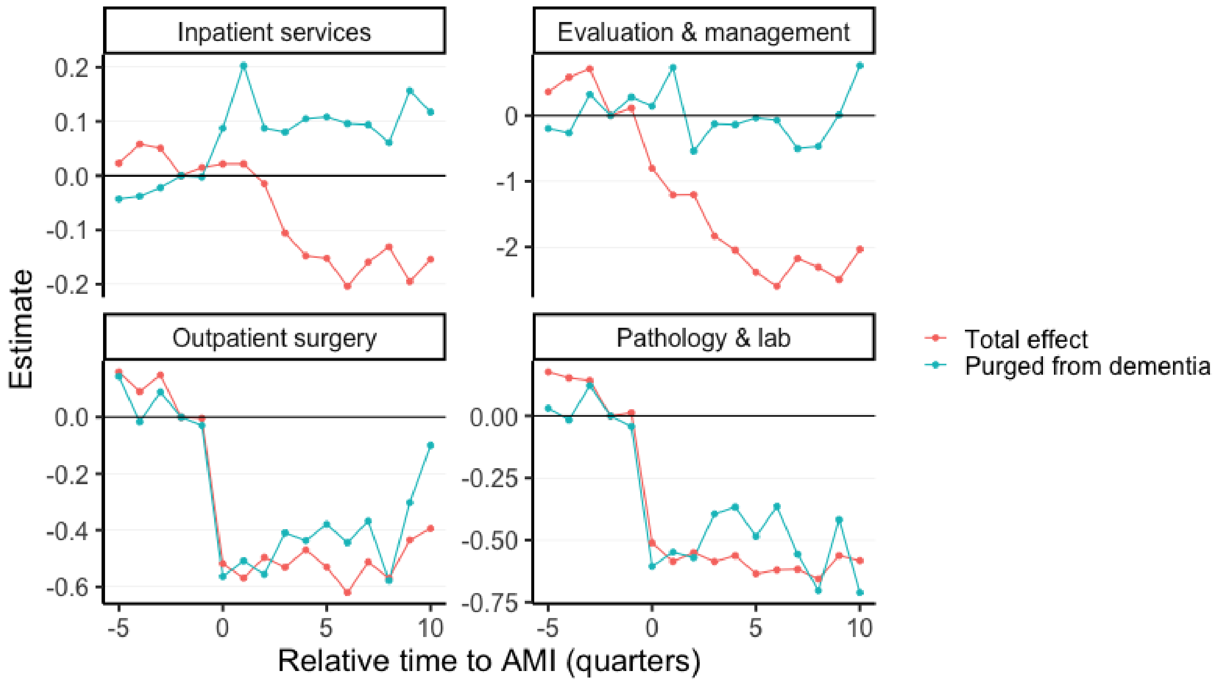
In this section I show that decreases in frequency of care caused by dementia are driven by outpatient care, such as pathology and lab (e.g., blood tests) or surgeries that do not require admission to the hospital. In Figure 1.7 I plot the overall estimated effect for various types of care together with the lower bound on the treatment effect that cannot be attributed to dementia treatment (i.e., estimates for the subgroup that had a dementia diagnosis at $|r| > 10$). The results are qualitatively similar for care that is typically associated with dementia:

---

[19] Admittedly, two countervailing forces are at play here: those who had dementia diagnosed further in the past are presumably on average more sick now, as the disease evolved over a longer period of time. However, the sample is restricted to those with a heart attack, so the "more than 10" group could be healthier as they by definition survived longer on average than those in the "6-10" group.

inpatient claims and evaluation and management (i.e., outpatient physician consultations). More specifically, for both of these types of care there is no sharp decrease at the onset of heart attack, and the effect becomes null or switches sign if we purge it from dementia treatment (blue line). Moreover, although long-term decreases of 0.2 and 2 days of inpatient and evaluation and management care, respectively, are relatively large (i.e., 66% and 40% of pre-AMI levels), the "dementia first" group utilizes much more of such care in the pre-AMI period (0.2 and 1.1 days more, respectively, see Panel B of Table 1.2), making the estimated negative treatment effects not as striking.

Figure 1.7: Effect on days with claims for specific types of care



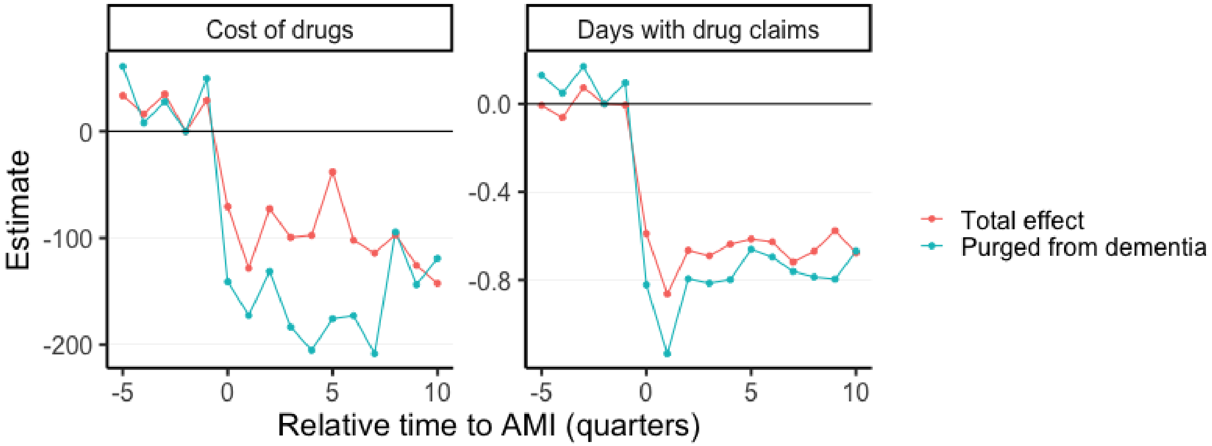
Notes. "Purged from dementia" refers to estimates for those whose dementia diagnosis is at least 10 quarters away from heart attack onset (i.e., this is the "more than 10" group from Figure 1.6).

The negative treatment effects are more pronounced and robust when it comes to outpatient surgeries and pathology and lab. In particular, these types of care are used substantially less at the heart attack onset and this change persists for at least two years afterwards. Moreover, the estimates are strikingly similar when purged from possible effect of dementia treatment, and amount to around 0.4 and 0.7 less days of claims for outpatient surgery and laboratory tests, respectively. In relative terms, these effects correspond to roughly 40% and 54% of pre-AMI levels, which are also initially similar for the "dementia first" and "AMI first" groups. Importantly, although patients have substantial discretion over both of these types of care, they are otherwise quite different: pathology and lab could be part of routine follow-up care after a heart attack, while outpatient surgeries are presumably unrelated. Moreover, reductions in outpatient surgery use might be a natural solution for a patient that is sick enough that elective procedures bring more cost than benefit. However, the estimated decreases concern care frequency, while effects on total cost are mostly insignificant (see Appendix Figure 1.D.1). At the same time, the observed increase in frequency of inpatient claims suggests that dementia might be causing substitution of non-invasive outpatient care into invasive inpatient stays. Decreased frequency of pathology and lab might also reflect less health monitoring that could result in avoidable hospitalizations.

1.3.4 Effect on prescription drugs

Up to now I focused on the use of medical services, which does not include prescription drugs. This distinction is motivated by important differences between these two categories of health-related spending. First, prescription drug claims record purchases instead of actual use of the underlying drugs, so changes in utilization could reflect purchase behavior that is irrelevant to a person's health. Second, not all insurance plans cover prescription drugs and every quarter only one third of Medicare beneficiaries in the HCCI data have a prescription drug claim, which suggests incomplete coverage. Third, dementia is negatively correlated with utilization of prescription drugs, while the opposite is true for medical services (see Table 1.2). Finally, prescription drug use could shed light on the treatment adherence mechanism, as it is unlikely that dementia patients receive markedly less prescriptions upon suffering a health shock, especially as certain medications can alleviate dementia symptoms. Thus, decreases in use are presumably indicative of changes in patient purchasing behavior, which might provide an opportunity for policy interventions.

Figure 1.8: Effect cost and frequency of prescription drug claims



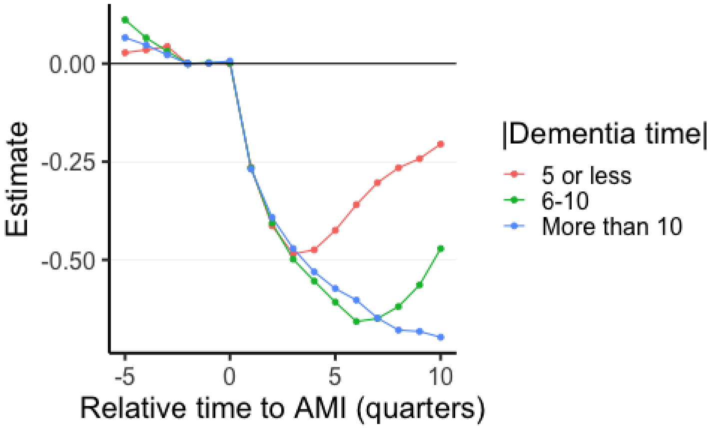
Notes. “Purged from dementia” refers to estimates for those whose dementia diagnosis is at least 10 quarters away from heart attack onset (i.e., this is the “more than 10” group from Figure 1.6). See the underlying event-study trajectories (i.e., β_r and $\beta_r + \gamma_r$) in Appendix Figure 1.D.2.

In Figure 1.8 I show estimates of the total effect as well as the lower bound that cannot be attributed to dementia treatment (i.e., estimates for the subgroup that had a dementia diagnosis at $|r| > 10$). Similarly to medical services, claim frequency drops sharply at the onset of heart attack, and the decrease persists for at least 2.5 years afterwards. However, unlike for medical services, prescription drugs also exhibit a substantial decrease in cost and the estimated effects are even bigger when restricted to the period of time before dementia onset of the “AMI first” group. When it comes to relative magnitude of the overall effects two years after the heart attack, both the 100$ decrease in drug cost and 0.65 decrease in the number of days with drug claims per quarter amount to 31% of the corresponding pre-AMI averages. Although slightly smaller in relative terms than the effects for outpatient medical services discussed in the previous section, these effects go in the same direction as pre-AMI differences between the “dementia first” and “AMI first” groups. To be more specific, before the heart attack the “dementia first” have 90$ less drug spending per quarter and this difference more than doubles within two years after the heart attack.

1.3.5 Estimates of attrition

In this section I investigate differences in attrition between the “dementia first” and “AMI first” groups. Although I cannot distinguish between attrition caused by switching insurers from that reflecting deaths, I consider the latter to be of primary concern in this context due to the fact that I study heart attacks among the elderly, which carry substantial mortality risk (e.g., 12% in the first 30 days according to Krumholz, Normand, and Wang, 2019). Moreover, using relative time of AMI and dementia diagnosis to define comparison groups in my analysis introduces mechanical differences in attrition that could partially drive the estimated treatment effects. In particular, “AMI first” survive at least 3 quarters after the heart attack, as they are, by definition, diagnosed with dementia at $r > 2$. Thus, the most deadly cases of AMI will not be included in “AMI first”, but they will be in “dementia first”, which could explain some of the estimated decreases in frequency of health care utilization, if non-invasive but frequent care management is not implemented for the most gravely ill patients[20]. At the same time, the most serious cases of dementia will not be included among the "dementia first", as being included in that group implies surviving long enough to experience a heart attack.

Figure 1.9: Differences in attrition between "dementia first" and "AMI first"



Notes. Estimates of γ_r from a regression with the indicator of being present in the panel.

In Figure 1.9 I plot the difference in attrition trajectories between the two groups by relative time to dementia. For example, the estimate of -0.25 at $r = 1$, uniform across the three plotted lines, implies that one quarter after heart attack "dementia first" are 25 percentage points less likely to remain in the panel than "AMI first". As "AMI first" are all by definition alive at this point in time this implies, barring meaningful differences in insurance switching, at most 25% mortality of the "dementia first" within a quarter of the heart attack, which is in the ballpark of unconditional AMI-related mortality estimates from the literature. Moreover, differences in attrition between the two groups become smaller

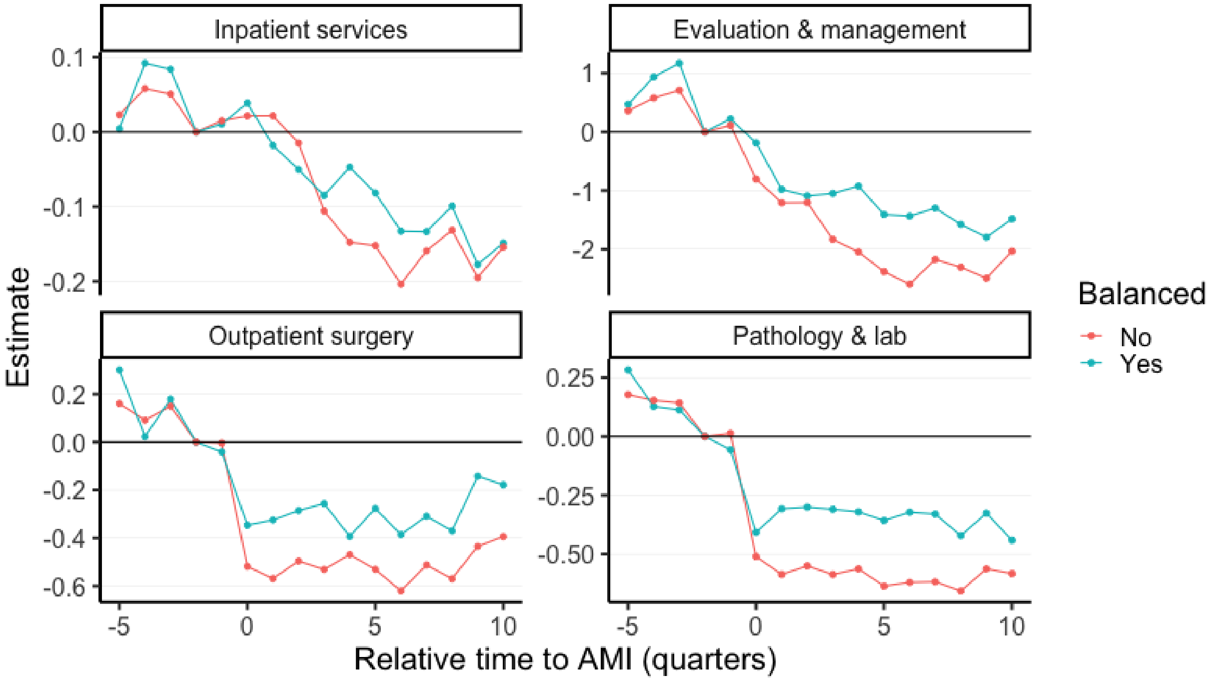
---

[20] Empirically, death is associated with sharp increases in health care utilization, so we would not expect deaths to explain decreases in health care utilization. Moreover, types of care associated with high quality of life for a dying person, such as palliative care, would presumably exhibit high frequency for the same cost as opposed to, say, hospitalizations.

once we relax restrictions on survival imposed on "AMI first" by focusing on those who were diagnosed with dementia sooner (e.g., red vs green line). However, substantial differences in attrition between "dementia first" and "AMI first" remain even if we restrict attention to a comparison between those who had dementia at most 5 quarters before/after the heart attack, which imposes the least survival restrictions for either of the group. Crucially, increased attrition of "dementia first" caused by either the aforementioned positive selection of heart attack cases among "AMI first" and "dementia first" being on average older and sicker do not constitute treatment effects of interest and should therefore be considered confounders in my analysis. On the other hand, other underlying reasons of differences in attrition are valid treatment effects – for example, if higher mortality of "dementia first" is *caused* by the reductions in heath care frequency documented in this paper. As it is impossible to distinguish between these two types of attrition (i.e., mechanical consequence of group definition vs treatment effect of dementia), I consider changes to my estimates when I restrict attention to a balanced panel.

In Figure 1.10 I show my main estimates for frequency of various kinds of care together with analogous estimates for the balanced panel. Clearly, the new estimates are smaller in absolute terms, suggesting that treatment effects are larger among those who died or switched insurers within 10 quarters of having a heart attack. Importantly, these changes are much more pronounced for care unrelated to dementia (i.e., outpatient surgery and laboratory tests), for which the balanced panel restriction leads to 30-50% smaller treatment effect estimates. At least three possible explanations arise. First, larger treatment effects on the unbalanced panel could be directly driven by differences in mortality, as outpatient

Figure 1.10: Estimate comparison between the balanced and unbalanced panels



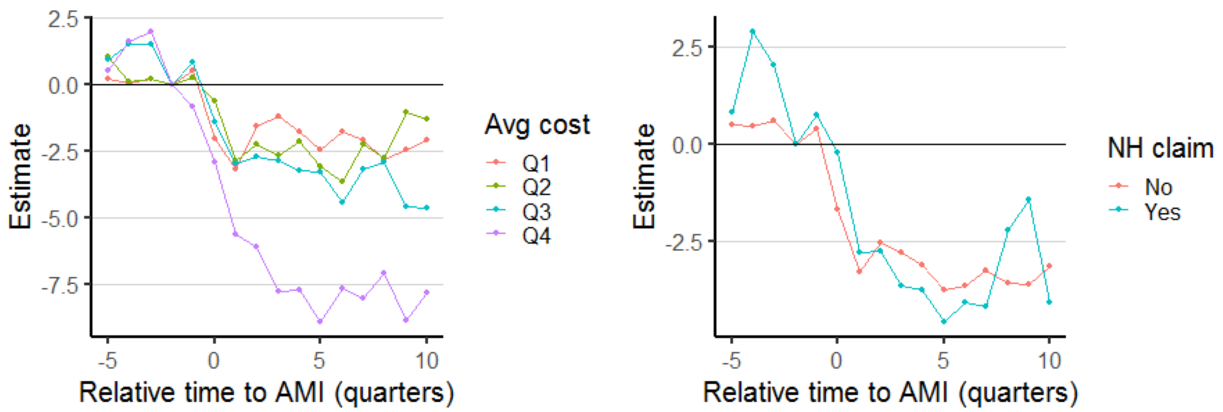
Notes. Estimates correspond to γ_r parameters with days of care of each type as the outcome variable. Balanced panel refers to a sample that is observed in the HCCI data for all $r = -5, \ldots, 10$ (this restriction does not apply to the pure control group for whom time-to-event is not observed).

services are presumably used less around the time of death. However, in this case we would also expect more utilization of invasive care around the time of death, while the direction of change when going from balanced to unbalanced panel is the same for inpatient services. Second, reductions in diagnostic testing and outpatient surgeries could be a rational response to the increased risk of mortality of the "dementia first" population and the estimates for the balanced panel bound the extent to which this could be the case. Finally, decreases of these kinds of care could be part of the cause of increased mortality for the unbalanced panel, in which case they are of primary interest to my analysis. Unfortunately, I am unable to distinguish between these three scenarios so the key takeaway from this section is that taking away the effect of attrition leads to similar, although smaller in magnitude, treatment effects.

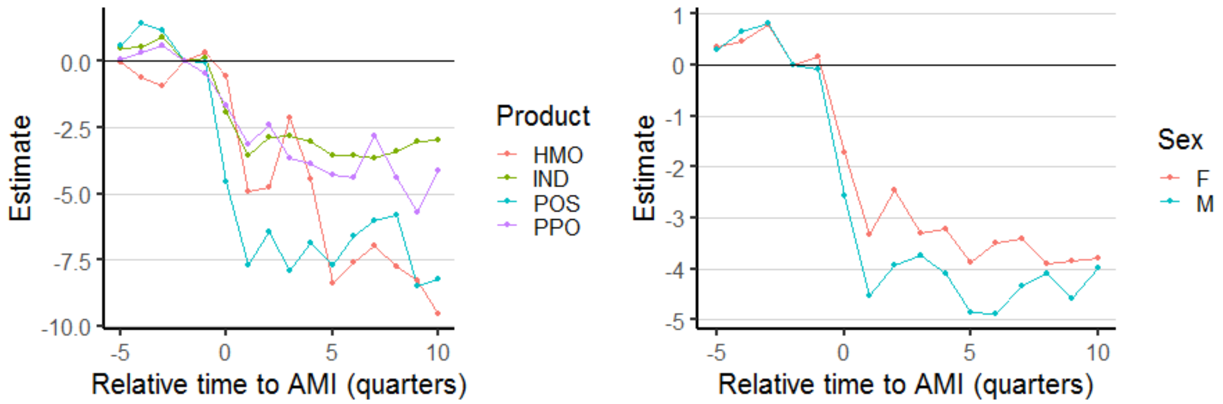
1.3.6 Treatment effect heterogeneity

Figure 1.11: Treatment effect heterogeneity

(a) By average pre-period cost quartile **(b)** By nursing home claim indicator



(c) By type of insurance **(d)** By sex



Notes. Each panel shows treatment effect estimates for the number of medical claim days by groups of beneficiaries based on a specific variable. Panel (a) uses quartiles of average cost of medical services in the pre-period (i.e., $r = -5, \dots, -2$), panel (b) uses an indicator for whether the patient had a nursing home claim in the pre-period, panel (c) splits by type of insurance at $r = -2$ (or at entry into the database for the pure control group), panel (d) splits by sex.

In Figure 1.11 I plot treatment effects by various heterogeneity dimensions. In panel (a) we can see that potential violations of the parallel trends assumption are driven by high-cost individuals. At the same time, estimate trajectories seem quite uniform for all but the highest-cost individuals, for whom care frequency decreases almost three times as much as for other patients a year after the heart attack. When it comes to possible nursing home status, determined by the presence of nursing home claims in the pre-period, panel (b) shows no consistent direction of heterogeneity, suggesting that nursing home agency is not the primary driver nor does it counteract the observed decreases in care frequency due to dementia. In panel (c) we see that those enrolled in "point of service" (POS) insurance plans have treatment effects that are twice as large as those for indemnity (IND) and PPO plans, whereas enrollees of HMOs fall somewhat in between. An important characteristic of both POS and HMO insurance plans is the fact that they restrict provider networks available to the patient, which makes access to care more complicated and can thus exacerbate issues with treatment adherence. Finally, in panel (d) I show that treatment effects are somewhat larger for men despite no gender differences in pre-period estimates. In addition, Figure 1.D.3 in the Appendix shows that attrition patterns are uniform across all of the group divisions discussed above, so none of the treatment effect heterogeneity shown in Figure 1.11 is driven by differences in mortality or insurance switching.

1.4 Estimates for other conditions

In this section I demonstrate how estimated treatment effects of dementia change when health shocks other than heart attack are considered. In section 1.4.1 I show the distribution of the main estimate of interest, i.e., treatment effect of dementia on care frequency, across health shocks and point to the fact that although the magnitude tends to be smaller than for heart attack, the sign remains the same, as dementia causes long-term decreases in care frequency following any of the conditions considered. Section 1.4.2 is devoted to using those who had the condition but not dementia (the "no dementia" group) as a control group in order to corroborate the finding for heart attack that at most 50% of my estimates can be explained by dementia treatment. Finally, in section 1.4.3 I present results concerning prescription drugs that reveal substantial negative treatment effects on both cost and purchase frequency across all conditions.

1.4.1 Distribution of estimates across conditions

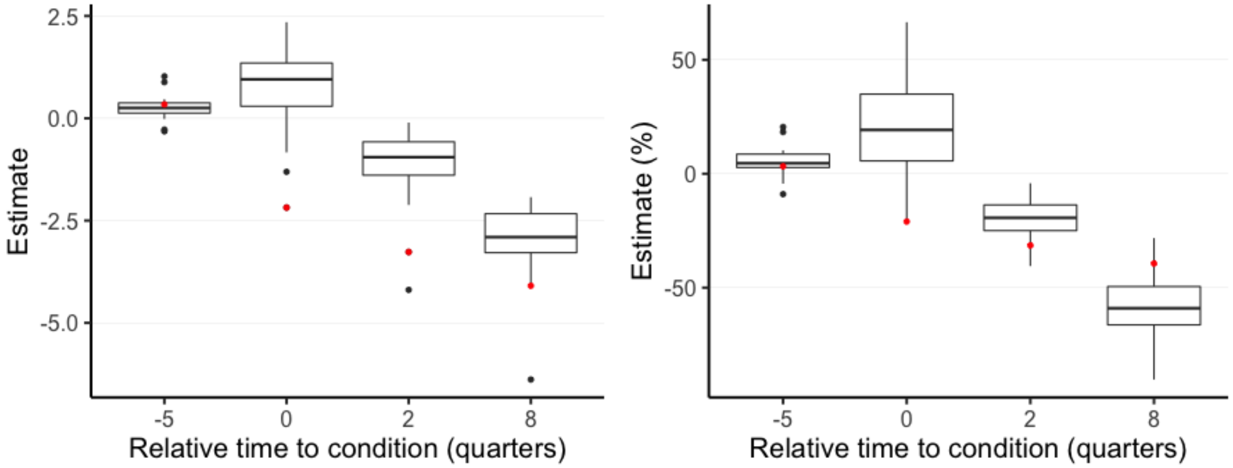
In Figure 1.12 I plot the distribution of estimated treatment effects on care frequency across 27 health shocks considered in my analysis[21] with heart attack estimates highlighted in red. For brevity, I focus on a subset of the relative time-to-event scale, that speaks to possible violations of the parallel trends assumption ($r = -5$), immediate effect ($r = 0$), effect after 2 quarters from the health shock ($r = 2$) and long-term effect ($r = 8$). In Figure 1.12a we see that the positive pre-trend for AMI at $r = -5$ can also be found for most other conditions, but it is generally small in magnitude. When it comes to the immediate effect, dementia leads to higher frequency of care at the time of the health shock for the vast majority of conditions, and AMI's negative estimate of almost -2.5 is a clear outlier. Two quarters after the health shock, which is by definition still before anyone from the "condition first"

---

[21] As discussed in section 1.2.3, my analysis includes 27 chronic conditions defined by CCW, of which I exclude dementia and Alzheimer's disease. I also exclude endometrial cancer due to insufficient sample size, as well as cataract and glaucoma that are defined only in terms of outpatient claims and are thus not of primary interest to my analysis. In addition, I define the following 6 health shocks based on diagnosis codes: fall, leg fracture, hip fracture, pneumonia, influenza and traffic accident. The CCW list of conditions also includes hip fracture, but I use my definition instead, which is slightly broader; the results are very similar for either definition.

group is diagnosed with dementia, all conditions exhibit a negative effect of dementia on care frequency. Importantly, the distribution of this effect is highly concentrated around -1, so that the effect of -3.3 for heart attack is again an outlier, exceeded in absolute terms only by lung cancer with an estimated effect of -4.2. However, the case of AMI is much more representative[22] if we consider the effect at two years after the health shock, when the entire distribution of estimates is below -1.9 with a median of -2.9 days of care[23].

Figure 1.12: Distribution of treatment effects for care frequency across conditions



(a) Raw estimate

(b) Estimate relative to β_0

Notes. Both panels show the distribution of estimates across 27 health shocks considered for $r \in \{-5, 0, 2, 8\}$ with days with medical claims as dependent variable (see Appendix Figure 1.E.1 for cost of medical care). Panel (a) uses raw estimates of γ_r, while panel (b) scales these estimates by the estimated increase in care frequency at the time of event for the "condition first" group (i.e., it plots $\gamma_r/\beta_0 \times 100$), which is meant to account for differences in severity across health shocks. Red dots correspond to estimates for heart attack that were discussed in detail in section 1.3.

In order to account for the fact that the 27 health shocks considered vary in terms of severity, in Figure 1.12b I scale the estimated effects by the immediate effect for the

---

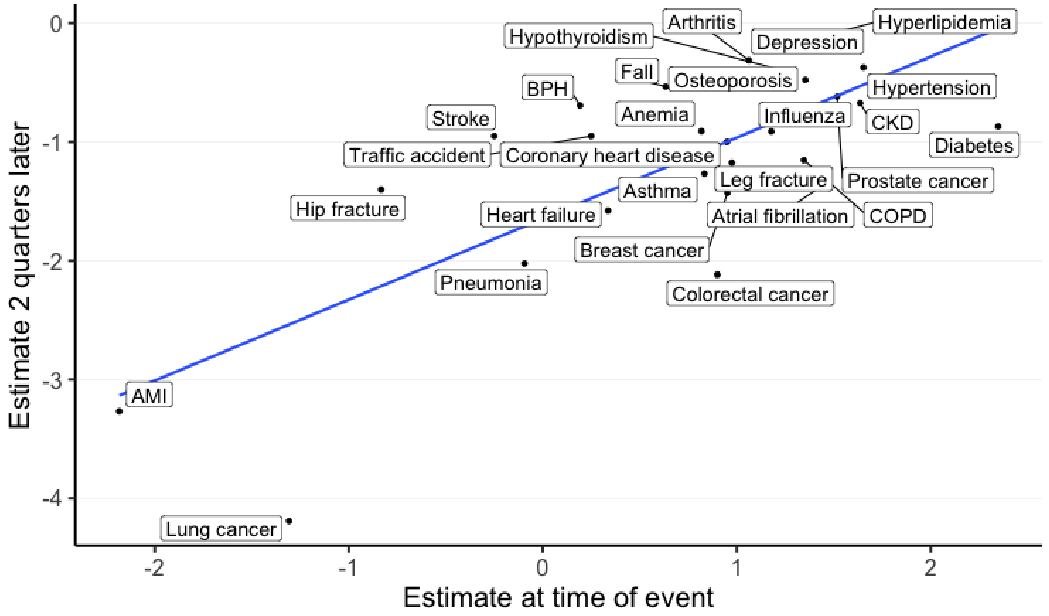
[22] Crucially, my choice of AMI as the main condition of interest was motivated by quasi-randomness of its onset, as well as no evidence in the medical literature concerning its biological relationship with dementia.

[23] For comparison, in the Appendix Table 1.A.2 I show that the pre-period average difference in care frequency between the "condition first" and "dementia first" groups ranges from 1.3 to 2 days of care for the 5 conditions for which I have that statistic.

"condition first" group, β_0. This scaling reflects the idea that long-term changes in care frequency for a serious condition such as lung cancer (around 10 more days of care at onset) are expected to be higher in absolute terms than those for relatively minor conditions like hyperlipidemia (only 2.5 more days at onset among "condition first"). In such relative terms, the distribution of immediate effects becomes more dispersed, while effects at 2 quarters and 2 years remain concentrated with medians of -20% and -59%, respectively. Moreover, with respect to immediate changes in care frequency at condition onset among the control group (i.e., "condition first"), the long-term estimates for AMI no longer seem extreme. Furthermore, the scaled distribution illustrates the fact, that, with respect to the "cost" of treating the condition in terms of days of care, measured by β_0, the decreases in care frequency for the "dementia first" group are of substantial magnitude. When it comes to the dollar cost of care, in Appendix Figure 1.E.1 I show that dementia leads to a reduction in the quarterly cost of medical services 2 years after a health shock, with reductions ranging from \$684 to \$6353, depending on the condition.

Figure 1.12 raises the possibility of a short-term vs long-term care trade-off, as for many conditions dementia patients receive more care at the time of event (both in terms of cost and frequency), and less in the long run. As dementia causes confusion and forgetfulness, it might also lead physicians to prescribe low-maintenance treatments in anticipation of issues with treatment adherence. Going back to the prostate cancer example from section 1.2.6, physicians would be reluctant to advise low-intensity watchful waiting to dementia patients, as it requires frequent follow-up visits, and would instead be more likely to prescribe surgery that leads to high immediate costs, but possibly less frequent care in the long term.

Figure 1.13: Immediate and 2-quarters-later treatment effect for days of care



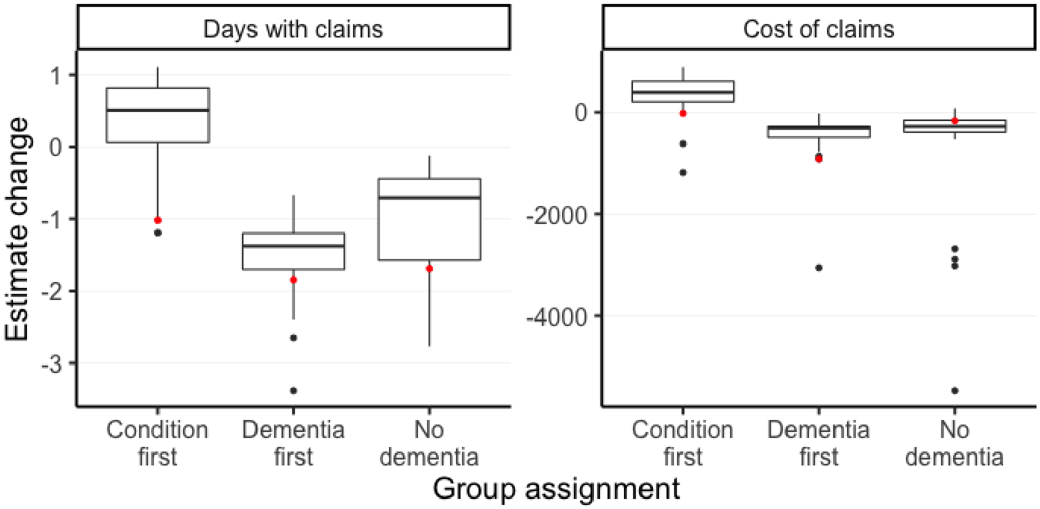
Notes. The plot shows the relationship between γ_0 and γ_2 from a regression of days with medical claims as the dependent variable across 27 health shocks considered (see Appendix Figure 1.E.2 for cost of medical care). The linear regression line shown in blue has a slope of 0.6821 (standard error 0.1244) that is significantly different than zero.

To investigate whether more short-term care is associated with less long-term care, in Figure 1.13 I plot the relationship between the immediate and 2-quarters-after treatment effects across health shocks. Although most conditions are associated with a positive treatment effect at the time of event, all of them lead to negative treatment effects 2 quarters after, and the relationship between the two estimates is strongly positive (correlation of 0.68). This positive relationship is robust to considering the 2-years-after estimate instead of the 2-quarters-after one (correlation of 0.57) as well as looking at cost of medical care instead of care frequency (correlation of 0.38, see Appendix Figure 1.E.2). Of course these correlations are only suggestive, as they do not account for various characteristics of the conditions in question, while weighing them equally. For example, the extent to which dementia can affect treatment decisions for a condition is an important omitted variable, as we expect both immediate and long-term estimates to be low for such conditions. However, the mechanism leading to long-term decreases in care frequency of dementia patients following another health shock seems to be to some extent universal across health shocks (e.g., the distribution of long-term effects is much more concentrated than that of short-term effects), and the aforementioned correlation is inconsistent with this care being substituted by more immediate medical interventions. Moreover, for most conditions reductions in long-term care use are larger than increases in immediate care use caused by dementia, and they are more persistent, as negative treatment effects are usually present throughout the time horizon for $r = 2, \ldots, 8$. Thus, cumulative treatment effects of dementia amount to substantial decreases in both the frequency and cost of medical services used in response to an unrelated health shock.

1.4.2 Effect sizes with respect to the "no dementia" sample

In this section I compare health care utilization trajectories of the "condition first" and "dementia first" groups to the analogous estimates for the "no dementia" sample, that come from a two-way fixed effects event-study regression for those who have never had dementia. The "no dementia" sample is excluded from my original design due to selection concerns, but it can serve as a benchmark when it comes to disentangling the effect of dementia via mechanisms of interest from its direct effect due to treatment of dementia itself. In the case of heart attack, I looked at heterogeneity by the time of dementia onset to show that at most 50% of the estimated difference in care frequency between the "dementia first" and "heart attack first" groups can be attributed to dementia being diagnosed in the post-period for the latter group (section 1.3.2). Here I complement that approach with the comparison to "no dementia" across the 27 health shocks included in my analysis.

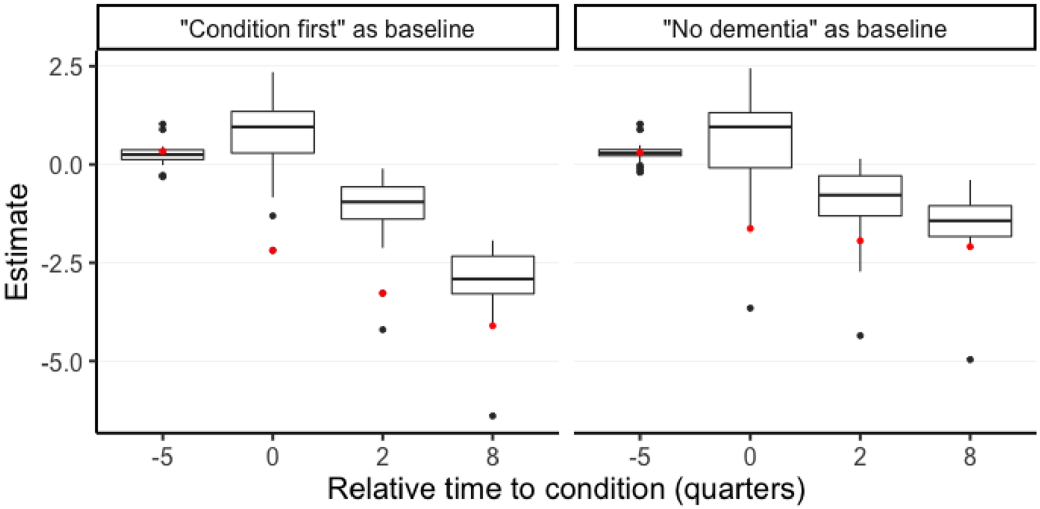
Figure 1.14: Change in treatment effect from 2 to 8 quarters after health shock



Notes. The plot shows the distribution across health shocks of the change of estimated event-study coefficient from 2 to 8 quarters after the health shock by relative time of dementia diagnosis. For "condition first" this corresponds to $\beta_8 - \beta_2$, for "dementia first" to $(\beta_8 + \gamma_8) - (\beta_2 + \gamma_2)$, and is analogous for the "no dementia" group for which the two-way fixed effects event study regression is estimated separately (for details see Appendix 1.C). Red dots highlight values for heart attack.

In Figure 1.14 I plot the difference between event-study coefficients for quarters 8 and 2 after the health shock depending on the relative time of dementia onset, which shows attenuation of the estimated effect of each health shock. Crucially, estimates for quarter 2 correspond to a time somewhat after the health shock onset at which members of the "condition first" group are not yet diagnosed with dementia, and so by construction cannot be driven by the diagnosis and treatment of dementia itself. Unsurprisingly, the "no dementia" group exhibits smaller treatment effects further away from the health shock, both in terms of frequency and cost of care, for almost all health shocks considered. The analogous distribution for “dementia first” is more compressed and shifted downwards, which is consistent with dementia patients receiving less treatment for unrelated health shocks. However, for “condition first” both cost and care frequency increase as one gets further away from the health shock for more than three quarters of the conditions considered, which can only be explained by the direct effect of dementia diagnoses between 2 and 8 quarters after the health shock in question.

Figure 1.15: Main estimate compared to using “no dementia” patients as control group



Notes. Both panels show the distribution of estimates across 27 health shocks considered for $r \in \{-5, 0, 2, 8\}$ with days with medical claims as dependent variable (see Appendix Figure 1.E.3 for cost of medical care). The right pane replaces “condition first” with “no dementia” patients as the control group, while the left pane shows the baseline estimates for reference. Red dots highlight estimates for heart attack.

Figure 1.15 shows the distribution of treatment effect estimates depending on the definition of control group. The left pane corresponds to the original specification, whereas the right pane uses “no dementia” patients as controls (instead of “condition first”), so does not combine the aforementioned direct and indirect effects of dementia[24]. Clearly, the distribu-

---

[24] Of course, this benefit comes at the cost of exacerbated selection bias, as the “no dementia” sample is

tion of immediate effect estimates (i.e., $r = 0$) has a similar range but is somewhat more dispersed when "no dementia" serves as a control group, which is perhaps a consequence of larger differences in health status with respect to the treated group (i.e., "dementia first"). While estimates for 2 quarters after the health shock are similar across the two approaches, those for 8 quarters afterwards are smaller in absolute terms when the control group is modified as the treatment effect distribution shifts upwards. In the case of heart attack the original specification implied a 4.1 decrease in days of care at $r = 8$, while changing the control group results in a decrease of only 2.1 days of care. Thus, purging the estimate from direct effect of dementia (via its treatment and diagnosis) leads to halving of the estimated treatment effect which is consistent with the prior approach of investigating heterogeneity by the time of dementia onset. Across the 27 conditions considered, the median effect at $r = 8$ changes from -2.9 to -1.4, and is robustly negative with a maximum of -0.4 in the case of hypertension.

1.4.3 Treatment effects for prescription drug usage

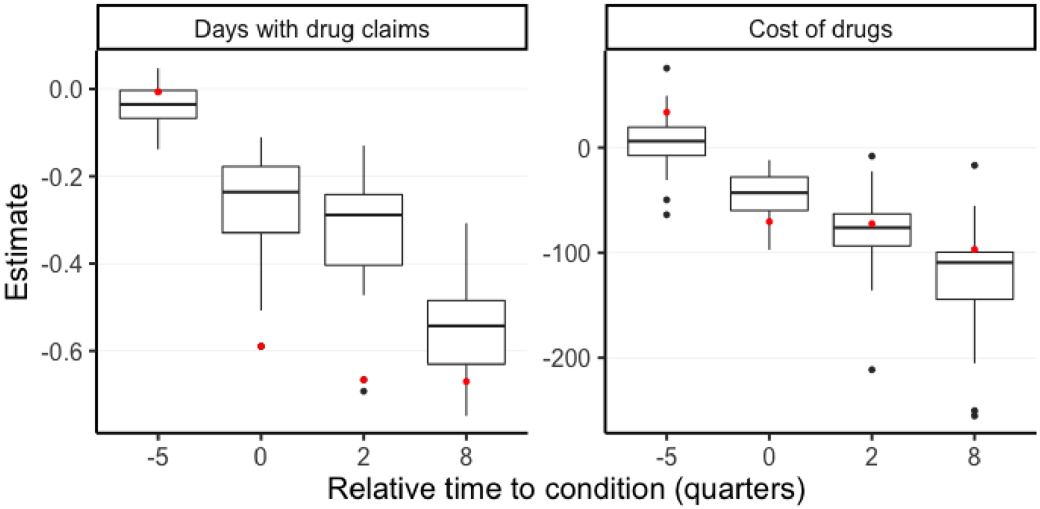
As in the case of heart attack, I separately investigate the effect of dementia on prescription drug utilization following other health shocks. In Figure 1.16 I plot the distribution of estimates for both the cost and frequency of drug purchases across conditions. Similarly to medical services, dementia leads to a persistent decrease in the frequency of prescription drug claims for all health shocks considered, which increases in absolute terms as the time

---

presumably younger and healthier than the "dementia first" one. However, this is likely to be more of a concern for immediate treatment effects that reflect medical interventions at the time of onset of the health condition in question. Hopefully, follow-up care varies less with one's health status so that selection bias is less important for medium- and long-term effects, which are the focus of this section.

horizon expands. Unlike for medical services, for drugs a similarly robust negative effect is present for the cost of these claims, and the immediate treatment effect is negative for both cost and frequency. However, these estimates should be interpreted with caution, as there is evidence of possible violations of the parallel trends assumption for some conditions. In particular, the range of estimates for $r = -5$ is substantially smaller with respect to that for $r \geq 0$ in the case of medical services, which is not true for prescription drugs.

Figure 1.16: Distribution of treatment effects for prescription drug utilization



Notes. The panels show the distribution of estimates across 27 health shocks considered for $r \in \{-5, 0, 2, 8\}$ for different outcomes. Red dots highlight estimates for heart attack.

Furthermore, I consider the possibility of some of these effects being driven by the treatment and diagnosis of dementia itself by using "no dementia" patients as a control group in a similar fashion as I did for medical services in section 1.4.2. This results in treatment effects that are somewhat bigger in absolute terms, for both cost and drug purchase frequency (see Appendix Figure 1.E.4). Importantly, this is consistent with what I found in the case of heart attack by restricting the control group to those who have not yet been diagnosed with dementia. This reinforces the idea that dementia affects utilization of prescription drugs, possibly through the medication adherence channel, which potentially affects people well before their dementia diagnosis, as drug utilization of the "condition first" patients is somewhat lower than that of the "no dementia" patients despite their worse overall health status.

1.5 Conclusion

This paper uses variation in relative timing of dementia diagnosis with respect to onset of another health shock to show that having dementia at the time when medical treatment decisions are made leads to less future health care utilization, which continues for at least two years afterwards. Although dementia patients tend to receive more treatment at the time of the health shock, subsequent and persistent decreases in care use result in an overall negative treatment effect of dementia on health care utilization. I estimate such a negative long-term effect of dementia on treatment of all 27 health shocks considered, and the effect is driven by decreases in frequency of care.

Specifically, for the median health shock dementia leads to 2.9 less days with claims per quarter, of which at most 50% can be explained by dementia care, with the remainder attributed to the effect of dementia on treatment decisions concerning the unrelated health shock in question. My evidence suggests that dementia patients receive more invasive but infrequent care as decreases in care frequency are driven by outpatient and diagnostic care while a countervailing effect is present for inpatient procedures. I provide separate estimates for prescription drugs, showing a median decrease of drug spending of $100 per quarter, more than one third of average drug spending of a dementia patient, none of which can be explained by dementia treatment.

I provide first causal estimates of the effect of dementia on health care utilization, which are robustly negative and stand in contrast to previous literature showing positive correlation between dementia and medical spending. Naturally, causal interpretation of my estimates comes at a cost, as I consider the specific setting of treatment of unrelated health shocks. Furthermore, I use Medicare Advantage claims which are limited to services covered by Medicare and only include a fraction of nursing home care, which is a major component of health care spending of dementia patients. Therefore, my analysis rules out dementia causing increased medical spending when it comes to Medicare-covered services, but cannot speak to the effect on overall health care utilization. Thus, policy interventions attempting to curb costs of dementia patients should focus on non-covered care, while investigating the causal relationship between dementia, cost and nursing home care is a crucial direction for future research.

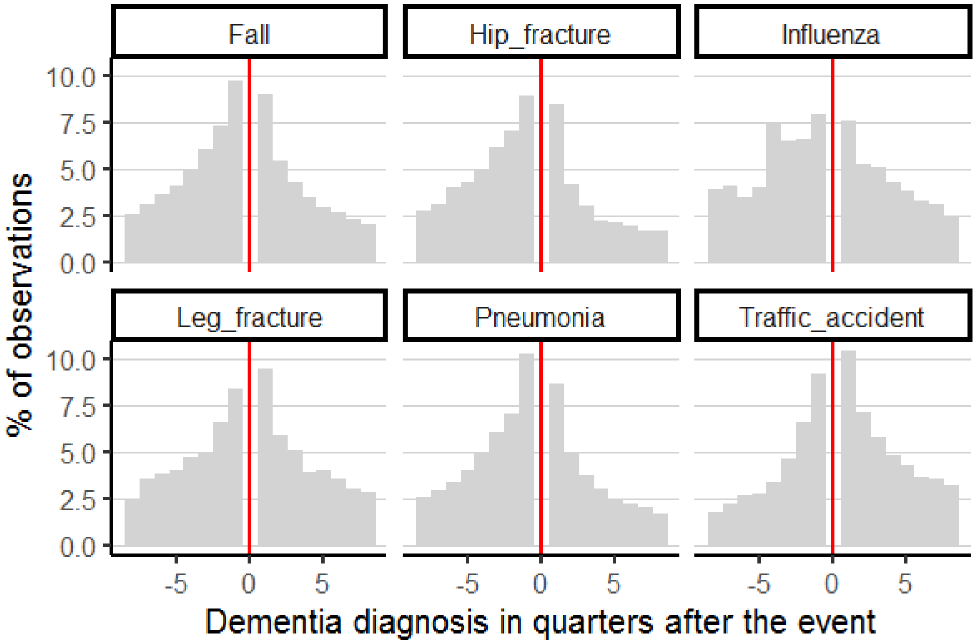
A major limitation of this paper is that it does not allow to draw conclusions concerning consequences for patient utility. Changes in health care utilization documented here could reflect under-utilization of care among dementia patients as well as over-utilization of care among non-dementia patients. However, these treatment effects do not seem to be a rational response to lower life-expectancy of dementia patients, as they are similar upon restricting attention to a balanced panel (i.e., those who survived at least 10 quarters after the health shock). Thus, unless differences in timing of dementia diagnosis are mirrored by stark differences in patient preferences, my estimates corroborate the hypothesis that dementia diagnosis affects medical decision-making in a way that should perhaps be counteracted by policy. Further evidence concerning the effect on objective outcomes such as mortality and wasteful medical interventions could help shed light on the extent to which this is the case.

The paper documents that dementia patients purchase less prescription drugs on average than the not-yet dementia patients and that the onset of a health shock leads to a trend-break in their drug use, resulting in further utilization decreases. Qualitatively, this result holds regardless of the health shock considered, which is surprising due to the fact that health shocks vary in the importance of prescription drugs in their treatment. This points to medication adherence as a crucial mechanism behind the results, unless physicians systematically simplify and limit the use of drugs for dementia patients when they encounter additional health problems. Investigating these mechanisms further is a promising avenue for future work, especially if long-standing prescriptions for pre-existing chronic conditions (e.g., hypertension) can be identified. Assuming that at least part of the estimated effect arises due to the medication adherence channel, it seems that simplifying medication regimes or encouraging adherence are two possible policy interventions that can improve care of dementia patients.

Appendix

1.A Additional summary statistics

Figure 1.A.1: Distribution of relative timing of injury and dementia diagnosis



Notes. Each bar indicates the percentage of people that had dementia diagnosed a specific number of quarters after the condition in question. Diagnoses in the same quarter (0 on horizontal axis) are omitted. Negative (positive) values indicate dementia diagnoses before (after) the condition. Injuries are identified using ICD-10 and ICD-9 diagnosis codes based on their description.

Table 1.A.1: Additional statistics by relative time of dementia and AMI

	No AMI (1)	AMI first (2)	Dem. first (3)	Diff. (4)
A. Cost per day with a health care claim from each category ($)				
Ambulance transportation	258.7	294.8	191.9	-102.9 **
Evaluation and management	59.9	56.7	49.9	-6.8 ***
Inpatient claims	4268.9	3815.4	2796.6	-1018.8 *
Nursing home claims	1373.7	1700.1	1413.8	-286.3
Outpatient surgery	110.4	135.2	79.9	-55.3 **
Pathology and lab	49.3	44.5	35.9	-8.5 *
Radiology	125.4	96.6	79.0	-17.6
B. Days with a health care claim from each category				
Ambulance transportation	0.2	0.1	0.3	0.2 ***
Evaluation and management	3.9	3.7	4.9	1.1 ***
Inpatient claims	0.2	0.1	0.3	0.2 ***
Nursing home claims	0.2	0.1	0.2	0.2 ***
Outpatient surgery	1.3	1.6	1.6	0.0
Pathology and lab	1.0	1.2	1.3	0.0
Radiology	0.9	0.8	1.0	0.2 ***
C. Pre-existing conditions diagnosed (%)				
Acquired hypothyroidism	21.1	17.0	26.8	9.8 ***
Anemia	35.4	32.7	51.1	18.4 ***
Asthma	6.4	5.6	8.5	2.9 ***
Atrial fibrillation	17.9	16.5	24.8	8.3 ***
Benign prostatic hyperplasia	0.7	0.3	1.2	0.9 ***
Breast cancer	3.9	2.4	3.3	0.9 **
Chronic kidney disease	28.4	30.6	48.8	18.3 ***
Chronic obstructive pulmonary disease	17.3	18.1	27.4	9.3 ***
Colorectal cancer	1.7	1.6	2.0	0.5
Depression	26.9	12.8	37.6	24.8 ***
Diabetes	28.7	40.6	41.7	1.1
Heart failure	21.7	22.3	40.5	18.1 ***
Hip/pelvic fracture	4.4	1.3	6.2	4.8 ***
Hyperlipidemia	55.1	57.8	66.2	8.5 ***
Hypertension	74.1	79.1	89.3	10.2 ***
Ischemic heart disease	34.4	48.2	56.2	8.0 ***
Lung cancer	1.1	1.0	1.3	0.3
Osteoporosis	10.7	6.5	12.0	5.5 ***
Prostate cancer	5.0	5.4	6.5	1.1 *
Rheumatoid arthritis/osteoarthritis	39.7	37.7	50.8	13.0 ***
Stroke/transient ischemic attack	0.9	0.4	2.2	1.8 ***

Notes. Extension of Table 1.2. Reference period is two quarters before AMI diagnosis.

Table 1.A.2: Differences in health care use by condition timing for various conditions

	No AMI (1)	AMI first (2)	Dem. first (3)	Diff. (4)	
A. Use of medical services in the reference period (\$)					
AMI	1980.4	1700.2	2048.3	348.1	*
Breast cancer	2062.8	916.5	1000.5	84.0	
Prostate cancer	2024.7	1122.1	1525.1	403.0	**
Leg fracture	2022.4	1531.0	2019.5	488.5	**
Traffic accident	2039.6	1329.3	1758.3	429.0	**
B. Cost per day with medical claims in the reference period (\$)					
AMI	201.5	186.4	162.4	-24.0	*
Breast cancer	205.5	128.3	128.2	-0.1	
Prostate cancer	203.1	159.5	184.3	24.7	
Leg fracture	203.9	162.4	170.2	7.8	
Traffic accident	203.9	152.6	153.7	1.1	
C. Days with medical claims in the reference period					
AMI	6.8	6.9	8.9	2.0	***
Breast cancer	6.9	5.5	6.9	1.4	***
Prostate cancer	6.8	5.5	7.2	1.7	***
Leg fracture	6.8	7.3	8.6	1.3	***
Traffic accident	6.8	6.4	8.0	1.6	***

Notes. Reference period is two quarters before condition diagnosis. Cost per day is calculated only among those who had non-zero cost. Column 4 is the difference between columns 3 and 2.

1.B Estimation of propensity score weights

In order to account for unobserved differences between the "condition first" and "dementia first" groups I use a machine learning algorithm to estimate the predicted probability of each patient being in the "dementia first" group (i.e., propensity score $p(x)$). Then, following Abadie and Cattaneo (2018), when running regressions each patient is weighted by the inverse of the predicted probability of belonging to its realized group (i.e., "dementia first" are weighted with $\frac{1}{p(x)}$ and "condition first" are weighted with $\frac{1}{1-p(x)}$). The remainder of this section provides details of the propensity score model used, in particular: selection and processing of predictors, tuning of the predictive model and model diagnostics. I use heart attack as the unrelated condition to showcase my approach[25].

I consider "dementia first" and "AMI first" groups, with the outcome variable of interest equal to 1 for the former and 0 for the latter group. For each patient, I restrict attention to the pre-period (i.e., two or more quarters before her heart attack), and collect the following information: diagnosis codes from all inpatient, outpatient and physician claims, Current Procedural Terminology (CPT) codes that correspond to outpatient services provided to the patient, Diagnosis-Related Group (DRG) codes that reflect hospital care provided to the patient. For diagnosis codes, I restrict attention to the first 3 digits, which broadly define the diagnosis in question, and only include full 5-digit codes for ten most common 3-digit diagnosis codes. Then, for each patient I create indicators for having each code

---

[25] Fine tuning of the predictive model is particularly time-consuming, so I only did this step for heart attack. However, reasonable predictive performance can be achieved without that step, which is what I did when estimating propensity scores for breast cancer, prostate cancer and leg fracture. Similarly to the case of heart attack, reweighting the sample did not markedly alter my results for these conditions, so I use the unweighted model throughout the paper.

(diagnosis, CPT or DRG) throughout the pre-period, as well as having it in the reference period, $r = -2$. This creates two binary variables for each code present in the pre-period claims of the "dementia and AMI" patients. In addition, for each group of codes (diagnosis, CPT and DRG) I create an indicator for the patient not having any of the code of that kind, that adds up to a total of 17876 predictors. As there are 7611 patients in my sample, the sheer number of predictors with respect to the number of observations makes this setting suitable for machine learning prediction models.

I consider two types of predictive models: random forest and tree-based gradient boosting. The first one is easy to use as it only requires two tuning parameters: the number of trees and m which determines how many predictors are randomly drawn for consideration at each tree split. I consider 500 (default) or 1000 trees, and $m \in \{67, 134, 268\}$, where 134 is the default (i.e., square root of the number of predictors). Tree-based gradient boosting requires the following additional parameters: maximal tree depth d and learning rate η. I also specify an early stopping parameter which stops the algorithm if the specified number of iterations did not improve model fit. For tuning the model, I use 500 or 1000 trees, all predictors (i.e., $m = 17876$), and 10 combinations of the remaining three parameters, drawn from their respective ranges using a latin hypercube that attempts to distribute the draws evenly across the three-dimensional space in question. The full set of tuning parameters considered is shown in Table 1.B.1.

I train the model using a 75% training sample using 10-fold cross-validation, which means that the training sample is split into 10 even parts, stratifying by the outcome variable, so that accuracy metrics can be calculated out-of-sample, with each of the 10 parts treated as

Table 1.B.1: Fit statistics of predictive models for various tuning parameters

Forest?	Trees	m	d	η	Stop	ROC	Accuracy	κ	Class. cost
	1000	17876	7	0.0105572454	25	0.8841	0.8067	0.5814	0.5735
	500	17876	15	0.0221981101	53	0.8830	0.8076	0.5810	0.5508
	1000	17876	14	0.0045263945	47	0.8801	0.8029	0.5728	0.5660
	1000	17876	20	0.0433353415	53	0.8765	0.8032	0.5689	0.5632
✓	500	268				0.8691	0.7931	0.5305	0.6748
✓	1000	268				0.8684	0.7934	0.5307	0.6764
✓	1000	134				0.8639	0.7883	0.5191	0.6953
✓	500	134				0.8629	0.7901	0.5230	0.6957
	1000	17876	10	0.0011092198	35	0.8600	0.7936	0.5537	0.7503
✓	1000	67				0.8570	0.7864	0.5137	0.7192
✓	500	67				0.8564	0.7864	0.5140	0.7191
	500	17876	5	0.0017473951	25	0.8480	0.7773	0.5071	0.8132
	1000	17876	7	0.0002140129	41	0.8383	0.7694	0.5055	0.9340
	500	17876	11	0.0003804997	47	0.8331	0.7652	0.4976	0.9371
	500	17876	7	0.0000302713	35	0.8297	0.7638	0.4928	0.9947
	1000	17876	5	0.0000015165	31	0.8242	0.7594	0.4655	0.9995
	500	17876	5	0.0000015660	41	0.8170	0.7517	0.4483	0.9998
	1000	17876	17	0.0000582476	52	0.8150	0.7489	0.4581	0.9791
	1000	17876	12	0.0000136423	36	0.8118	0.7498	0.4701	0.9951
	500	17876	13	0.0000001505	52	0.8110	0.7521	0.4687	1.0000
	1000	17876	2	0.0004810259	59	0.8076	0.7403	0.4045	0.9196
	1000	17876	15	0.0000087014	23	0.8069	0.7421	0.4485	0.9969
	500	17876	1	0.0000067283	59	0.6938	0.6751	0.3529	0.9995
	500	17876	11	0.0000000049	23	0.5000	0.6313	0.0000	1.0000
	500	17876	3	0.0000000002	31	0.5000	0.6313	0.0000	1.0000
	500	17876	9	0.0000000110	36	0.5000	0.6313	0.0000	1.0000

Notes. The first column indicates the type of model: random forest or gradient boosted tree. Trees indicates the number of trees, m is the "mtry" parameter that determines how many predictors are randomly drawn for consideration at each split (out of 17876), d is maximal tree depth, η is the learning rate, "Stop" gives the number of consecutive iterations of the algorithm that trigger the early stopping rule if no improvement in model fit was obtained. The remaining columns correspond to measures of out-of-sample predictive accuracy of the model, which are averaged over the 10 testing samples coming from cross-validation. ROC is the area under the ROC curve, accuracy is the fraction of correctly predicted observations, κ is a measure of accuracy that corrects for unbalanced classes, classification cost is equal to $1/p$ for false negatives and $1/(1-p)$ for false positives, where p is the empirical probability of being in the "dementia first" group.

an out-of-sample prediction in turn. Thus, the final model is selected based on out-of-sample performance, while the remaining 25% of observations (i.e., testing sample), are only used for a final assessment of model fit. I use the area under the ROC curve as the main criterion for model selection, but also report accuracy,, as well as two measures of fit meant to account for the fact that the "dementia first" and "AMI first" groups are not of equal size: κ (a measure of accuracy accounting for unbalanced categories), and misclassification cost, equal to $1/p$ for false negatives and $1/(1-p)$ for false positives, where p is the empirical probability of being in the "dementia first" group.

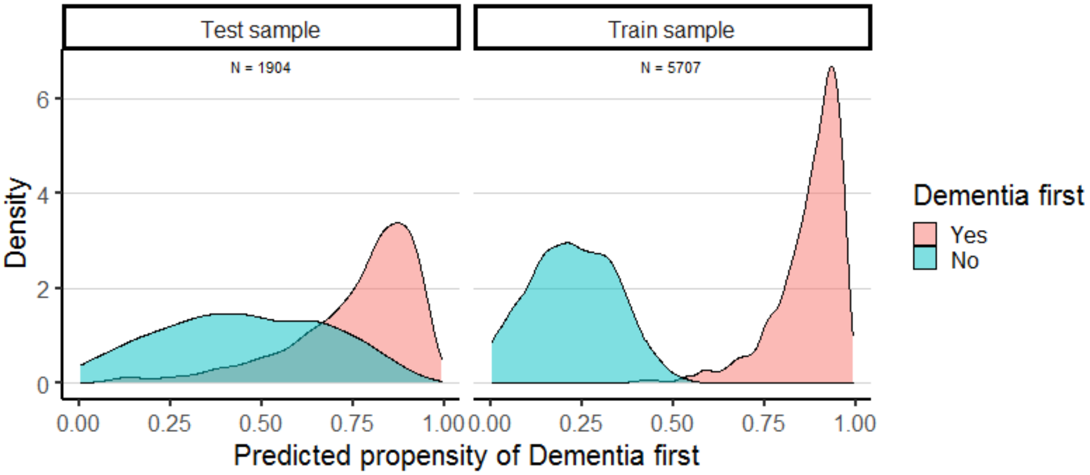
I present results from model tuning in Table 1.B.1, which shows that the best gradient boosting models outperform the random forest, but the poorly tuned ones do not[26]. In the paper, I used the best gradient boosting model as well as the best random forest model for generating weights and verify that my results are robust to the chosen type of predictive model. Figure 1.B.1 shows the distribution of propensity scores by actual group assignment. In the case of random forest, the supports of propensity score distributions have little overlap between the true group assignment when we look at the training sample, while a lot of overlap is present for the test sample. At the same time the plots for the test and training samples look similar if we consider gradient boosting, which suggests less over-fitting.

---

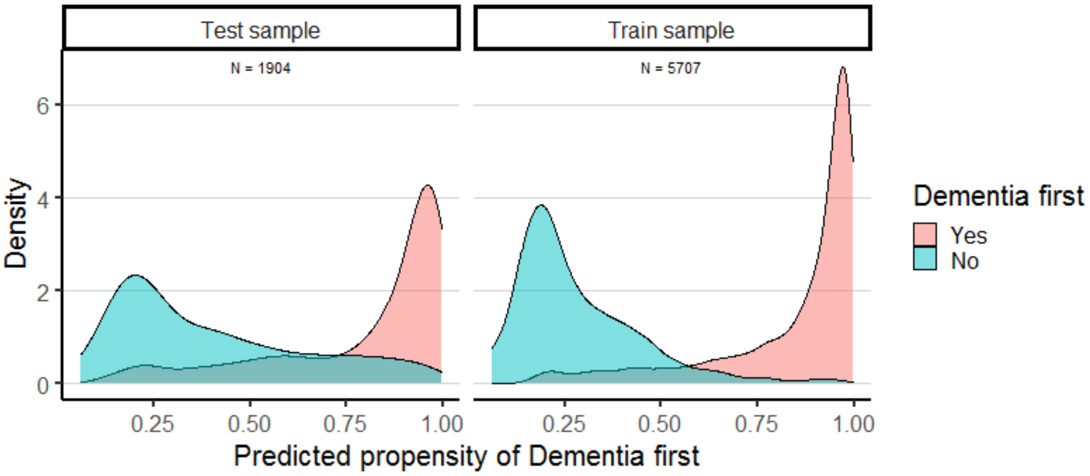
[26] In particular, the three bottom rows of the table contain the worst combinations of tuning parameters for gradient boosting and lead to the area under the ROC curve equal to 0.5, which indicates no predictive ability of the model. This confirms common wisdom that more complex machine learning tools require substantial tuning if they are to outperform the more popular methods such as an out-of-the-box random forest.

Figure 1.B.1: Density of estimated propensity scores by type of model

(a) Random forest



(b) Gradient boosting



Notes. Propensity score is the predicted probability of being in the "dementia first" group based on pre-heart attack health information. Colors correspond to the observed group assignment and panes show the difference in predictive accuracy between the training and test sample.

1.C Event-study estimates for "no dementia" sample

Event-study estimates for the "no dementia" sample come from the following regression

$$\text{Health care use}_{it} = \alpha_i + \tau_t + \sum_{r \neq -2} \delta_r + \varepsilon_{it}, \tag{1.2}$$

where r measures relative time to the health shock in question in quarters, all other components mirror regression (1.1) and the sample is restricted to those who never had dementia.

Figure 1.C.1 shows the trajectory of use of medical services around the time of AMI onset for the "no dementia" sample. Clearly, overall cost of medical services is characterized by a small anticipatory effect one quarter before the heart attack, a sharp increase at the time of onset, which stabilizes at a much lower level within two quarters of onset. The long-term effect (i.e., for $r > 2$) is significant, positive and does not fade out for the entire horizon considered. This pattern is qualitatively similar to the analogous event-study trajectory for "AMI first" but not for "dementia first" patients, as shown in Figure 1.3. However, two crucial differences arise. First, both the immediate and long-term effects are larger for the "no dementia" sample, which is presumably due to its different composition, particularly when it comes to health status[27], reinforcing the idea that the "no dementia" sample is not an ideal comparison group for "dementia first". Second, for "no dementia" the intensive margin also seems to play a role, as cost per day with claim remains slightly, but significantly, elevated throughout the time horizon. This stands in contrast to the analogous estimate for

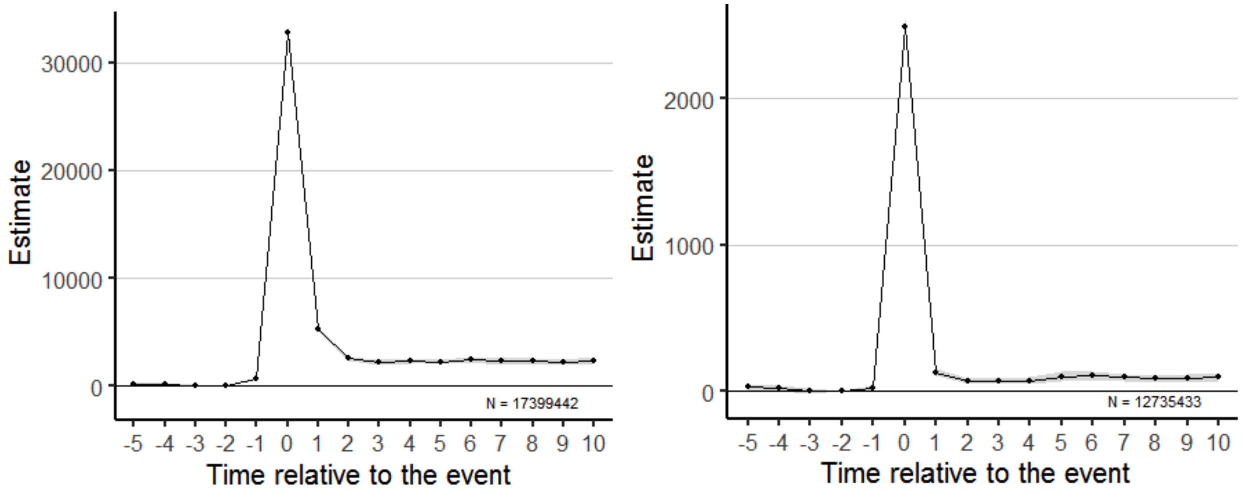
---

[27] For example, Zeltzer, Einav, Finkelstein, Shir, Stemmer, and Balicer (2020) show that among people with the same cancer type and predicted mortality, end-of-life spending is higher among the person who is younger, which could be due to preferences for care intensity varying by age.

“AMI first”, which is smaller in magnitude and insignificant for most values of r.

In Figure 1.C.2 I show event-study trajectories for prescription drug use for the “no dementia” sample. The comparison to estimates for “AMI first” and ”dementia first” (see Figure 1.D.2) is somewhat different here, as neither of the dementia groups exhibit any significant increases in the cost of drugs purchased nor in their cost per day, while “no dementia” patients increase both (although the estimate for cost per day is small and not consistently significant). At the same time, the event-study trajectory for days with drug claims is remarkably similar for “no dementia” and “AMI first” groups (i.e., sharp increase that tapers off slowly throughout the horizon), while estimates for “dementia first” are consistently negative or close to zero.

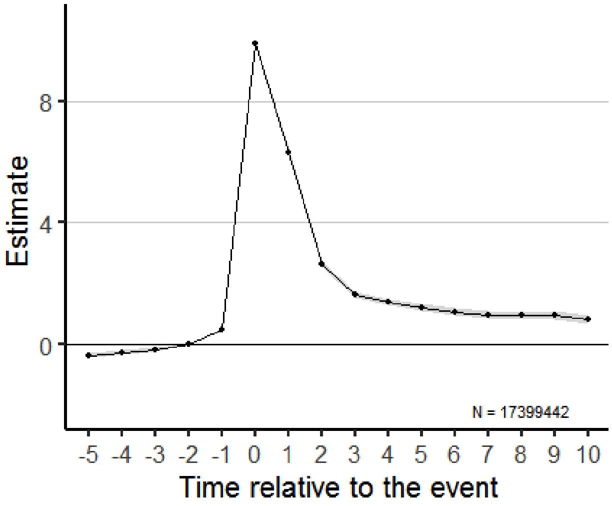
Figure 1.C.1: Use of medical services at onset of AMI among “no dementia” patients



(a) Cost ($)

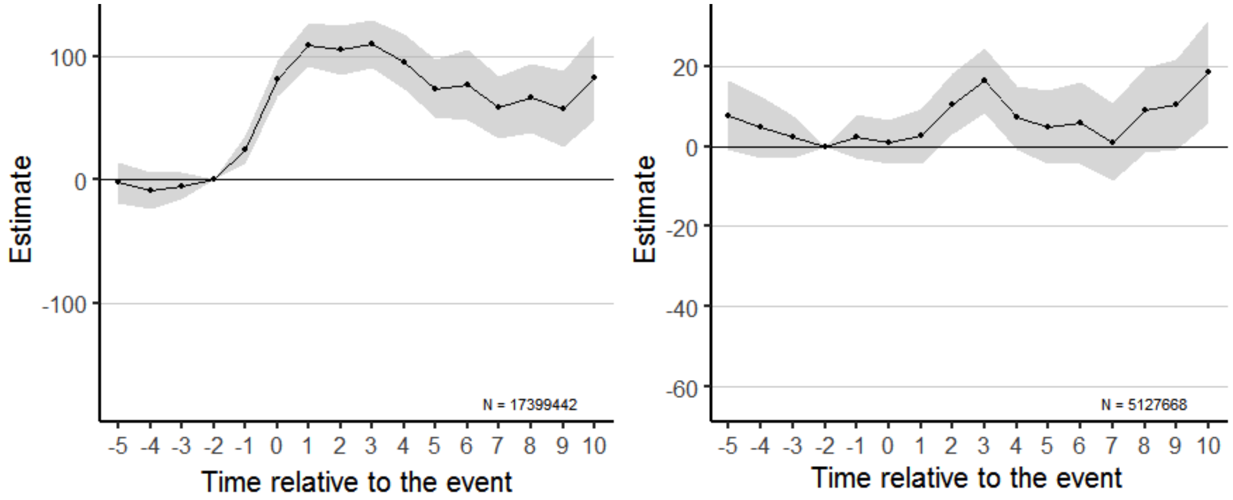
(b) Cost per day with claim ($)

(c) Days with claims



Notes. Each plot shows estimates of event-study coefficients δ_r from equation (1.2) with 95% confidence intervals for various health use measures as the dependent variable.

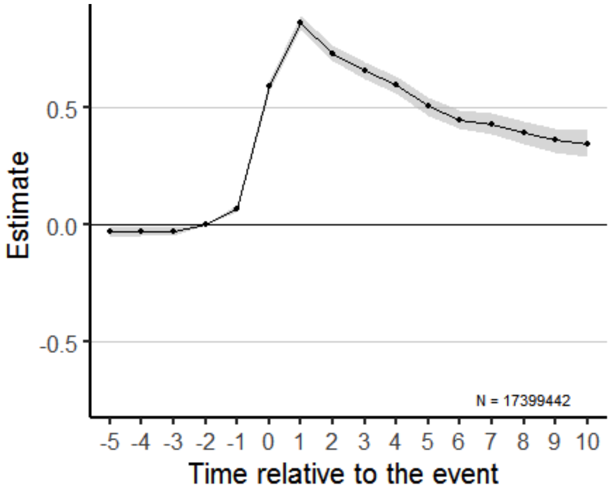
Figure 1.C.2: Use of prescription drugs at onset of AMI among "no dementia" patients



(a) Drug cost ($)

(b) Cost per day with drug claim ($)

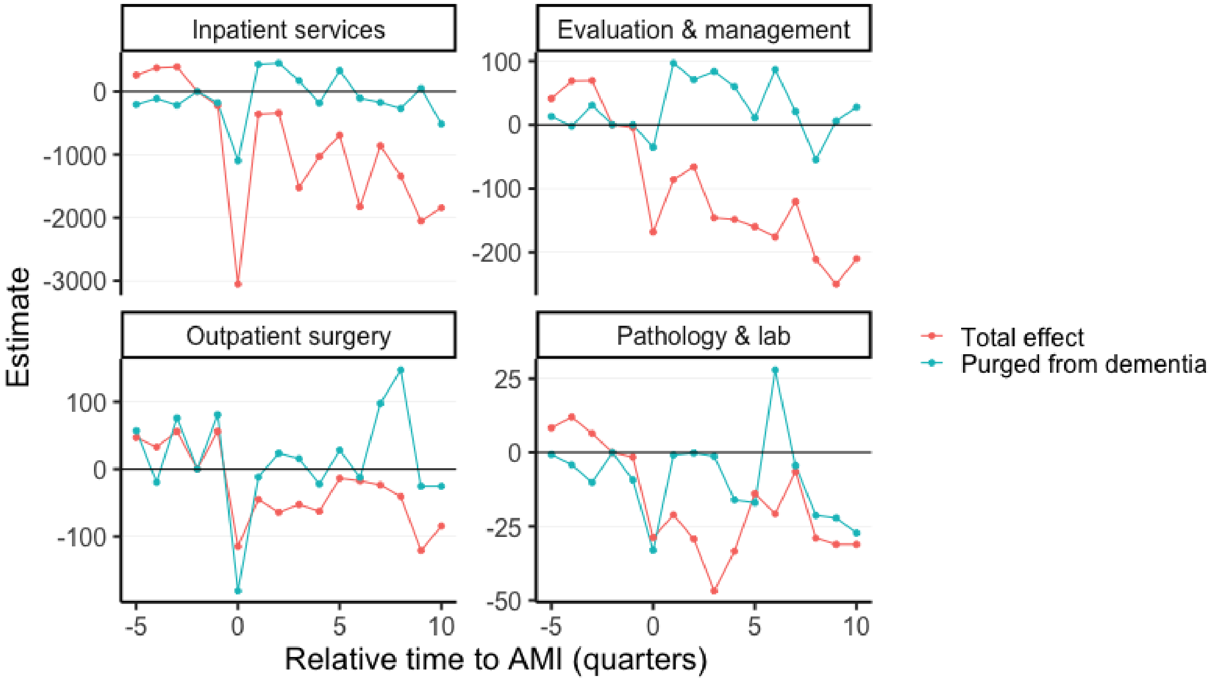
(c) Days with drug claims



Notes. Each plot shows estimates of event-study coefficients δ_r from equation (1.2) with 95% confidence intervals for various health use measures as the dependent variable.

1.D Additional event-study estimates for AMI onset

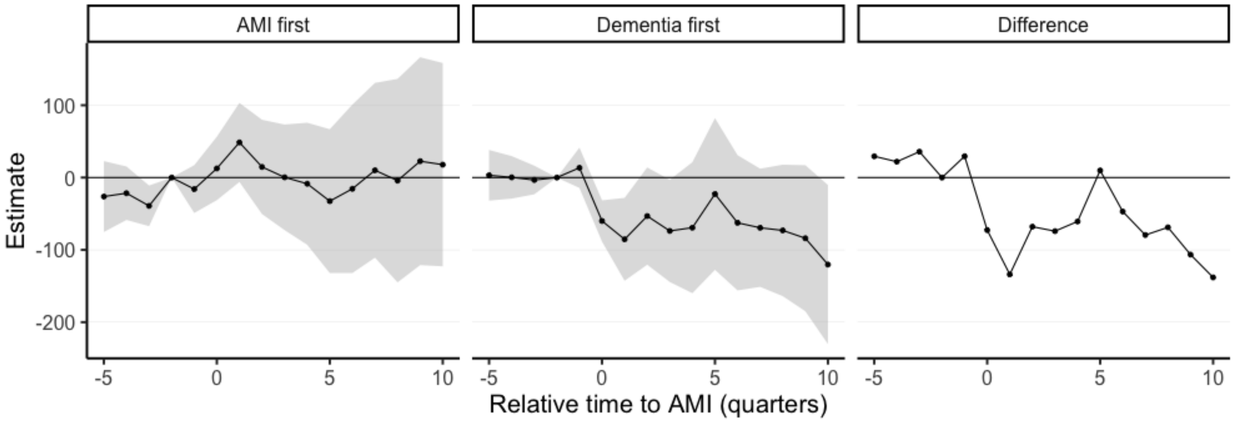
Figure 1.D.1: Effect on cost of specific types of care



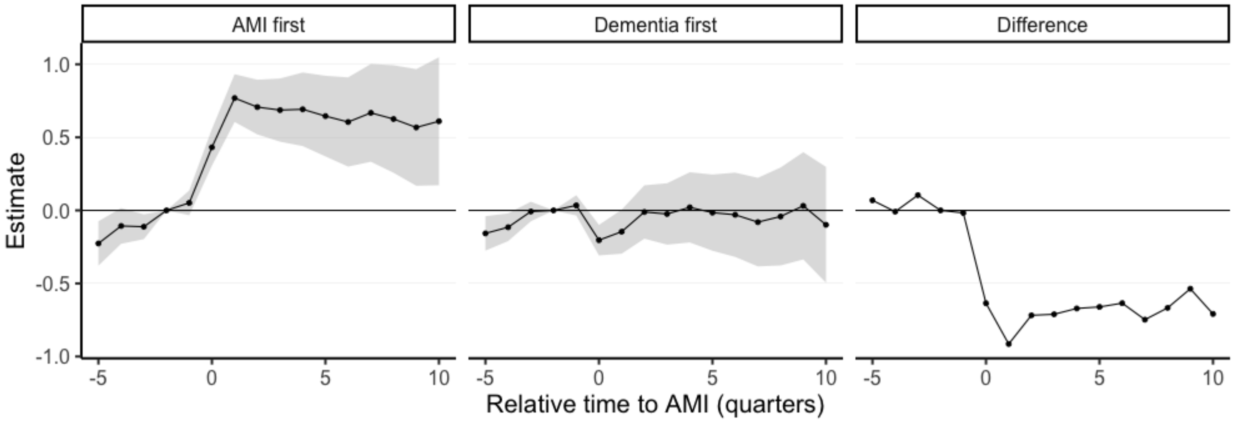
Notes. “Purged from dementia” refers to estimates for those whose dementia diagnosis is at least 10 quarters away from heart attack onset (i.e., this is the “more than 10” group from Figure 1.6).

Figure 1.D.2: Event-study estimates for utilization of prescription drugs

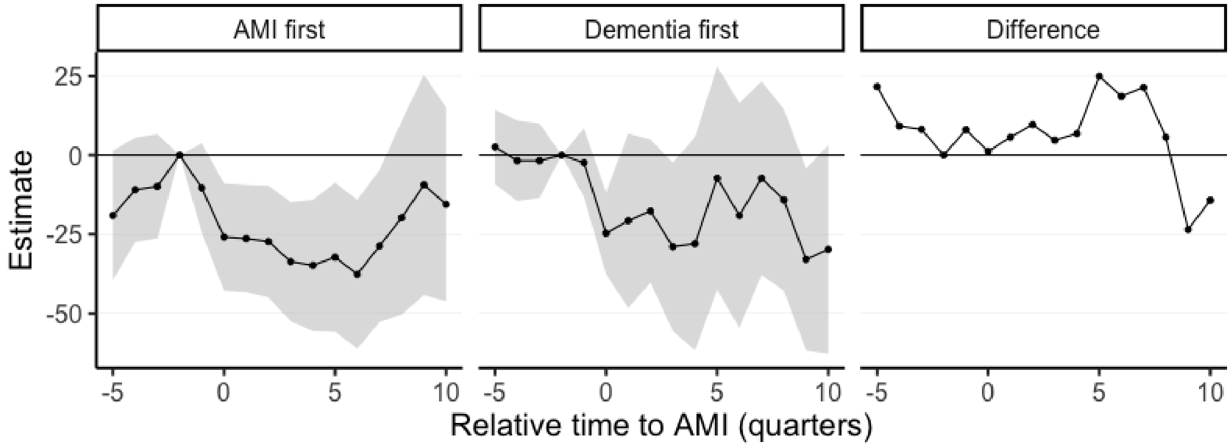
(a) Drug cost ($)



(b) Days with drug claims



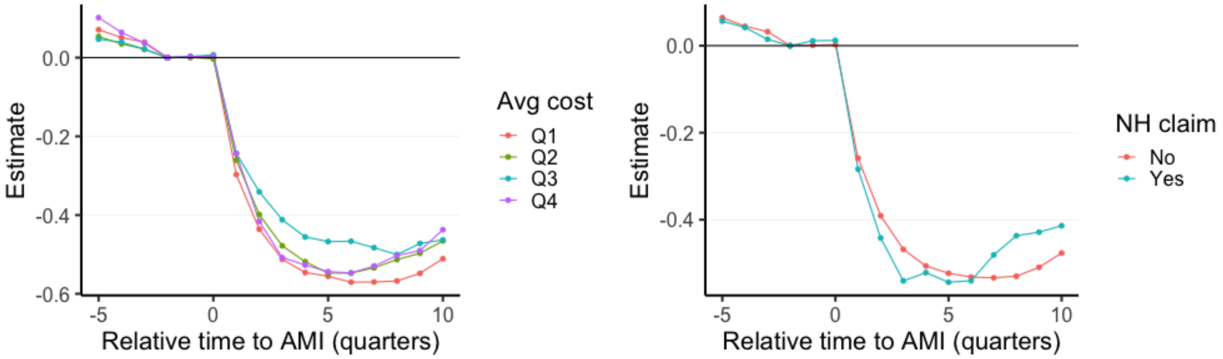
(c) Cost per day with drug claim ($)



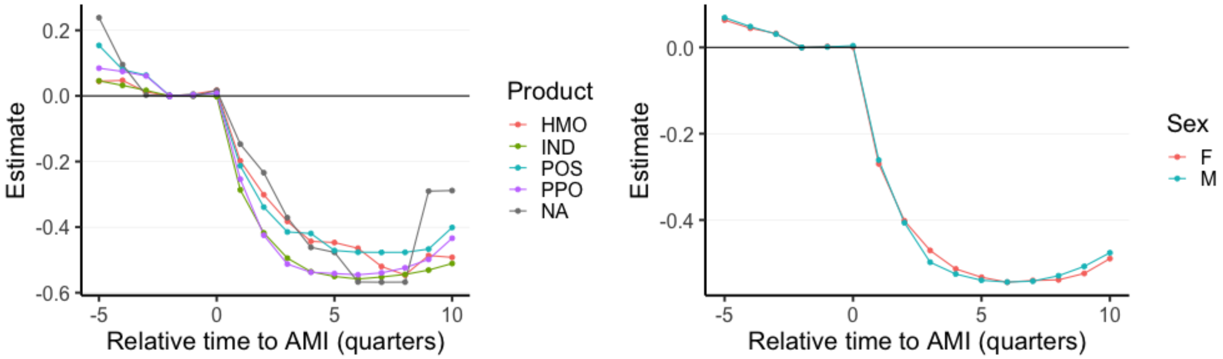
Notes. Panels "AMI first", "Dementia first" and "Difference" show estimates of β_r, $\beta_r + \gamma_r$ and γ_r, respectively. Standard errors for γ_r will be added upon re-accessing the data enclave.

Figure 1.D.3: Heterogeneity of attrition estimates

(a) By average pre-period cost quartile **(b)** By nursing home claim indicator



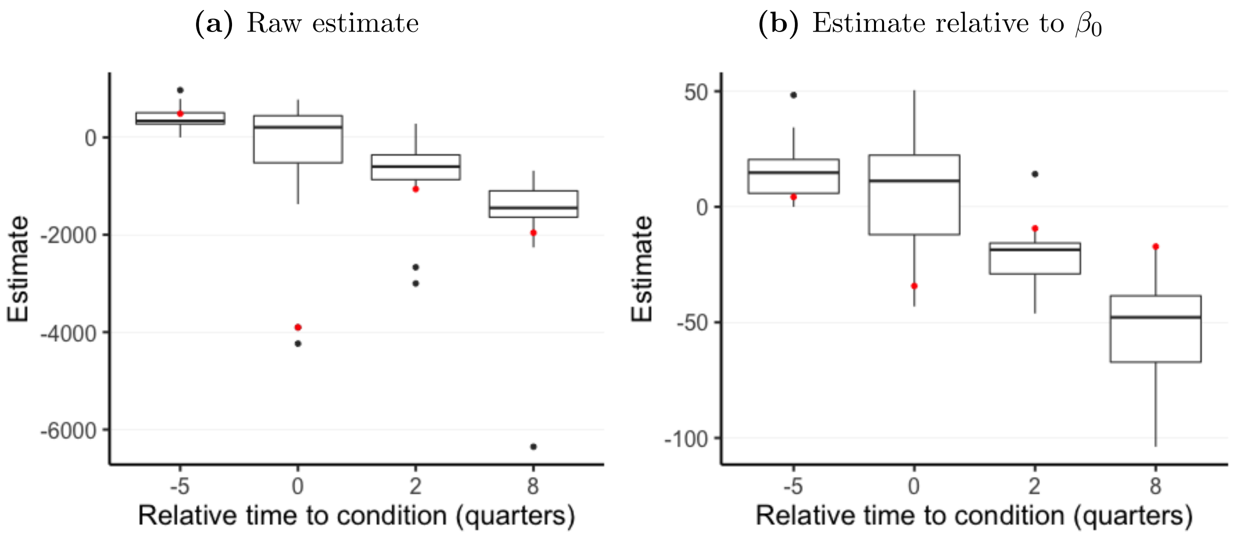
(c) By type of insurance **(d)** By sex



Notes. Each panel shows treatment effect estimates for the "present in data" dummy by groups of beneficiaries based on a specific variable. Panel (a) uses quartiles of average cost of medical services in the pre-period (i.e., $r = -5, \ldots, -2$), panel (b) uses an indicator for whether the patient had a nursing home claim in the pre-period, panel (c) splits by type of insurance at $r = -2$ (or at entry into the database for the pure control group), panel (d) splits by sex.

1.E Additional estimation results across conditions

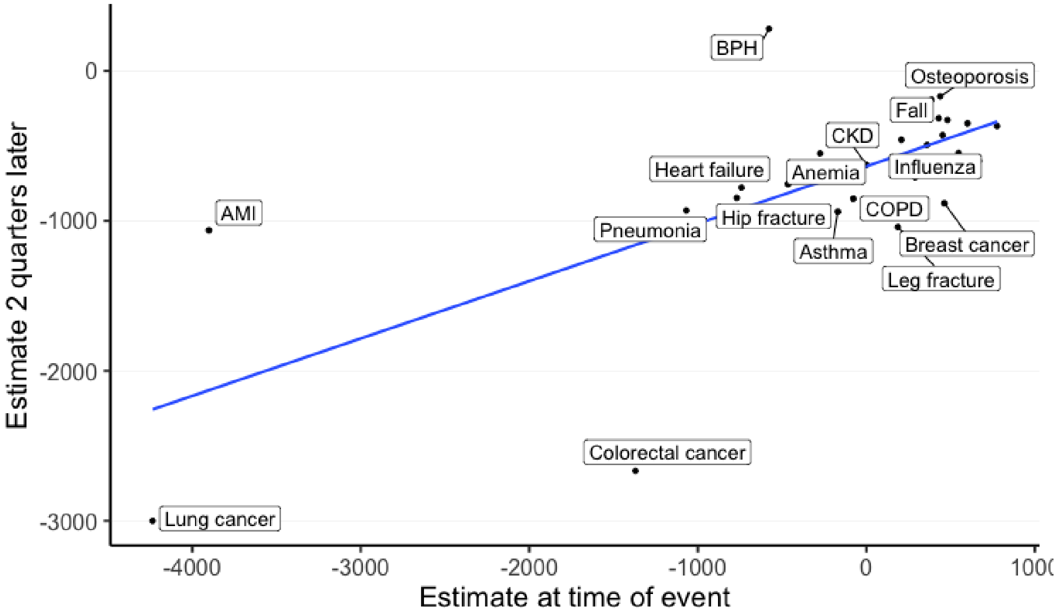
Figure 1.E.1: Distribution of treatment effects for cost of medical services across conditions



(a) Raw estimate **(b)** Estimate relative to β_0

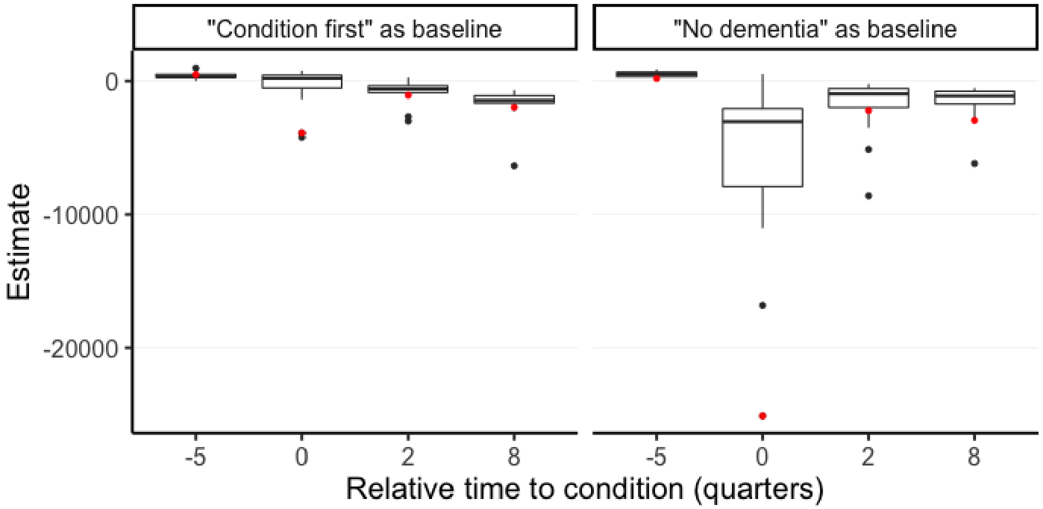
Notes. Both panels show the distribution of estimates across 27 health shocks considered for $r \in \{-5, 0, 2, 8\}$ with cost of medical services as the dependent variable. Panel (a) uses raw estimates of γ_r, while panel (b) scales these estimates by the estimated increase in care frequency at the time of event for the "condition first" group (i.e., it plots $\gamma_r/\beta_0 \times 100$), which is meant to account for differences in severity across health shocks. Red dots correspond to estimates for heart attack.

Figure 1.E.2: Immediate and 2-quarters-later treatment effect for cost of medical services



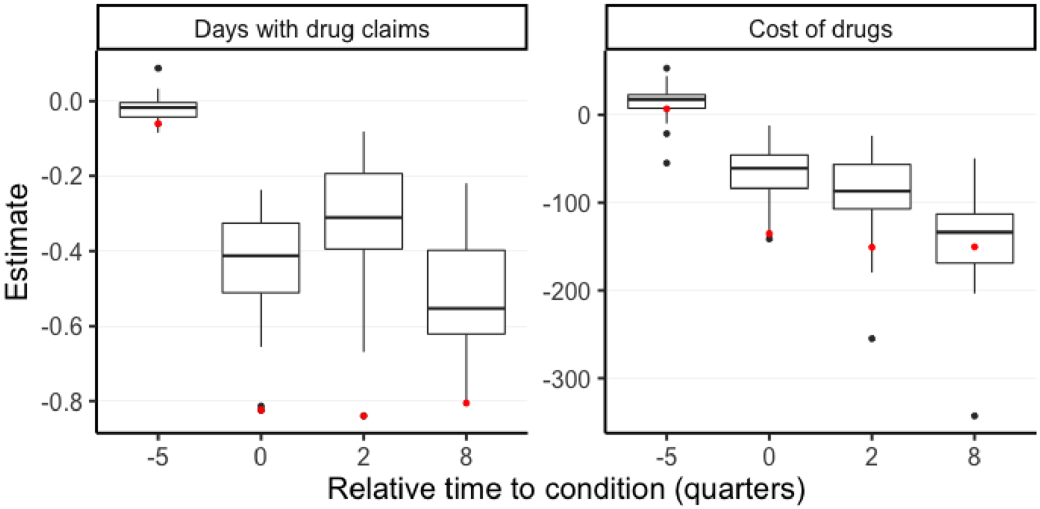
Notes. The plot shows the relationship between γ_0 and γ_2 from a regression of cost of medical services as the dependent variable across 27 health shocks considered. The linear regression line shown in blue has a slope of 0.3826 (standard error 0.0796) that is significantly different than zero.

Figure 1.E.3: Main estimate compared to using "no dementia" patients as control group for cost of medical care



Notes. Both panels show the distribution of estimates across 27 health shocks considered for $r \in \{-5, 0, 2, 8\}$ with cost of medical care as dependent variable. The right pane replaces "condition first" with "no dementia" patients as the control group, while the left pane shows the baseline estimates for reference. Red dots highlight estimates for heart attack.

Figure 1.E.4: Distribution of treatment effects for prescription drug utilization with "no dementia" patients as control group



Notes. The panels show the distribution of estimates across 27 health shocks considered for $r \in \{-5, 0, 2, 8\}$ for different outcomes with "no dementia" patients used as a control group (instead of "condition first"). Red dots highlight estimates for heart attack.

Chapter 2

Do procedure codes matter? A case study of Advance Care Planning

2.1 Introduction

Medical providers use procedure codes to record health care services supplied to the patient. In principle, each service has a unique code allowing for uniform reporting, that in turn lets insurers determine the appropriate reimbursement. Although physicians presumably aim at maximizing health benefits to patients when deciding about the quantity and type of care that they provide, financial incentives determined by the codes available and the corresponding reimbursements surely play a role. For example, Advance Care Planning (ACP), a service meant to elicit patient preferences concerning future medical care, became recognized as a procedure in 2015 when the American Medical Association (AMA) introduced a specific code for it, followed by dedicated Medicare reimbursement in 2016. As a counseling service, ACP could have been provided before 2015 under a more general procedure code[1], and the new reimbursement policy altered physicians' financial incentive to do so. This leads to the following research question: how can an introduction of a new code for an existing procedure affect cost and provision of medical services?

Little is known about how procedure code introductions affect health care cost and utilization, perhaps due to the fact that most such introductions either reflect marginal changes such as annual influenza vaccination variations, or breakthrough technologies for which the effect of the innovation cannot be easily distinguished from the effect of coding changes[2]. In the context of new medical procedures, Dranove, Garthwaite, Heard, and Wu (2022) find that when a medical procedure gets assigned a permanent and reimbursable Current Procedural Technology (CPT) code its utilization increases compared to when it was billed under a temporary non-reimbursable one[3]. Their analysis raises the possibility that financial incentives of reimbursable codes contribute to this effect, as they find no evidence of treatment effect heterogeneity when it comes to the amount of scientific evidence supporting the use of the new procedure (i.e., new codes do not seem to be treated as positive signals of efficacy).

When it comes to code introductions that concern existing procedures, such as ACP, the core challenge lies in the impossibility of measuring use of the procedure before the new code becomes available, which precludes inference concerning the *actual* use of the procedure. In

---

[1] For example, as part of the Initial Preventive Physical Examination (Barnato, O'Malley, Skinner, and Birkmeyer, 2019), also known as the "Welcome to Medicare Preventive Visit", which is a preventive visit for new Medicare beneficiaries that is only covered by Medicare once for each beneficiary.

[2] For example, in 2009 new procedure codes for transcatheter aortic valve replacement (TAVR) have been introduced due to their substantial survival benefits for people suffering from severe aortic stenosis that revolutionized the care of this condition.

[3] New medical procedures first appear under Category III CPT codes and only become reimbursable if subsequently assigned a Category I CPT code.

recent years a few such codes have been introduced[4] with the hope of encouraging provision of specific counseling services. So far, the medical literature focused on quantifying the use and cost of these procedures when they are explicitly billed using the new code (Belanger, Loomer, Teno, Mitchell, Adhikari, and Gozalo, 2019; Li, Andy, and Mitchell, 2021). However, such analyses are not informative about the aggregate effect on health care cost, as reporting provision of these services explicitly probably leads to reductions in the utilization of more general billing codes that have been used to report them before. Moreover, as such codes are usually introduced in conjunction with dedicated reimbursements, this results in changes to physicians' financial incentives and so can also affect provision of other medical services.

In this paper I quantify the effect of physicians adopting a new procedure code in the specific case of ACP by estimating the change in physicians' revenue upon adoption of the new code using a dynamic difference-in-differences design (event study). ACP is a 30 minute face-to-face counseling session during which the physician and patient discuss advance directives that document the patient's preferences concerning medical care in the event of a serious illness or a lack of capacity for decision-making. In 2015 the AMA introduced a CPT code for ACP and in 2016 Medicare started reimbursing it. I assume that before this policy took effect, the procedure was performed under other Evaluation and Management (E&M) codes that are used for recording other counseling services. Then, I restrict attention to physician specialties who are most likely to use the new code and show that those who did and did not adopt it exhibited parallel trends of E&M revenue before adoption, suggesting that from the perspective of E&M revenue physicians who did not use ACP are a good

---

[4] CPT code 99490 for Chronic Care Management was introduced in 2015, CPT code 99483 for Cognitive Assessment and Care Planning Services was introduced in 2018.

comparison group for those who did and allowing for causal inference using difference-in-differences.

For physicians specializing in hospice and palliative care, geriatric medicine, and hospitalists, I find that ACP adoption leads to an increase in E&M revenue equal to $802 per quarter which corresponds to 16% of the baseline mean. What is more, each quarter adopters provide ACP to 2 patients on average, which accounts for only $151 of the estimated increase in revenue, while the remaining $651 reflects increased use of other E&M services, such as optional screening for depression and alcohol abuse. I show that 75% of this amount can be attributed to care of patients who receive ACP from the physician in question. Moreover, although the aggregate effect of ACP adoption on E&M revenue fades over time and becomes insignificantly different from zero after 2.5 years, physicians increase the amount of revenue that they obtain from ACP patients over time. Specifically, patients who receive ACP more than a year after their physician's adoption of the procedure exhibit higher utilization increases than those who received it within a year of adoption. Finally, by leveraging differences in timing of ACP between patients of the same physician I estimate small decreases in their use of non-E&M services in the long term.

To identify the causal effect of a physician starting to use dedicated billing codes for ACP on their provision of medical services I assume that conditional on observed characteristics of the physician, the timing of adoption of the code is as good as randomly assigned. In 2015 the AMA introduced the new CPT codes in question: 99497 and 99498 for the first and each subsequent 30 minutes of ACP services, respectively. However, until a year later the codes were not covered by Medicare, and scarcely used. In 2016 Medicare introduced a reimbursement policy for ACP, which makes all of its beneficiaries eligible to receive the service each year. Crucially, if provided in conjunction with an annual wellness visit, a preventive service, ACP is free to the patient. Despite the generosity of this policy, between 2016 and 2019 only 3% of all providers billed for ACP, raising the possibility of selection of physicians into using the code, that would invalidate a comparison between adopters and non-adopters. Thus, I restrict attention to the three specialties that are most likely to adopt ACP (23% adoption rate), and show that they exhibit parallel trends in E&M revenue before adopting the code, giving credence to the assumption that adoption among them is as good as random. Moreover, even among physicians who specialize in hospice and palliative care of whom almost half adopt the code, adoption is staggered throughout the study period, pointing to the possibility that physicians cannot freely choose when to use if (e.g., it could be driven by the type of patient that a physician sees[5]).

In my analysis, I use the 20% random sample of Carrier claims of Traditional Medicare beneficiaries that reflects the use of physician services among the relevant population. I focus on E&M revenue as my main outcome variable, as it encompasses all services that are likely substitutes of ACP, so could have been used to bill for the service before the specific code was available. In my event study analysis I define physician's time of adoption as the first quarter in which they used a CPT code for ACP, regardless of whether they received reimbursement for it. I estimate the trajectory of E&M revenue with and without ACP around the time of adoption to show that using the new code leads to substantial increases in revenue from

---

[5] Some patients might be more "appropriate" for receiving ACP, for example due to their health state – diagnosis codes for hypertension, hyperlipidemia and chronic obstructive pulmonary disease are among the most common codes on ACP claims.

E&M services other than ACP. Further, I split each physician's patients into groups based on the timing of their ACP visit, and show that the aforementioned increases are driven by the recipients of ACP services around the time when they received the service. Moreover, I find no evidence of pre-ACP differences between patients who received the service at different times. In other words, all patients of a physician who ever received ACP exhibit parallel trends in their utilization before the service was provided, allowing for inference based on a comparison between patients who receive the service at different times with respect to a physician's adoption time.

A major concern when relying on a comparison between physicians who did and did not adopt ACP comes from potential selection of physicians into treatment (i.e., adoption) as the reasons for adoption are poorly understood. To address this issue, I estimate a propensity score model for ACP adoption in order to make the treated and control groups of physicians more comparable using inverse propensity score weights. Alternatively, I remove physicians who did not adopt ACP from the design, and reduce the number of fixed effects included in the model in order to identify the treatment effect in question. I show that for the full population of physicians neither of these approaches lead to a fair comparison between adopters and non-adopters as the two groups are significantly different in the pre-period. Thus, I focus on a subset of three physician specialties for whom the parallel trends assumption does not seem violated and show that both the baseline and "adopters only" specifications yield similar results. The estimated effect is somewhat smaller but qualitatively similar if I only consider hospice and palliative care specialists, for whom balance on covariates is satisfied for a wider variety of outcomes in the reference period.

Physician pay depends on the services they provide and the characteristics of their patients as age or health status can heavily impact the complexity of any given procedure. Extensive work studies the practice of quantifying a patient's health state through diagnosis codes. In particular, Dafny (2005) shows that changes in reimbursement for patients of higher complexity leads to hospitals "upcoding" patients (i.e., exaggerating their complexity), a way of gaming the system to extract more funds from the insurer. The extent to which providers engage in such revenue-enhancing practices varies across hospital types (e.g., for-profit vs others) and can be facilitated by, for example, the use of health information technology (Agha, 2014). However, Sacarny (2018) uses a 2008 reform that allowed providers to raise revenue by providing more detailed documentation to show that while around a third of providers adopted this practice right after the reform, others were slow to follow suit (after three years only 50% had adopted it), suggesting that hospitals' ability to maximize revenue is far from perfect. Thus, the literature documents the fact that providers respond to financial incentives determined by the relationship between diagnosis codes and reimbursement, though large variance in the timing of such responses persist, with some physicians failing to do so at all.

When it comes to procedure codes, they are meant to enable uniform reporting of provision of medical services, but interest groups who advocate for introduction of new codes are likely motivated by either reimbursement or insurance coverage motives. When arguing in favor of introducing the specific code for ACP, the Illinois State Medical Society refers to evidence of communication gaps between patients and physicians (Tulsky, 2005) and the cost-saving potential of ACP (Teno, 2000), which suggests that they expected the code introduction and subsequent reimbursement to affect provision of it. In fact, both the AMA and the American Academy of Palliative and Hospice Medicine stated that this change will encourage physicians to engage in ACP (Zeitoun, 2015). However, Belanger et al. (2019) show that utilization patterns of the procedure remain low, especially among specialties that typically care for seriously ill patients, which underscores the possibility that ACP use relies on less specific billing codes, calling into question the effect that introducing the specific code had on actual provision of the service. At the beneficiary level, Reich, Jin, Gupta, Kim, Lipstiz, Prigerson, Tjia, Ladin, Halpern, Cooper, and Weissman (2020) show that Medicare patients who are most likely to receive ACP services are both seriously ill and frail, yet even among them the take-up of this service is only 5.2%. However, provision of ACP does seem targeted at a specific type of patient, as Barnato et al. (2019) found that in 2017 48% of beneficiaries who had ACP for more than 30 minutes died within a year.

The ACP service itself consists of a discussion of a patient's wishes when it comes to end-of-life care, and care in the event of loss of decision-making ability, which are recorded using advance directives. In the United States, the two major types of such documents are living will and durable health care power of attorney. A living will specifies preferences concerning the use of life-sustaining and other medical interventions. Durable health care power of attorney designates a surrogate decision-maker that can act on behalf of the patient. Yadav, Gabler, Cooney, Kent, Kim, Herbst, Mante, Halpern, and Courtright (2017) estimate that about one third of adults in the US completed either a living will or durable health care power of attorney, although the enforceability of such documents and the resulting compliance varies widely across states (Kessler and McClellan, 2004). Moreover, various controversies surround the use of advance directives. For example, Fagerlin and Schneider (2004) oppose the use of living wills, as they argue that people rarely have the expertise to make informed decisions about hypothetical health-related scenarios, while question phrasing affects answers to such an extent, that it is doubtful that living wills reflect actual preferences. Byrne and Thompson (2000) develop a theoretical model that leads them to conclude that binding advance directives are medically unethical as in some situations they will lead to underprovision of care with respect to the patient's true preferences.

In this project I investigate the role of new procedure codes in the provision of medical services using ACP as a case study, which contributes to the literature on physician decision-making and financial incentives (Dafny, 2005; Dranove et al., 2022), by identifying three core mechanisms through which procedure codes affect outcomes. First, if a dedicated reimbursement for a specific service changes physician's net revenue when providing the service under the new code as opposed to the old one, physicians might increase their revenue by changing the way they allocate time to the provision of other services. Second, if the new code in question is associated with different insurance coverage, physicians might increase their provision of the service even if the corresponding reimbursement remains constant with respect to using a previous code. In particular, services with lower beneficiary copayment are favored as they are perceived as less costly by the patient and so are less likely to lead to changes in patient demand for future elective care. Finally, either of the aforementioned effects can lead to changes in the actual provision of the service underlying the new billing code, instead of merely recording its provision with a specific code instead of a general one. In turn, provision of counseling services such as ACP can affect future health care utilization. For example, filling out advance directives might lead the patient to contemplate her preferences for medical care and opt for less invasive elective procedures in the future.

Second, my analysis contributes to the recent literature concerning Medicare's dedicated reimbursement of ACP (Barnato et al., 2019; Belanger et al., 2019), as it the first to provide causal estimates of the effect of this policy. First, using a stylized example, I demonstrate how the reimbursement in question can change physician's revenue for a standard office visit involving ACP, which can lead to the provision of additional E&M services. Second, I document large increases in the use of E&M services other than ACP at the time of ACP adoption, that point to the importance of investigating potential complementarities between various services when quantifying the effect of new reimbursement policies on health care cost. This paper provides a framework for such analysis in the context of other recent Medicare policies, such as reimbursement of cognitive assessment and care planning services studied by Li et al. (2021).

Third, I contribute to the vast literature studying the effect of the provision of ACP services on health care utilization. Brinkman-Stoppelenburg, Rietjens, and Heide (2014) conduct a systematic review of past studies to show that the bulk of existing evidence relies on observational studies performed in institutional settings (e.g., hospitals, ICUs, nursing homes), often focusing on specific kinds of advance directives such as do-not-resuscitate or do-not-hospitalize orders. I use plausibly exogenous variation in the timing of one's ACP visit to show increases in E&M care utilization in the short-term and possible decreases in other types of physician care in the long-term. My estimates concern the effect of a very broadly defined ACP service among a specific population of Medicare beneficiaries that did eventually receive the service. Moreover, the effect on outcomes of interest such as hospital deaths or hospice use is beyond the scope of my analysis.

Finally, I contribute to the substantial literature investigating the drivers of treatment decisions and spending at the end of life, which aims at establishing whether better physician-patient communication could improve resource allocation. In particular, Cutler et al. (2019) use nationally representative survey data to show that although physicians' subjective beliefs about appropriateness of different procedures explain a lot of geographic variation in health care spending at the end of life, patients' preferences for care intensity seem not to matter. This is surprising, as in general, care over which patients have more discretion (e.g. elective procedures) tends to correlate with patient preferences more strongly than other care (e.g. emergency care), as documented by Finkelstein et al. (2016) and there seems to be a consensus that patients should be more involved in medical decision-making at the end of life (Gawande, 2016). In a recent study, Lustbader, Mudra, Romano, Lukoski, Chang, Mittelberger, Scherr, and Cooper (2017) show that enrollment in a home-based palliative care program that involved conversations about goals of care is associated with substantial cost savings and higher hospice utilization. I show that patients who receive ACP exhibit slight decreases in non-E&M care in the long term, pointing to possible reductions in the use or intensity of subsequent elective procedures.

The remainder of the paper is structured as follows. In section 2.2 I provide the institutional background as well as details of the reimbursement policy in question, the data environment, and summary statistics of ACP adoption. I report the results of my empirical analysis in section 2.3, including estimates for aggregate physician revenue, as well estimates specific to patients who benefited from ACP. Section 2.4 concludes.

2.2 Data and methods

This section sets the stage for my empirical analysis. I start in section 2.2.1 by providing background on Traditional Medicare and the sample of its providers that I use in my analysis. Then, in section 2.2.2 I discuss the purpose and importance of procedure codes. Section 2.2.3 defines Advance Care Planning (ACP) and briefly discusses former literature concerning its effect on health care utilization. In section 2.2.4 I describe Medicare's policy of reimbursing ACP that is central to my analysis. Finally, in sections 2.2.5 and 2.2.6 I outline my research design and the possible mechanisms at play, respectively.

2.2.1 Traditional Medicare claims data

This paper uses Traditional Medicare (TM) claims data from the Centers for Medicare and Medicaid Services[6] (CMS) that includes claim-level information for all TM enrollees in 2010-2019, which encompasses around 37 million people each year. TM is a federally administered health insurance plan for the Medicare-eligible population (i.e., those aged 65 and older or suffering from certain disabilities) that consists of parts A and B that cover hospital and medical care, respectively. The key feature of TM is that it operates on a fee-for-service (FFS) basis, so that the plan reimburses health care providers for each medical service performed based on a pre-determined fee schedule. Historically, TM served the vast

---

[6] Access to CMS data for this project has been granted through the National Bureau of Economic Research (NBER) under DUA #57953 that is reusing data included in NBERs DUA #16702 for the larger study titled "Improving Health Outcomes for an Aging Population".

majority of Medicare enrollees, although their share has been steadily decreasing throughout the study period from 73% in 2013 to 63% in 2019 (KFF, 2023) with the rise of privately administered insurance plans known as Medicare Advantage. This decrease presumably reflects the ongoing effort to move away from the FFS model in order to address the concern that it contributes to exorbitant health care costs in the US by incentivising provision of medical services without regards to their necessity or quality (Cutler, 2015). However, despite its loss of importance, TM remains the largest health insurance plan for the elderly in terms of enrollment and studying the way it affects provision of various services remains a crucial policy concern.

As my analysis concerns Advance Care Planning (ACP) and other medical services provided by physicians it relies on a 20% random sample of TM Carrier claims, which include claims submitted by professional providers such as physicians, physician assistants and nurse practitioners. The sample for years 2010-2019 contains records of 1.75 million unique providers[7], of whom I exclude 73 thousand that have a specialty that is irrelevant for my analysis[8]. For ease of computations I further restrict attention to 53 thousand physicians that used ACP at least once[9] and a random sample of 212 thousand of the remaining physicians such that the final analysis sample contains 265 095 providers, of which 20% adopted ACP in the time frame of my analysis. For thus selected physicians I construct a quarterly

---

[7] The TM database uses National Provider Identifiers (NPIs) to identify providers that can correspond to individual physicians as well as clinics, laboratories, institutional providers or suppliers of medical equipment.

[8] I compile a list of specialties appearing on ACP claims and restrict attention to physicians who appeared under one of those specialties for at least one quarter (specialty assignment to physicians is weighted by the number of claims in a quarter with that specialty). The 73 thousand physicians who did not satisfy this condition filed the majority of their claims under specialties such as optometrists, audiologists, speech language pathologists or oral surgeons, which are clearly irrelevant for ACP provision.

[9] I exclude one provider 77 providers who billed for it in 2015 when the ACP procedure code existed without a dedicated reimbursement.

panel based on the first and last quarter in which each physician appears in the database, and fill it with quarterly aggregates of service provision when available and zeros otherwise.

The data come with three important caveats. First, I only observe a 20% sample of Carrier claims, so at any point in time I am likely misclassifying some physicians who used ACP as one's that did not[10], which attenuates any estimates based on a comparison between the two, and so my estimates should be interpreted as lower bounds on the underlying treatment effects. Second, my analysis is restricted to TM and cannot speak to the possible effect of changes in financial incentives in TM on the relative allocation of time and effort between TM and non-TM patients of the same physician. Thus, my conclusions concerning the effect of ACP on health care cost are not informative of aggregate health care cost as increases in health care utilization among TM patients could be offset by symmetric decreases among non-TM patients. Finally, the use of insurance claims leads to an inherently limited understanding of changes in provision of health care services. In particular, my analysis documents changes in their use as reported for reimbursement purposes, but cannot speak to how this related to actual services nor their quality. Despite these caveats, claims data offer the most accurate way of measuring health care cost so are uniquely suited for investigating the effect of ACP reimbursement on TM cost.

---

[10] I expect this misclassification rate to be much lower than 80% as physicians treat multiple patients and it is enough for one of their ACP claims to be in the 20% sample for me to classify them as an ACP-user in that quarter.

2.2.2 Procedure codes

Outside of hospitals[11], medical providers report procedures, supplies, products and services administered to patients using the Healthcare Common Procedure Coding System (HCPCS), which is divided into two coding subsystems: level I consists of Current Procedural Terminology (CPT) codes that mostly identify medical services and procedures provided by physicians, and level II codes corresponding to products, supplies and services not included in level I, such as ambulance services and durable medical equipment (CMS, 2023). As the focus of this paper is provision of medical services by physicians, I concentrate on CPT codes, issued and maintained by the American Medical Association (AMA). Importantly, procedure codes enable consistent reporting of health care provision that is necessary for its reimbursement. Crucially, insurers rarely cover procedures that do not fall under existing codes (Dranove et al., 2022), so reimbursement can be thought of as the primary reason for the existence of CPT codes despite their stated purpose of offering a "uniform language (...) to streamline reporting" (AMA, 2023a).

As the sole administrator of its proprietary CPT codes database, the AMA meets three times per year to review requests for code introductions and modifications, brought forward by various entities such as medical societies. As the codes are only meant to reflect the use of procedures rather than validate their appropriateness[12], approval of new codes does not parallel that of drug approval by the FDA where efficacy must be corroborated in a sequence of clinical trials that satisfy specific criteria (Dranove et al., 2022). When it comes

---

[11] Inpatient care is recorded using a different coding system.

[12] In fact, the stated criteria for a code to be approved mostly concern showing its current or intended utilization, while proof of potential efficacy is optional and can constitute of merely a description of an ongoing trial that investigates it (AMA, 2023b).

to the interest groups that request code changes, one could expect that their motivation goes beyond improving the precision of medical billing records and involves intended changes to reimbursement or insurance coverage, or both. In fact, the Illinois State Medical Society that applied for the introduction of CPT codes for ACP, lists reimbursement as the leading reason for their request (AMA, 2013).

2.2.3 Advance care planning

Advance care planning (ACP) is the process of preparing for future decisions concerning medical care in case of a serious illness or loss of decision-making capacity. It can entail informal discussions of one's preferences for care as well as completion of formal advance directives that are legal documents recording such preferences. Yadav et al. (2017) estimate that approximately one third of adults in the United States have completed an advance directive in the form of a living will (i.e., record of one's desire for withholding death-delaying procedures in the event of a terminal condition) or durable power of attorney for health care (i.e., designation of another person who can make decisions concerning medical treatment if one is no longer able to do so). In addition, one can choose from a variety of documents specifying one's preferences for the use of specific medical procedures such as resuscitation (do not resuscitate order), hospitalization (do not hospitalize order) or life-sustaining treatment (physician/medical order for life-sustaining treatment, or POLST/MOLST).

A crucial consideration in the context of advance directives is that incentives to follow them and penalties for failing to do so vary across states. Kessler and McClellan (2004) track states' adoption of laws that enforce compliance with health directives (e.g., deem them legally binding or specify penalties for disregarding them) or require delegation of decision-making concerning medical treatment if no advance directive is present. Although by 1992 all states passed some form of law incentivising compliance, practical implications of these laws vary widely. For example, Oregon also legalized assisted dying and pioneered the use of a POLST registry that allows health professionals from various health care settings to access one's advance directives, which possibly underlies the fact that the majority of deaths in Oregon happen at home, unlike in the remaining states (Tolle and Teno, 2017). However, in many settings existing advance directives are not accessible or not followed by decision-makers, contributing to vocal criticism of laws that encourage their use (Fagerlin and Schneider, 2004).

Although a sizable literature attempts to establish the link between ACP, treatment decisions and health care cost, no consensus has been reached. Brinkman-Stoppelenburg et al. (2014) perform a systematic literature review to show that ACP is associated with decreases in life-sustaining treatment and hospitalizations, but increases in the use of hospice and palliative care. However, a causal interpretation of these findings is questionable, as 95% of the studies reviewed were observational, and many of them use convenience samples that are hardly representative of the general population. Moreover, experimental evidence from the large-scale Study Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT) failed to show an effect of increasing documentation of pre-existing advance directives on hospital resource use, casting doubt on the potential of ACP to limit health care costs (Teno, Lynn, Connors Jr., Wenger, Phillips, Alzola, Murphy, Desbiens, and Knaus, 1997). Kessler and McClellan (2004) corroborate this by using adoption of laws related to advance directives to show that such laws do not reduce medical expenditures, and vary in their effect on health care outcomes: laws encouraging compliance lead to less hospital deaths while laws requiring appointment of surrogate decision-makers increase the provision of acute care at the end of life.

2.2.4 ACP procedure code

On January 1st, 2016 Medicare started reimbursing ACP under CPT codes 99497 (first 30 minutes) and 99498 (each additional 30 minutes) (Zeitoun, 2015). Although these CPT codes were approved by the AMA a year earlier, without dedicated reimbursement from Medicare they were scarcely used[13]. In order to report the use of ACP under the code 99497, the physician is required to spend at least 16 minutes face-to-face with the patient, family member or surrogate, while explaining and discussing advance directives and potentially completing them (Medicare Learning Network, 2023). Although using the procedure requires specific documentation (e.g., specifying its length, listing those present), it can be performed on any Medicare beneficiary, regardless of their health and the accompanying diagnosis. Moreover, Medicare waives patient cost-sharing for ACP if it is provided as part of one's Annual Wellness Visit[14] as a preventive service. Unsurprisingly, of the 1.1 million ACP claims present in my sample almost 45% also include procedure codes for the Annual Wellness Visit (see Appendix Table 2.A.1).

---

[13] In the 20% Carrier claims sample used for this analysis these codes were used 276 times in 2015.

[14] The Annual Wellness Visit (HCPCS codes G0438 for the first and G0439 for a subsequent visit) is an annual appointment with one's primary care provider aimed at updating a personalized prevention plan, which is fully reimbursed by Medicare (but deductible and coinsurance may apply to additional services provided at that visit).

Table 2.2.1: Use of the ACP procedure code by year

Year	2016	2017	2018	2019
Providers ('000)	14.21	21.89	28.42	34.49
Patients ('000)	120.38	201.13	257.90	317.53
Average patient age	75.98	75.94	75.98	76.22
Total allowed amount ($M)	10.50	17.62	23.94	31.10
Allowed amount per paid procedure ($)	82.98	79.53	81.98	82.08
Allowed amount share of whole claim (%)	30.90	30.21	30.76	30.36
Claims with zero allowed amount (%)	13.78	11.53	10.55	9.33

Notes. Statistics based on claims for Advance Care Planning (CPT codes 99497 and 99498). Allowed amount per paid procedure only includes CPT code 99497 (first 30 minutes) that had a non-zero allowed amount. Allowed share of claim indicates the average percentage of a medical claim that is attributed to ACP among claims that included ACP.

I report summary statistics on the use of ACP in Table 2.2.1. Patients who receive ACP are 76 years old on average and each ACP service is priced by Medicare at around $82 on average. Moreover, a significant but declining fraction of ACP claims appear with an allowed amount of $0, implying improper use of the procedure, which suggests that providers are learning how to properly bill for ACP. Strikingly, only a small fraction of all providers use ACP, which is in line with other early tabulations of the use of this procedure, often met with disappointment from researchers expecting the procedure to be widely adopted the introduction of Medicare's reimbursement policy (Wright, 2018; Belanger et al., 2019). However, it is important to note that each year 6-7 thousand new providers adopt the procedure, which resulted in the doubling of the number of adopters between 2016 and 2018 and suggests that further growth can be expected. Nevertheless, even among the 2.8 thousand physicians in my data who specialize in hospice and palliative care only 48% billed for ACP at least once during the study period (see Appendix Table 2.A.2). Thus, as noted by Belanger et al. (2019), using dedicated CPT codes for ACP seems likely to be a poor proxy for actual use of ACP services as physicians might discuss goals of care with their patients as part of other interactions, possibly under more general CPT codes such as those for a regular office visit.

2.2.5 Empirical strategy

The goal of my empirical analysis is to identify the effect of ACP adoption (i.e., starting to use CPT codes for ACP) on a physician's[15] provision of health care services. To achieve this, I rely on a two-way fixed effects event-study design that leverages variation in time of ACP adoption between physicians to estimate the trajectory of differences in health care provision between adopters and non-adopters around the time of adoption. I primarily focus on provision of Evaluation and Management (E&M) services other than ACP as physicians can substitute ACP for more general CPT codes and at the same time take advantage from possible complementarities between ACP and other E&M procedures (see discussion in section 2.2.6). I account for time-invariant physician characteristics and time trends common to all physicians by including individual and calendar-time fixed effects, respectively. This leads to the following regression equation:

$$\text{E\&M revenue}_{it} = \alpha_i + \tau_t + \sum_{r \neq -1} \beta_r + \varepsilon_{it}, \tag{2.1}$$

where i indexes physicians, t calendar time (year-quarters), r corresponds to the relative time to ACP adoption with $r = -1$ as the reference period, and standard errors are clustered at

---

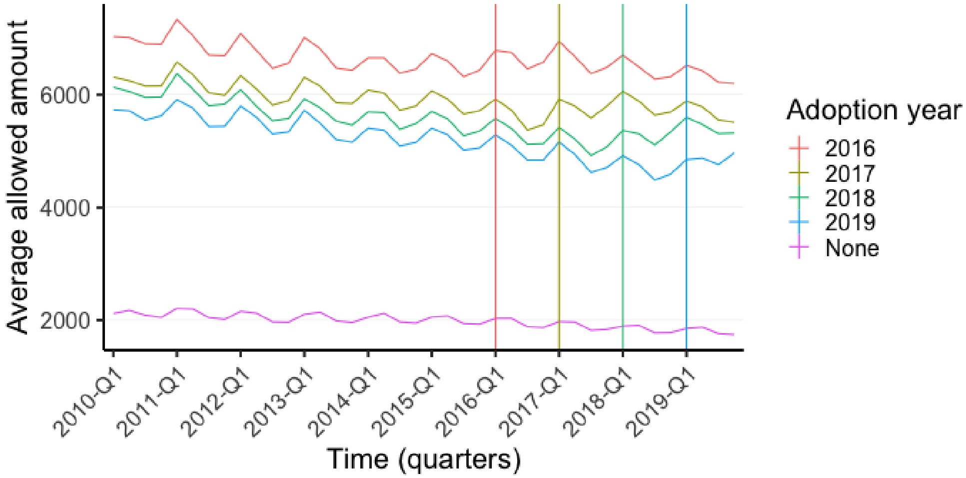
[15] Although I include both individual physicians and other types of health care providers (e.g., clinics), I use the term "physician" and "provider" interchangeably.

physician level. Importantly, in order to be able to identify both calendar-time and relative-time fixed effects, I also include physicians who never used ACP (i.e., the control group) and assign them to the reference period such that their revenue patterns contribute to the estimation of calendar-time but no relative-time fixed effects. As a robustness check, I also estimate regression (2.1) without the control group, which precludes inclusion of all three types of fixed effects, so instead of physician fixed effects I control for their 5-digit zip code. E&M revenue is the sum of allowed amounts for E&M services *excluding ACP* billed by physician i in quarter t. To accurately capture possible decreases in provision of services, I impute zeros for all quarters between the first and last time that each physician appeared in the database. In further analysis I also consider other outcome variables, such as the number of patients or provision of specific E&M services such as routine office visits.

A causal interpretation of my event-study estimates relies on three assumptions: no anticipation, parallel trends in baseline outcomes and treatment effect homogeneity across different times of ACP adoption. Clearly, beyond the plausibly exogenous introduction of the Medicare reimbursement policy in 2016, physicians choose whether and when to use the new CPT codes which could violate the aforementioned assumptions. First, no anticipation means that physicians do not experience systematic changes to their medical practice before starting to use ACP, which would be violated if prior decreases in E&M provision lead physicians to look for further sources of revenue and learn about the new code. Second, the parallel trends assumption implies that the revenue trajectories of ACP adopters would have remained parallel to those of non-adopters if they have not adopted the procedure. This is violated if adoption is driven by unobservables correlated with one's provision of E&M

services. For example, if E&M services are only a small part of one's revenue, staying up to date with coding procedures might not be worth the effort. Finally, physicians who adopt ACP earlier could be the ones with the highest number of patients who would benefit from the procedure, which would make the response of their provision to ACP adoption strongest, violating the assumption of homogenous treatment effects.

Figure 2.2.1: Average quarterly E&M revenue by year of ACP adoption



Notes. Total E&M revenue, including ACP. Adoption year is the year of a physician's first claim that included the CPT code 99497 or 99498.

In Figure 2.2.1 I plot the evolution of E&M revenue by year of ACP adoption. Before the introduction of Medicare's reimbursement policy the graphs shown in the plot seem to evolve in parallel, and this pattern remains true for adopters before their time of adoption. Although the parallel trends assumption concerns counterfactual outcomes, which are unobserved, parallel trends in realized outcomes before adoption are reassuring. Moreover, the fact that earliest adopters also have the highest baseline levels of E&M provision raises the possibility that the use of ACP, presumably profitable only for a very specific kind of patient, is driven by the physician interacting with the right kind of patient, which likely happens more often among physicians who see the most E&M patients. Crucially, E&M revenue for physicians who never adopted ACP is much lower than that of adopters, which contributes to concerns about using them as a comparison group. In order to make adopters and non-adopters more comparable, I consider sample restrictions based on physician specialty as well as weighting the sample using inverse propensity score weights.

In Table 2.2.2 I show average differences between physicians who did and did not adopt ACP. Clearly, these two groups of physicians differ significantly along all dimensions considered. In particular, adopters see almost 50% more patients than non-adopters and they provide much more E&M services, while somewhat less of other outpatient services. Moreover, their patients are on average almost 70 years old, compared to only 61 for non-adopters. Among adopters' patients more than four times as many are diagnosed with conditions associated with ACP[16] such as hypertension and chronic obstructive pulmonary disease, which can be driven by differences in health status of their patients as well as differences in diagnostic practices of the physicians in question. All of these differences shrink and become insignificant or of opposite sign once I restrict attention to physicians who specialize in hospice and palliative care, suggesting that the comparison between adopters and non-adopters is less problematic. Unlike specialty restrictions, using inverse propensity score weights does not lead to major improvements when it comes to balancing the two groups on observables so I defer the relevant discussion to section 2.B of the Appendix.

Of course, none of the assumptions of my analysis can be verified directly, and the summary statistics reported above are only suggestive. I complement them by formal tests

---

[16] See Appendix Table 2.A.1 for a list of diagnosis codes that are most common on ACP claims.

Table 2.2.2: Differences between adopters and non-adopters

	Control	Adopter	Diff.	
A. All physicians				
E&M revenue ($)	1691.29	5448.04	3756.75	***
Non-E&M revenue ($)	3739.27	3274.27	−465.01	***
Patients	23.93	32.66	8.73	***
Patient average age	61.34	69.85	8.52	***
Patients with COPD	0.68	3.16	2.48	***
Patients with hyperlipidemia	1.50	5.84	4.34	***
Patients with hypertension	4.63	20.71	16.08	***
B. Top 3 specialties among ACP adopters				
E&M revenue ($)	3687.62	4883.02	1195.40	***
Non-E&M revenue ($)	501.63	628.11	126.48	
Patients	18.68	24.62	5.94	***
Patient average age	68.84	74.83	5.99	***
Patients with COPD	1.80	2.26	0.47	
Patients with hyperlipidemia	2.91	3.37	0.46	
Patients with hypertension	12.73	14.76	2.02	
C. Only hospice and palliative care specialists				
E&M revenue ($)	2061.20	2491.05	429.85	
Non-E&M revenue ($)	292.80	58.43	−234.37	
Patients	10.25	10.12	−0.14	
Patient average age	64.38	70.75	6.38	*
Patients with COPD	0.97	0.51	−0.45	
Patients with hyperlipidemia	0.31	0.23	−0.08	
Patients with hypertension	1.16	1.48	0.31	

Notes. Based on values for last quarter of 2015, immediately before ACP became reimbursed by Medicare. Top 3 specialties among ACP adopters refers to specialties with the highest propensity of adoption, which are: hospice and palliative care, geriatric medicine and hospitalist (see Appendix Table 2.A.2). Patients who have hyperlipidemia, hypertension and COPD (chronic obstructive pulmonary disease) are counted using diagnosis codes E785, I10 and J449, respectively. For more outcomes see Appendix Table 2.A.3.

of anticipation, parallel trends in observed outcomes as well as treatment effect homogeneity using my regression framework. Whenever I find evidence of assumption violations, I leverage differences in the timing of ACP across patients of the same physician to estimate the causal effect that ACP use has on provision of E&M services for the specific kind of patient. To do this, I estimate the following regression equation:

$$\text{E\&M revenue}_{itg} = \tilde{\alpha}_i + \tilde{\tau}_t + \sum_{r \neq -1} \left(\tilde{\beta}_r + \tilde{\gamma}_r \text{Early ACP patients}_g + \tilde{\delta}_r \text{No ACP patients}_g \right) + \tilde{\theta}_1 \text{Early ACP patients}_g + \tilde{\theta}_2 \text{No ACP patients}_g + \tilde{\varepsilon}_{itg}, \quad (2.2)$$

where g indexes the following three patient groups: did not have ACP ("no ACP"), had ACP in the first year since their physician's adoption ("early ACP"), had ACP more than one year after their physician's adoption[17] ("late ACP"), which is the reference group in equation (2.2). The treatment effect of interest is captured by parameters $\tilde{\gamma}_r$ for $r \in \{0, \ldots, 3\}$ as this before "late ACP" patients had ACP and isolates the difference in health care provision between patients who are already affected by ACP and those who will only be affected in the future. The underlying identification assumption states that the exact timing of a patient's ACP visit with respect to physician's adoption time is as good as randomly assigned, making the "late ACP" patients a good comparison group for "early ACP" patients (e.g., ACP could be driven by plausibly exogenous times at which patients show up for health check-ups). However, if physicians provide ACP first to patients who provide the best opportunity for raising revenue, this approach will provide estimates that are biased upwards as the revenue-raising potential of ACP is exaggerated. Using standard tests for possible violations of the parallel trends assumption I show that this seems unlikely.

Regression (2.2) serves two purposes: it uses a different source of variation for causal estimates and sheds light on mechanisms underlying ACP's effect on provision of health care

---

[17] I also estimate a variation of regression (2.2) that splits "late ACP" patients into those who had ACP between 1 and 2 years after the physician's adoption and those who had it more than two years afterwards.

services. First, in settings where the parallel trends assumption for regression (2.1) fails, I provide causal estimates of the effect of having ACP earlier versus later, which demonstrates how ACP use affects provision patterns of different patients – a better identified result for a more specific treatment effect. Second, after changing the reference group to "no ACP" I show that for patients who had ACP the timing of increases in provision of health care services aligns precisely with when they had ACP. This sheds light on the mechanism by which the use of ACP leads to increases in the use of other services. In particular, I corroborate the hypothesis that ACP compements other E&M services performed on the same patient, as discussed in the following section.

2.2.6 Mechanisms

In principle, before the CPT code for ACP was introduced, physicians could still provide this service and receive payments for it by using a more general CPT code such as that for a routine office visit. Thus, the introduction of a specific code for ACP could *only* affect record precision. In fact, reporting is the stated goal of the CPT coding system, so with the new code in place physicians could hold their provision of ACP constant and only change what is being reported. However, improving the precision of one's billing records is presumably costly to physicians as it requires them to remain up to date with the coding possibilities as well as documentation requirements for the new codes. In other words, in the absence of additional incentives, whether financial of otherwise, to use the new code, we can hardly expect physicians to start using it. In fact, although it was possible to record ACP provision with the new code without reimbursement already in 2015, almost no physicians did it.

Table 2.2.3: Reimbursement for routine office visits and ACP in 2016

CPT code	Time (min)	Revenue ($)	Change
A. Established patient office visit			
99213	20-29	73.40	
99214	30-39	108.13	34.73
99215	40-54	145.72	37.59
B. Advance care planning			
99497	16-30	85.93	
99498	16-30	74.83	

Notes. Revenue is equal to the non-facility national payment amount for each CPT code taken from the 2016 Physician Fee Schedule (CMS, 2023). The "change" column shows the difference in revenue compared to the previous row.

The dedicated reimbursement of ACP changes physicians' financial incentives, which could in turn increase the *actual* provision of the service. Specifically, as the stated purpose of introducing the new code was encouraging provision of ACP, I expect the effective net revenue of performing ACP and using the code to be higher than that of performing the same service but under a more general code. To illustrate this, in Table 2.2.3 I report national averages of physician revenue for three kinds of office visits as well as ACP. Suppose that a physician performs ACP as part of an established patient visit. If this takes her less than 16 minutes, then she cannot use the ACP code, and must bill for both ACP and the remainder of the visit using office visit codes only. However, assume that she spends 16 minutes on ACP and 24 minutes on other care. Then, she can earn $146 for the most expensive visit (40 minutes) or bill for ACP explicitly, resulting in revenue of $159 (73+86). Moreover, before the ACP reimbursement, extending the office visit by an additional 6 minutes would not change the total reimbursement, while now it would qualify the visit for a higher reimbursement, adding \$35 to the revenue. What is more, if she does this for four patients, she can earn almost twice as much as if she were to devote that time to another patient (\$140 vs \$73 for a 24 minute visit). Thus, it is possible for the dedicated ACP reimbursement to affect actual ACP provision as well as provision of other services.

Apart from reimbursing ACP, Medicare also waives patient copayment for it in certain circumstances, which could encourage the use of the new code even in the absence of direct financial incentives. In particular, holding revenue constant, physicians might prefer to be paid by Medicare than by the patient via copayment. First, Medicare is funded by the government, so is unlikely to renege on its responsibility to pay for a service. Second, copayments increase the price of a service from the patient's perspective and can thus alter her demand for future health care as well as affect her satisfaction with the physician, resulting in negative reviews of the physician in question.

The main hypothesis of this paper is that reimbursing ACP leads to increased provision of other E&M services. I expect changes in net revenue and better reimbursement to encourage provision of ACP, as physicians can still provide it under a more general code but benefit from using the new code in certain situations. However, this conjecture cannot be verified using the data available, as *actual* provision of ACP is unobservable. Instead, I quantify changes in the provision of other E&M services around the time of adoption of the ACP code to establish which of the following countervailing effects dominates. On one hand, shifting how physicians record provision of ACP from general to specific codes could lower the use of the general codes in question. On the other hand, as discussed above, the change to financial incentives might lead to increases in provision of other E&M services. In fact, many ACP claims also indicate the use of other E&M services that are optional during an annual wellness visit, such as depression screening, which raises the possibility of complementarities between ACP and these services (see Appendix Table 2.A.1 for a list of most common procedures performed on the same claim as ACP). Such complementarities, if they exist, would manifest themselves through increases in use of the services in question for patients who had ACP, which is what I show.

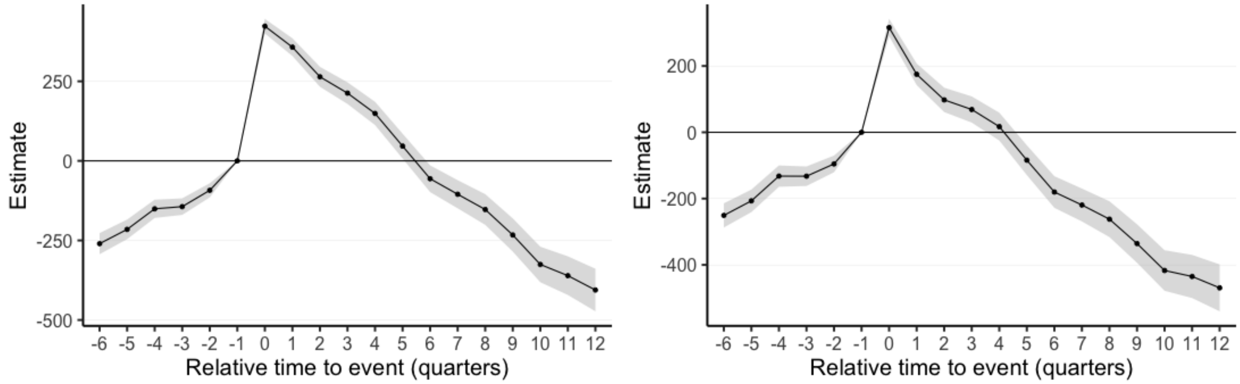
2.3 Estimates of the effect of ACP adoption

In this section I report the results of my empirical analysis. I start in section 2.3.1 by presenting estimates of regression (2.1) that use differences between physicians who did and did not adopt ACP to show persistent increases in E&M revenue following ACP adoption. In section 2.3.2 I show that about three quarters of this increase can be attributed to patients who receive ACP in the time period in question. I proceed to estimates of regression (2.2) in section 2.3.3 that show no evidence of selection of patients into "early" and "late ACP" groups. Finally, in section 2.3.4 I document that ACP adopters see more patients with conditions that are most strongly associated with ACP use, suggesting that differences in patient populations could be one of the drivers of ACP adoption.

2.3.1 Provision of services across patients

In Figure 2.3.1a I plot event-study estimates for my baseline specification that uses adoption of CPT codes for ACP as the event to trace out revenue from E&M services excluding ACP

Figure 2.3.1: Event-study estimates for E&M revenue excluding ACP



(a) Baseline specification, no weights **(b)** Weighted by inverse propensity scores

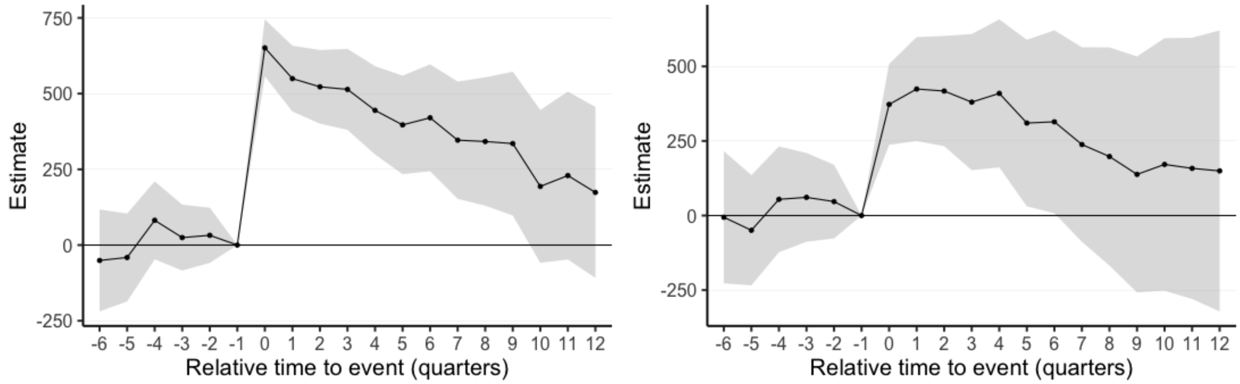
Notes. Estimates of regression equation (2.1) for the whole sample with adoption of ACP as the event. Inverse propensity score weights are estimated using a random forest as described in Appendix 2.B. Physicians with weights above 2 are excluded to ensure overlap of covariate distributions (using a less restrictive cutoff of 10 leads to similar results).

around the time of adoption. Clearly, significant and consistently negative estimates for $r < 0$ point to a violation of the parallel trends assumption that persists when I use inverse propensity score weights in order to make the adopters and non-adopters more comparable in Figure 2.3.1b. Thus, physicians who never adopted ACP are not a good comparison group for those who adopted it, even when only within-physician changes in E&M revenue are considered. In other words, despite including physician fixed effects, selection into ACP adoption remains an issue that does not allow for a causal interpretation of the estimates at hand[18]. However, there seems to be a positive relationship between provision of E&M services and ACP adoption in the short-term, but zero or negative correlation in the long term. In Appendix Figure 2.C.1 I show that removing non-adopters from the design leads

---

[18] In fact, in Appendix Figure 2.C.2 I show that the issue of significant pre-trends is equally severe among physicians who adopted ACP in 2016 versus later. Thus, using the introduction of Medicare's reimbursement policy as a source of possibly exogenous variation in ACP adoption would presumably not solve the issue of pre-trends.

Figure 2.3.2: Event-study estimates for E&M revenue excluding ACP with specialty restrictions



(a) Top 3 specialties by ACP propensity

(b) Specialists in hospice and palliative care

Notes. Estimates of regression equation (2.1) for physicians of specified specialties with adoption of ACP as the event, unweighted. Top 3 specialties by propensity are: hospice and palliative care, geriatric medicine and hospitalist

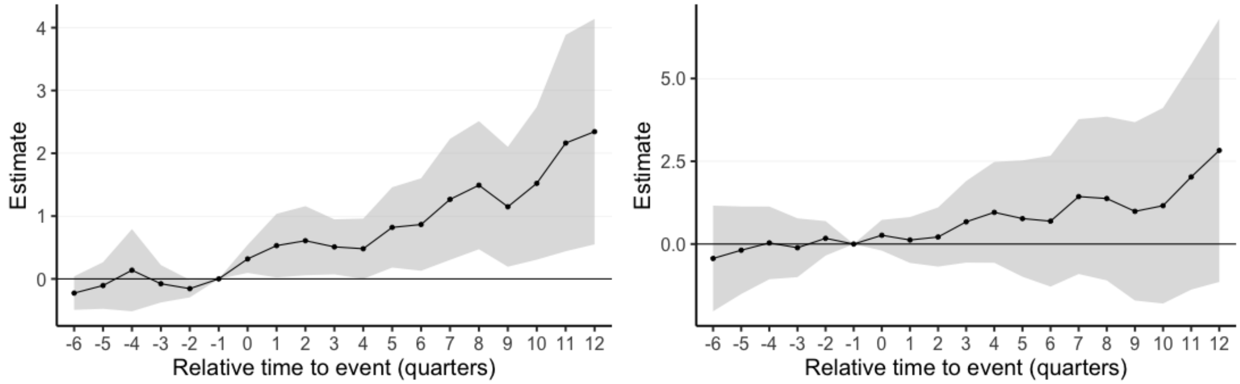
to a similar conclusion and does not resolve the issue of pre-trends either.

As ACP adoption is rare, it is not surprising that adopters seem very different from non-adopters in other ways than in terms of their revenue trajectory around the time of adoption. In order to arrive at a sample of physicians for whom the comparison by adoption status allows for valid inference, I estimate the propensity of ACP adoption by physician specialty (see Appendix Table 2.A.2). The three specialties with highest adoption propensity are: hospice and palliative care, geriatric medicine and hospitalist with adoption rates of 46%, 29% and 19%, respectively. In Figure 2.3.2 I show unweighted estimates of the baseline regression restricted to physicians of these three specialties (a) or to hospice and palliative care specialists only (b). Both plots point to an increase in quarterly E&M revenue that decreases over time but remains positive, although not statistically significant, three years after adoption. As I find no evidence of a violation of the parallel trends assumption for these two groups of physicians, I focus on the three aforementioned specialties for the remainder of my analysis.

The magnitude of the increase at $r = 0$ equals \$651 for the three specialties and decreases to \$514 when the control group is excluded, as reported in Appendix Figure 2.C.3 (for hospice and palliative care specialists these estimates are \$372 and \$309, respectively). Importantly, when I use total E&M revenue as the outcome variable (i.e., including ACP), these estimates increase to \$802 and \$668 (or \$544 and \$480 for hospice and palliative care specialists), respectively, which implies revenue of \$151-172 from the ACP procedure, and corresponds to the provision of around 2 of such procedures per quarter. Thus, the payment for ACP itself amounts to only 19-36% of the total increase in revenue caused by adopting the new CPT code, depending on specification. Unless this is offset by decreases in provision of medical services in the long-term, for which I find no evidence, failing to account for this spillover effect of one procedure on the use of others leads to a flawed projection of a reimbursement policy's future cost.

I attempt to identify specific procedures that are driving the aggregate changes in provision shown above. Using revenue for each of the procedures that are most commonly billed with ACP identified in Appendix Table 2.A.1 I find no evidence for changes in use of some procedures (e.g., routine electrocardiogram, routine venipunctures) and routine office visits, with the caveat that in the case of annual wellness visits the parallel trends assumption seems violated (see Appendix Figure 2.C.4). In Figure 2.3.3 I show estimates for two screening procedures, which point to possible increases in their use as a result of ACP adoption. Crucially, both of these procedures last 10-15 minutes and are covered by Medicare once

Figure 2.3.3: Event-study estimates for revenue from annual screening procedures



(a) Annual alcohol misuse screening

(b) Annual depression screening

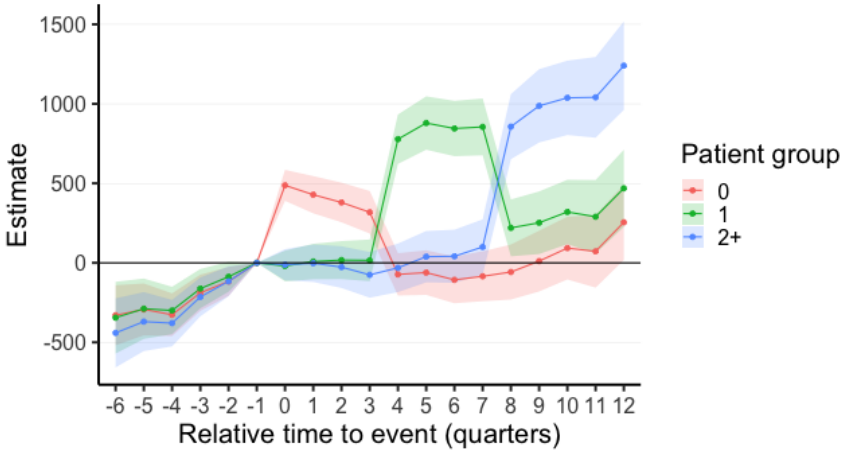
Notes. Estimates of regression equation (2.1) for physicians of top 3 specialties by ACP propensity with revenue from HCPCS code G0442 in panel (a) and G0444 in panel (b). See Appendix Figure 2.C.4 for routine office visits.

per year with an allowed amount of around \$18 (CMS, 2023). Therefore, it is possible that spending more time with the patient while performing ACP familiarizes the physician with a patient enough to make her efficient at providing such screening services, which otherwise might be too cheap to be worth a physician's time. However, it is worth bearing in mind that the magnitude of these estimates, which are in dollars, is very small and corresponds to a fraction of either screening procedure per quarter.

2.3.2 Heterogeneity by timing of ACP

In order to shed light on whether estimated increases in E&M provision are driven by patients who receive ACP, I split each physician's patients into groups based on the timing of their first ACP visit with respect to physician's ACP adoption. In this way, each patient of a physician belongs to one of the following groups: None, 0, 1, 2+, where "None" corresponds

Figure 2.3.4: Event-study estimates for E&M revenue excluding ACP by patient group



Notes. Estimates of modified regression equation (2.2) for physicians of top 3 specialties by ACP propensity. Patients are grouped by time between their ACP and the time of physician's ACP adoption in years (patients who never had ACP are the reference group).

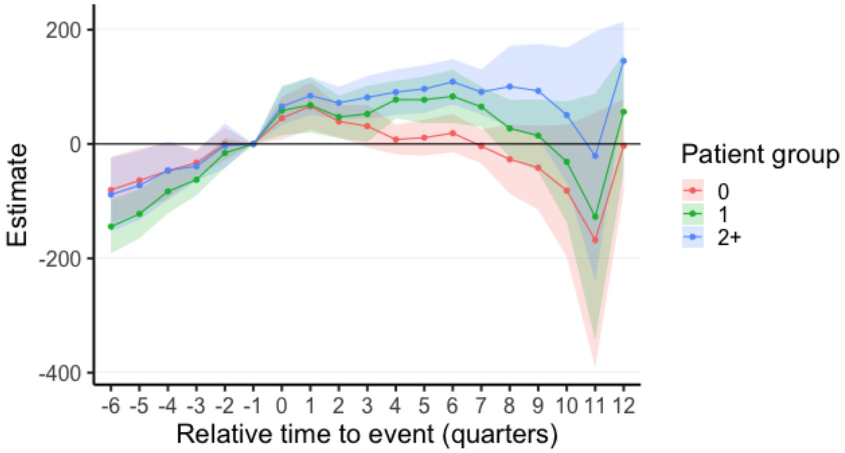
to patients who never had ACP[19], while 0 (1 or 2+) denotes those who had it within a year (1-2 years after or more than 2 years after, respectively) of physician's ACP adoption etc. As I will be comparing changes in physicians' revenue across these groups, possible differences in their size are important for interpretation. In Appendix Figure 2.C.6 I show that once physicians adopt ACP, their subsequent provision is quite stable over time at around 2 patients per quarter, with the exception of 2016 adopters for whom the number is closer to 2.5.

In Figure 2.3.4 I plot estimates of the effect of adoption on E&M revenue excluding ACP by patient group with "None" as the reference group. Although ACP patients are significantly different than non-ACP patients in terms of E&M revenue before ACP adoption, differences between them are not significant outside the times when they have ACP. For

---

[19] Note that all patients of physicians who never used ACP belong to this group, as group assignment is based on ACP occurring between a patient and the specific physician in question.

Figure 2.3.5: Event-study estimates for non-E&M revenue by patient group



Notes. Estimates of modified regression equation (2.2) for physicians of top 3 specialties by ACP propensity. Patients are grouped by time between their ACP and the time of physician's ACP adoption in years (patients who never had ACP are the reference group).

example, patients from group 0 are significantly different from patients from groups 1 and 2+ only for $r = 0, 1, 2, 3$ which is when they receive ACP services while the other groups do not. Moreover, for r between 0 and 3 estimates for patients in group 0 decrease from \$487 to \$318 which corresponds to 75% and 71% of the total estimated increase in revenue across all patients, respectively (estimated as \$651 and \$445 in Figure 2.3.2a). The remaining share of the estimated revenue increase can be attributed roughly equally to other groups as estimates for 1 and 2+ are not statistically different from those for the reference group "None" at $r = 0, 1, 2, 3$.

In Figure 2.3.5 I plot estimates for revenue for non-E&M services by patient group. Note that groups 1 and 2+ do not exhibit any significant differences between each other throughout the time horizon in question, whereas group 0 brings less non-E&M revenue than either of the other groups for $r = 4, 5, 6, 7$ with inconsistent significance. Moreover, similar though insignificant divergence emerges at $r = 8, 9, 10, 11$ when group 1 starts bringing less revenue than group 2+. In sum, after receiving ACP patients' utilization of non-E&M services drops to the level for patients who never had ACP. Crucially, this stands in stark contrast to the equivalent estimates for E&M revenue in terms of both decreases after ACP and no increases contemporaneous with ACP.

Taken together, my estimates by patient group show that providing ACP to a patients is associated with more provision of other E&M services at the same time, but no change in non-E&M services. I find no evidence that these changes reflect mere changes in the time of service provision, as the aforementioned increases in E&M utilization are not followed by corresponding decreases, which would be the case if optional screening procedures were simply performed earlier due to convenience of combining them with ACP. Moreover, among patients who receive ACP, increases in E&M revenue are highest for those who receive it further away from a physician's adoption, suggesting that physicians improve in their ability to maximize revenue while providing ACP services, especially that the number of ACP patients decreases as we go further away from a physician's adoption date[20] (see Appendix Figure 2.C.6).

When it comes to underlying mechanisms that could explain the difference in estimates for E&M and non-E&M services, it is possible that complementarities discussed in section 2.2.6 only arise for E&M. Unlike other outpatient services, E&M mostly entails various kinds of counseling services that can often be billed based on time spent with the patient and over which physicians have substantial discretion[21]. Thus, adjustments of visit lengths

---

[20] Of course, among physicians who adopted ACP in 2016 the group 2+ contains more patients, as it includes patients from years 3 and 4 after the physician's adoption date. In Appendix Figure 2.C.5 I plot estimates for revenue per beneficiary that account for this and lead to a similar conclusion.

[21] In fact, Fang and Gong (2017) use time requirements for CPT codes along with other estimates of time

to maximize revenue as described in section 2.2.6 might only be possible with E&M services, while the physician might become more efficient at providing care of the counseling type as the total visit time increases. With regards to slight decreases in the use of non-E&M services following ACP, it could be caused by a variety of factors, such as: patient's reluctance for elective care due to budget constraints or direct effect of ACP via recognition of one's priorities for future medical care.

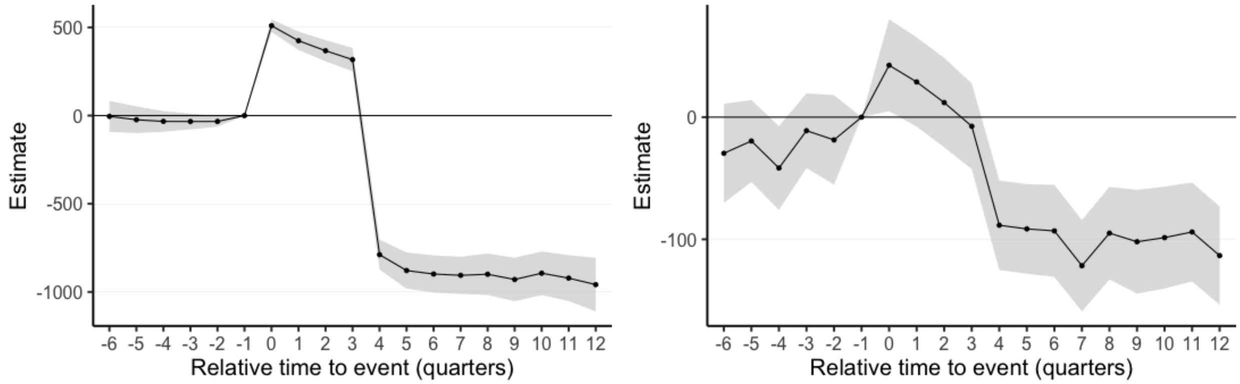
2.3.3 "Early" vs "late ACP" patients

In order to capture the effect of receiving ACP services within one year of a physician's adoption of the service I plot estimated parameters of regression (2.2) in Figure 2.3.6. In panel (a) we see no evidence of a violation of the parallel trends assumption, as patients who received ACP earlier are not significantly different from those who received it later in terms of the trend of revenue that they generate for the physician. "Early ACP" patients are responsible for up to \$500 more E&M revenue per quarter at $r = 0, 1, 2, 3$, but this increase is outweighed by a much larger decrease in revenue at $r > 3$. In panel (b) I plot estimates for revenue per patient to account for the fact that the two groups of patients are of different size. "Early ACP" patients bring on average \$100 less revenue per quarter at $r > 3$, which accounts for approximately \$200 of the difference in panel (a), as physicians provide ACP to 2 patients per quarter on average. In Appendix Figure 2.C.7 I provide analogous estimates for additional outcome variables to show that "early ACP" patients bring around \$15 more revenue per procedure at $r = 0, 1, 2, 3$ but \$23 less at $r > 3$ than

---

spent with patients to show that aggregating the implied time that physicians spend with their patients often leads to implausibly high totals (e.g., more than 100 hours per week only on care for Medicare patients).

Figure 2.3.6: Event-study estimates of differences between "early" and "late ACP" patients



(a) E&M revenue excluding ACP **(b)** Per patient E&M revenue excluding ACP

Notes. Estimates of regression equation (2.2) for physicians of top 3 specialties by ACP propensity.

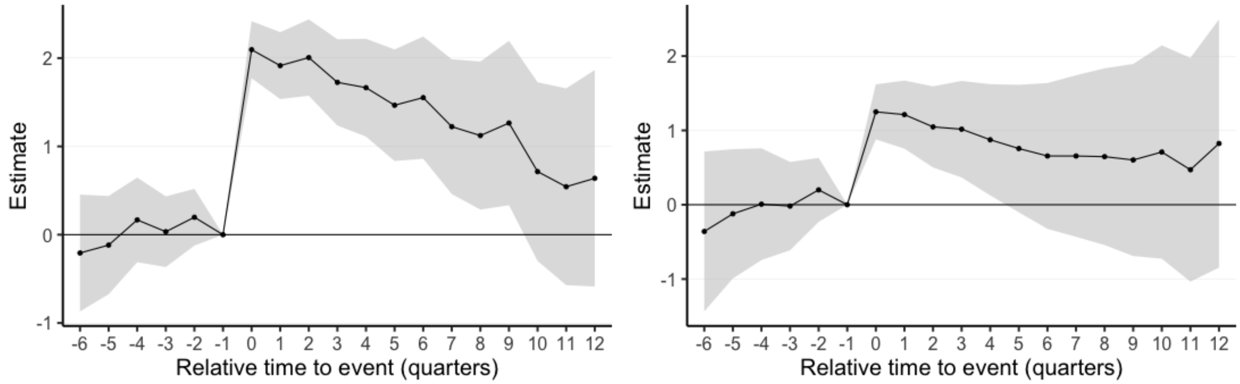
"late ACP" patients, suggesting that physicians transition towards using more expensive procedures after adopting ACP.

The fact that "early ACP" patients account for less revenue increases than "late ACP" patients could be the result of either a causal effect of the ACP procedure on the patient's behavior or a change in physicians' behavior as they converge towards their long-term strategy of ACP use. As patients who received ACP at different times exhibit parallel trends in health care utilization before ACP, it is unlikely that the estimated effects are driven by differences in patient characteristics or health status. In other words, I find no evidence of patient selection into time of ACP provision, conditional on receiving ACP at some point in time.

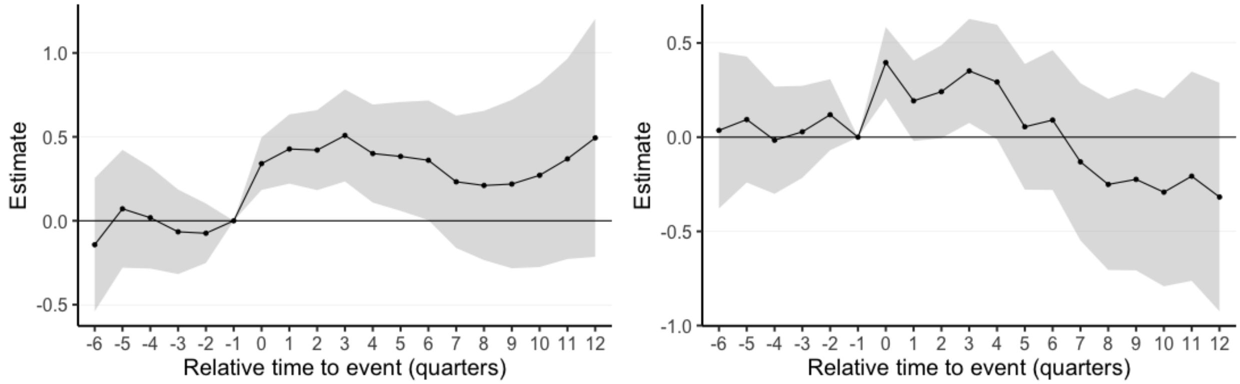
2.3.4 Characteristics of the patient pool

Increases in E&M revenue following ACP adoption could to some extent reflect changes in the kind of patients that physicians see. In Figure 2.3.7a I show the estimated effect of

Figure 2.3.7: Event-study estimates for the number of beneficiaries with a specific diagnosis



(a) Total number of patients **(b)** Essential (primary) hypertension



(c) Hyperlipidemia, unspecified **(d)** Type 2 diabetes mellitus w/o complications

Notes. Estimates of regression equation (2.1) for physicians of top 3 specialties by ACP propensity with the outcome variable corresponding to the total number of patients (a), and the number of patients who had a specific ICD-10 diagnosis code: I10 (b), E785 (c), E119 (d).

ACP adoption on the number of patients of a physician, which points to an increase of 2 patients per quarter at the time of adoption that attenuates over time to around 0.5 patient three years later. As the estimated increase in non-ACP E&M revenue at $r = 0$ for this specification equals \$651 with \$487 attributed to patients who receive ACP in the first year since adoption, the remaining \$164 could be driven by new patients that the physician is able to accommodate by reallocating her time as a response to the new financial incentives[22].

---

[22] Note that from panel B of Table 2.2.2 we know that this population of ACP adopters receives around \$198 of revenue per patient in the reference period.

The remaining panels of Figure 2.3.7 show estimates for the number of patients who had a specific diagnosis for three conditions that most commonly appear on ACP claims (Appendix Table 2.A.1). Importantly, these estimates do not necessarily reflect new patients nor new diagnoses of the respective conditions. For example, in the quarter of their first ACP claim, ACP adopters see on average one more patient with primary hypertension than non-adopters. It could be the case that such patients have high ACP needs and therefore the quarter in which the arrive for a visit is more likely to be the quarter in which the physician uses ACP. On the other hand, changes in the number of patients with a certain diagnosis code could reflect changes to diagnostic practices of the physician in question. Although ACP reimbursement is not conditional on specific diagnoses, it is possible that services that physicians bill for in conjunction with ACP to increase revenue do have such requirements[23]. Thus, changes in diagnosis prevalence following ACP adoption could point to the causal effect of ACP as well as to a correlation between a physician's patient pool characteristics and their propensity to provide the service.

2.4 Conclusion

This paper uses the availability of a new procedure code for advance care planning to showcase the way in which changes in reimbursement of one service can affect the provision of other services. Although the service in question costs \$80 and leads to an increase in health care utilization of only \$151 per quarter per physician, using it is associated with \$651 more

---

[23] For example, routine electrocardiograms which are often provided with ACP, could be deemed medically necessary for hypertension patients, which is a basis for Medicare reimbursement.

spending per physician on other Evaluation and Management services, most of them provided to recipients of ACP. This result is specific to physicians who specialize in hospice and palliative care, geriatric medicine and hospitalists, but I also estimate increases of similar magnitude for the entire population of physicians, although without causal interpretation due to selection concerns.

I explain the results at hand by possible complementarities between ACP and other E&M services. In particular, increasing the length of a patient visit with the addition of ACP might make it easier for physicians to also provide other counseling services, as they become more familiar with the patients and her needs. I find that my estimates are driven by recipients of ACP and show slight increases in the use of optional screening procedures that give credence to the aforementioned hypothesis. However, the latter estimates are minuscule in magnitude, whereas the treatment effects concerning aggregate use of E&M services fade over time, which is not in line with the hypothesis in question.

The main caveat of my analysis is the fact that I leverage adoption of ACP as a source of variation for causal estimates without understanding the reasons why some physicians adopt the code while others do not. My attempt at addressing this issue with a machine learning propensity score model fails at making the adopters and non-adopters comparable in terms of pre-ACP revenue. Thus, a crucial area of future research concerns investigating adoption patterns further to possibly validate my estimates using another source of variation in adoption of the code. Moreover, the event-study framework that I use could be applied in the context of other recent introductions of E&M codes in order to verify whether similar results can be found.

Appendix

2.A Additional summary statistics

Table 2.A.1: Most common diagnosis and procedure codes on ACP claims

Code	Description	Count
A. Principal diagnosis codes		
Z0000	Encounter for general adult medical examination without abnormal findings	291078
I10	Essential (primary) hypertension	89899
Z0001	Encounter for general adult medical examination with abnormal findings	69842
Z7189	Other specified counseling	31607
E119	Type 2 diabetes mellitus without complications	20308
Z515	Encounter for palliative care	19639
E785	Hyperlipidemia, unspecified	12561
J449	Chronic obstructive pulmonary disease, unspecified	12081
I2510	Atherosclerotic heart disease of native coronary artery without angina pectoris	10158
A419	Sepsis, unspecified organism	9031
B. Outpatient procedure codes		
G0439	Annual wellness visit, includes a personalized prevention plan of service (PPPS), subsequent visit	401396
99214	Established patient office visit, 30 to 39 minutes	207047
G0444	Annual depression screening, 5 to 15 minutes	193872
G0442	Annual alcohol misuse screening, 5 to 15 minutes	114924
99213	Established patient office or other outpatient visit, 20-29 minutes	112412
99233	Subsequent hospital visit, at least 50 minutes	101472
G8427	Eligible clinician attests to documenting in the medical record they obtained, updated, or reviewed the patient's current medications	98267
G0438	Annual wellness visit, includes a personalized prevention plan of service (PPPS), first visit	92909
93000	Routine electrocardiogram (ECG) with at least 12 leads, including the tracing, interpretation, and report	69726
36415	Routine venipunctures, not requiring the skill of a physician, for specimen collection	67852

Notes. The 20% sample of Carrier claims for years 2010-2019 contains 1.1 million claims that include ACP services. Panel A shows ten most common principle diagnosis codes appearing on ACP claims, whereas Panel B shows 10 most common outpatient procedures appearing on the same claim as ACP. Codes are ICD-10 for diagnoses and HCPCS for procedures.

Table 2.A.2: ACP adoption propensity by provider specialty

Specialty	Count	Estimate	SE	
Hospice and Palliative Care	2154	0.4646	0.0046	***
Geriatric medicine	3932	0.2887	0.0031	***
Hospitalist	17927	0.1901	0.0019	***
Internal medicine	167808	0.1316	0.0004	***
Family practice	128027	0.1042	0.0005	***
Nurse practitioner	206503	0.0612	0.0004	***
Hematology/oncology	13768	0.0608	0.0015	***
Medical oncology	6129	0.0561	0.0026	***
Pulmonary disease	14977	0.0506	0.0015	***
Nephrology	12028	0.0482	0.0016	***
Osteopathic manipulative therapy	1765	0.0446	0.0050	***
Advanced heart failure and transplant cardiology	654	0.0363	0.0119	***
General practice	19000	0.0362	0.0015	***
Certified clinical nurse specialist	4306	0.0362	0.0026	***
Hematology	1785	0.0342	0.0050	***
Gynecologist/oncologist	1599	0.0317	0.0046	***
Critical care (intensivists)	7141	0.0315	0.0025	***
Geriatric psychiatry colorectal surgery	404	0.0225	0.0110	**
Physician assistant	116098	0.0210	0.0005	***
Hematopoietic cell transplantation and cellular therapy	215	0.0208	0.0226	
Neurology	20396	0.0149	0.0012	***
Infectious disease	8559	0.0139	0.0019	***
Cardiology	33550	0.0134	0.0010	***
Emergency medicine	65312	0.0111	0.0007	***

Notes. Estimates from a regression of a physician's adopter dummy on her specialty. Only specialties with estimated probability of adoption of at least 1% are included. As physicians have multiple specialties, each I weigh each specialty by the fraction of claims of a physician that appeared with that specialty.

Table 2.A.3: Differences between adopters and non-adopters: additional variables

	Control (1)	Adopter (2)	Diff. (3)
A. All physicians			
Female patients (%)	59.24	59.73	0.49 ***
Black patients (%)	10.67	11.24	0.57 ***
Hispanic patients (%)	2.29	2.21	−0.08 *
Patients aged 65-75 (%)	41.23	37.04	−4.19 ***
Patients aged 75-85 (%)	24.59	27.69	3.10 ***
Patients aged over 85 (%)	11.34	18.98	7.64 ***
Revenue from annual alcohol misuse screening ($)	0.11	3.18	3.06 ***
Revenue from annual depression screening ($)	0.30	6.58	6.28 ***
Revenue from established patient office visit 30-39 min ($)	411.70	1467.76	1056.05 ***
Revenue from subsequent annual wellness visits ($)	19.75	211.90	192.15 ***
B. Top 3 specialties among ACP adopters			
Female patients (%)	59.81	61.06	1.25
Black patients (%)	15.00	13.01	−1.98
Hispanic patients (%)	2.41	1.61	−0.80
Patients aged 65-75 (%)	30.07	24.70	−5.37 ***
Patients aged 75-85 (%)	29.27	29.78	0.51
Patients aged over 85 (%)	29.11	34.91	5.81 ***
Revenue from annual alcohol misuse screening ($)	0.00	0.21	0.21 *
Revenue from annual depression screening ($)	0.22	1.69	1.47 ***
Revenue from established patient office visit 30-39 min ($)	545.04	777.33	232.29 **
Revenue from subsequent annual wellness visits ($)	46.11	81.82	35.71
C. Only hospice and palliative care specialists			
Female patients (%)	52.05	53.96	1.91
Black patients (%)	12.18	14.45	2.27
Hispanic patients (%)	1.67	1.45	−0.21
Patients aged 65-75 (%)	33.28	29.74	−3.54
Patients aged 75-85 (%)	27.81	27.50	−0.31
Patients aged over 85 (%)	23.84	25.86	2.02
Revenue from annual alcohol misuse screening ($)	0.00	0.00	0.00
Revenue from annual depression screening ($)	0.00	0.09	0.09
Revenue from established patient office visit 30-39 min ($)	152.08	93.81	−58.27
Revenue from subsequent annual wellness visits ($)	60.91	2.57	−58.34

Notes. This is an extension to Table 2.2.2. Based on values for last quarter of 2015, before ACP became reimbursed by Medicare. Top 3 specialties among ACP adopters refers to specialties with the highest propensity of adoption, which are: hospice and palliative care, geriatric medicine and hospitalist (see Appendix Table 2.A.2). The four sources of revenue are calculated based on the use of HCPCS codes G0442, G0444, 99214 and G0439, which are the four most common procedures that are billed together with ACP.

2.B Inverse propensity score weighting

In order to account for unobserved differences between physicians who did and did not adopt the new ACP codes, I use a machine learning algorithm to estimate the predicted probability of adoption for each physician (i.e., propensity score $p(x)$). Then, following the approach of Abadie and Cattaneo (2018), physicians are weighted by the inverse of the predicted probability of belonging to its observed group (i.e., adopters' weight is $\frac{1}{p(x)}$, while non-adopters' weight is $\frac{1}{1-p(x)}$). In this section I provide details on the way that I calculate propensity scores.

Propensity scores are calculated as fitted values from a predictive model that uses an indicator for physician's adopter status as the outcome variable, and various characteristics of the physician from year 2015 as predictors. I use information from year 2015, before Medicare introduced its ACP reimbursement policy, to ensure the policy in question had no effect on the inputs to the predictive model. In Table 2.B.1 I list all predictors used – there are four instances of each of them, one for each quarter of 2015, as the predictive model is static, so values from various quarters are treated as separate inputs. For identifying procedures and diagnoses that could be predictive of ACP adoption I look at ACP claims.

I consider two predictive models: LASSO (least absolute shrinkage and selection operator) and random forest. LASSO is a linear regression model with a penalty on the sum of absolute values of its coefficients, which results in many coefficients being shrunk to zero. Random forest is a tree-based machine learning algorithm that allows for interaction effects between predictors in a flexible way. To avoid overfitting, I use a 75% random sample of physicians

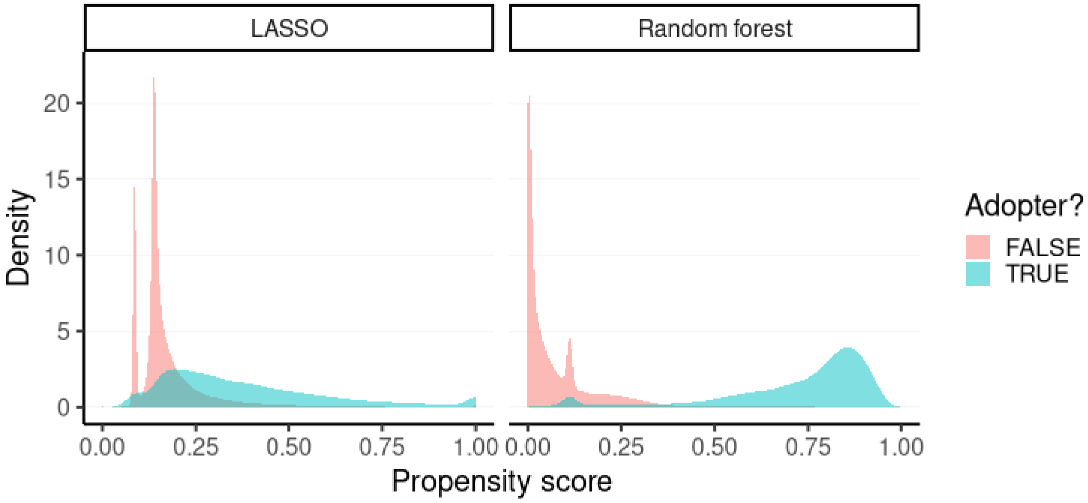
Table 2.B.1: Predictors for the propensity score model

Overall utilization measures
Number of services provided, Number of E&M services provided, Number of claims, Number of claim lines, Total revenue, Total revenue from E&M services, Amount paid by beneficiaries
Revenue by type of claim
Local carrier non-DMEPOS claims, Local carrier DMEPOS claims, DMERC and non-DMEPOS claim, DMERC and DMEPOS claim, Claims in which the provider participated, Claims for which all or some covered and allowed expenses are applied to deductible, Claims with 65% fee schedule (physician assistants assisting in surgery or nurse midwives), Claims with 75% fee schedule (physician assistants performing services in a hospital other than assisting surgery or nurse practitioners and clinical nurse specialists performing services in rural areas or clinical social worker services), Claims with 85% fee schedule (physician assistant services for other than assisting surgery or nurse practitioners services)
Patient pool characteristics
Number of patients, Female patients, Male patients, Average patient age, Patients under 55, Patients 55-65, Patients 65-75, Patients 75-85, Patients 85 or more, White patients, Black patients, Asian patients, Hispanic patients, Native American patients, Patients of other race
Utilization of specific procedures (measured by both service count and revenue)
1036F, 1090F, 1100F, 1101F, 1123F, 1124F, 1125F, 1126F, 1170F, 3017F, 3044F, 3288F, 36415, 3725F, 4040F, 80061, 81002, 81003, 83036, 85025, 90662, 90670, 90732, 93000, 96372, 99204, 99213, 99214, 99215, 99222, 99223, 99232, 99233, 99239, 99306, 99308, 99309, 99310, 99406, G0008, G0009, G0402, G0438, G0439, G0442, G0444, G0446, G0447, G8417, G8420, G8427, G8482, G8510, G8752, G8754, G8783
Number of patients with specific diagnosis
A419, C61, D649, E039, E1122, E1165, E119, E538, E559, E6601, E669, E7800, E782, E784, E785, F0390, F17210, F329, F411, F419, G20, G309, G4700, G4733, G8929, G894, I10, I119, I129, I2510, I480, I482, I4891, I509, I639, I739, J189, J309, J441, J449, J9601, K219, K5900, M150, M1990, M545, M6281, M810, M8580, N179, N183, N186, N189, N390, N400, N401, R05, R0602, R1310, R42, R531, R5381, R5383, R7301, Z0000, Z0001, Z1211, Z1231, Z125, Z1331, Z136, Z1389, Z23, Z515, Z66, Z6825, Z6826, Z6827, Z6830, Z713, Z7189, Z789, Z79899, Z87891, Z9181

Notes. DMEPOS stands for durable medical equipment, prosthetics, orthotics, and supplies. DMERC stands for durable medical equipment regional carrier.

to train the random forest model. In Figure 2.B.1 I plot the estimated propensity scores by model and true adopter status. Unsurprisingly given its flexibility, the random forest model performs much better, as the distribution of propensity scores is concentrated at higher values among adopters, while remaining low for non-adopters. Both models result in

Figure 2.B.1: Distribution of estimated propensity scores by model and adopter status



Notes. As LASSO is a linear model, fitted values outside of the $[0, 1]$ interval occur, which I censor at 0 and 1 to fall within the interval.

a bi-modal distribution of propensity scores among non-adopters, and both result in very few non-adopters having propensity scores higher than 0.3. Thus, it seems that pre-policy characteristics have a lot of potential when it comes to predicting a physician's future adopter status.

I present importance measures for top 10 predictors for each model in Table 2.B.2. For LASSO, I measure importance of a predictor with the absolute value of its coefficient, whereas for random forest I use the default importance measure based on the variance of the outcome variable. Only 72 out of 912 predictors receive a non-zero LASSO estimate (the remaining ones are all ranked 73), while 568 predictors have non-zero random forest importance. Clearly, LASSO identifies the use of specific procedures as most predictive of a physician's adoption status, while random forest points to overall measures of utilization and the characteristics of one's patient pool. Crucially, LASSO is a linear model, while random forests

Table 2.B.2: Variable importance measures for propensity score models

		LASSO		Random forest	
Variable	Quarter	Coef.	Rank	Coef.	Rank
Patients with I10 diagnosis	4	0.0000	73	523	1
Patients with E039 diagnosis	4	0.0000	73	419	2
E&M revenue	4	0.0000	51	366	3
E&M number of services	4	0.0000	73	310	4
Patients with E119 diagnosis	4	0.0000	73	284	5
E&M number of services	3	0.0000	73	258	6
Patients with E785 diagnosis	4	0.0000	73	247	7
E&M number of services	1	0.0000	73	245	8
Number of G0439 services	4	0.0091	7	237	9
E&M revenue	3	0.0000	69	229	10
Patients with Z515 diagnosis	4	0.0245	2	131	29
Number of 99239 services	4	0.0086	8	94	53
Number of G0438 services	4	0.0063	9	47	165
Number of 90732 services	4	0.0094	6	10	383
Number of G0402 services	2	0.0161	3	7	413
Number of 90732 services	1	0.0062	10	7	414
Number of G0402 services	4	0.0310	1	7	415
Number of G0402 services	3	0.0109	5	7	422
Number of 90732 services	2	0.0130	4	5	452

Notes. The table reports the variable importance coefficient and rank for 10 most important variables for each model. To measure variable importance I use regression coefficients in the case of LASSO, and node impurity in the case of random forest. Diagnosis codes denote: I10 hypertension, E039 hypothyroidism, E119 type 2 diabetes, E785 hyperlipidemia. Procedure codes G0438 and G0439 are Annual Wellness Visits (first and subsequent, respectively), 90732 denote pneumococcal vaccinations and flu shots, G0402 is an Initial Preventive Physical Examination, 99239 is a hospital inpatient or observation discharge day management service.

exploits interaction effects between predictors. Both models agree on the predictive value of the number of patients with diagnosis code Z515, which denotes an encounter for palliative care, suggesting that physicians who provide such care in the first place, whether due to their practice style or patient characteristics, are more likely to use ACP, which is also meant to improve patient care at the end of life. Moreover, the number of Annual Wellness Visits provided (G0439) ranks among the top 10 predictors for both models, which is not surprising given that such a visit is the primary way through which Medicare reimburses ACP provision.

Table 2.B.3: Differences between adopters and non-adopters by inverse propensity score weight used

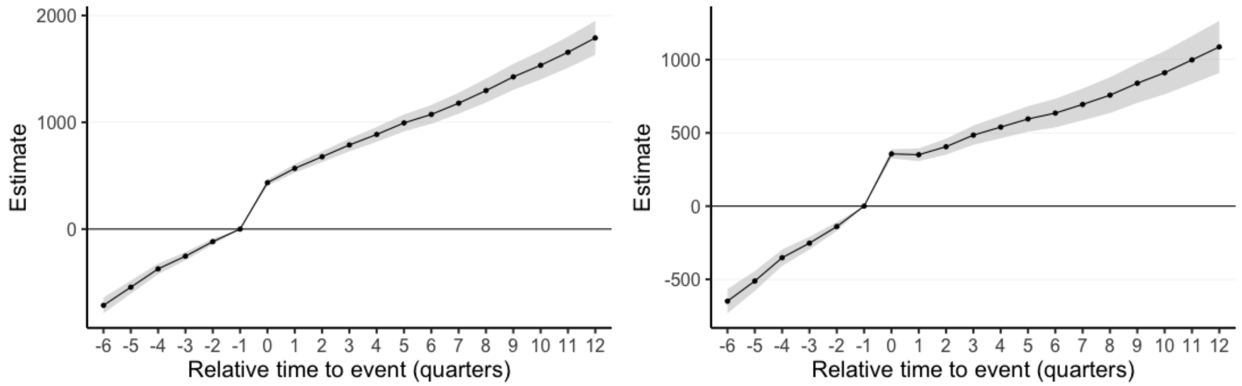
	Control	Adopter	Diff.	
A. Unweighted sample				
E&M revenue	3687.62	4883.02	1195.40	***
Non-E&M revenue	501.63	628.11	126.48	
Patient average age	68.84	74.83	5.99	***
Patients	18.68	24.62	5.94	***
Patients with COPD	1.80	2.26	0.47	
Patients with hyperlipidemia	2.91	3.37	0.46	
Patients with hypertension	12.73	14.76	2.02	
B. Random forest weights truncated at 2				
E&M revenue	3862.83	5029.99	1167.16	***
Non-E&M revenue	561.15	654.10	92.96	
Patient average age	69.67	77.24	7.57	***
Patients	19.70	25.42	5.72	***
Patients with COPD	1.89	2.32	0.42	
Patients with hyperlipidemia	3.20	3.49	0.30	
Patients with hypertension	13.85	15.18	1.33	
C. Random forest weights truncated at 10				
E&M revenue	4115.50	3880.93	-234.57	
Non-E&M revenue	525.93	487.20	-38.73	
Patient average age	71.25	64.76	-6.49	***
Patients	20.75	19.67	-1.07	
Patients with COPD	2.04	1.76	-0.28	
Patients with hyperlipidemia	3.08	2.60	-0.48	
Patients with hypertension	14.35	11.36	-2.99	*

Notes. Based on values for last quarter of 2015 for the main analysis sample including top 3 specialties among ACP adopters.

In Table 2.B.3 I show how differences in covariates between adopters and non-adopters change upon weighting using the random forest model. I focus on physicians from the 3 specialties used in my analysis (hospice and palliative care, geriatric medicine and hospitalist). Without weights, adopters receive more E&M revenue and have more patients who are substantially older on average (panel A). In panels B and C I use inverse propensity score weights, but exclude physicians whose weight is above 2 or 10, respectively. Large weights are a result of propensity scores close to 1 or 0, so excluding physicians with such weights is meant to exclude extreme cases to ensure overlap of characteristics distributions between the adopter and non-adopter groups. In panel B we see that using severely restricted weights brings the two groups somewhat closer to each other in terms of some covariates but not others. In panel C, with only the most extreme cases excluded, the average differences between the two groups switch signs and often become insignificant. Creating a similar table for all physicians leads to significant differences in all characteristics regardless of the kind of weight used. In sum, even on the restricted sample weighting does not ensure balance on covariates, so I use the unweighted sample in my main analysis and validate my results by restricting attention to palliative care physicians for whom balance on covariates seems satisfied. Moreover, my regression results do not change markedly when I use weights, so using the unweighted version as primary is of little consequence.

2.C Additional event-study estimates

Figure 2.C.1: Event-study estimates for E&M revenue excluding ACP without control group

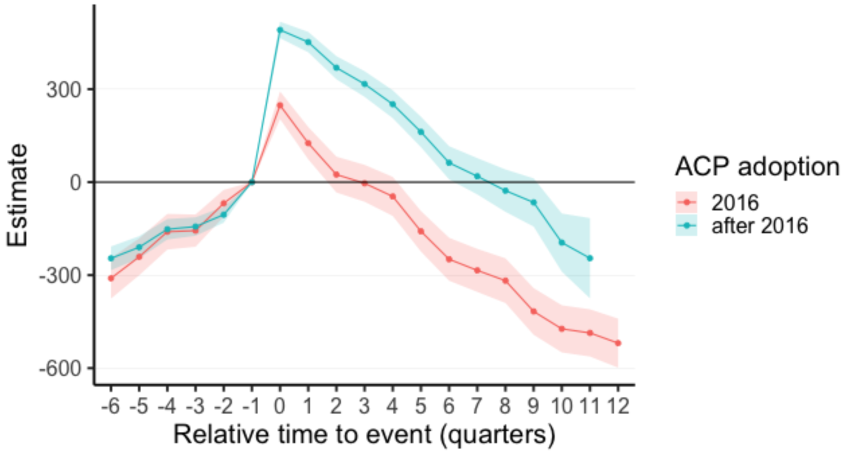


(a) Baseline specification, no weights

(b) Weighted by inverse propensity scores

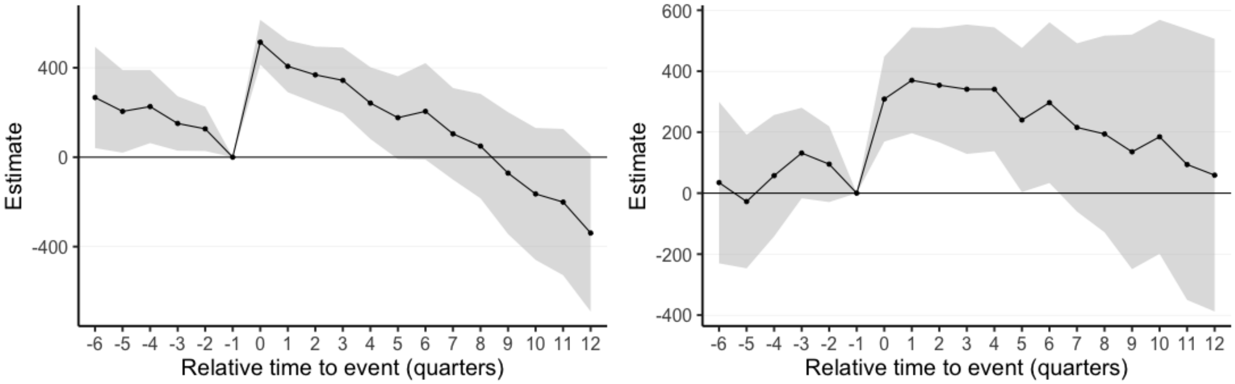
Notes. Estimates of regression equation (2.1) for adopters only with adoption of ACP as the event. Instead of physician fixed effects the specification includes 5-digit zip code fixed effects. Inverse propensity score weights are estimated using a random forest as described in Appendix 2.B. Physicians with weights above 2 are excluded to ensure overlap of covariate distributions (using a less restrictive cutoff of 10 leads to similar results).

Figure 2.C.2: Event-study estimates for E&M revenue excluding ACP by adoption year



Notes. Estimates of regression equation (2.1) separately for physicians who adopted ACP in 2016 and for those who did so later. In both specifications the control group remains the same.

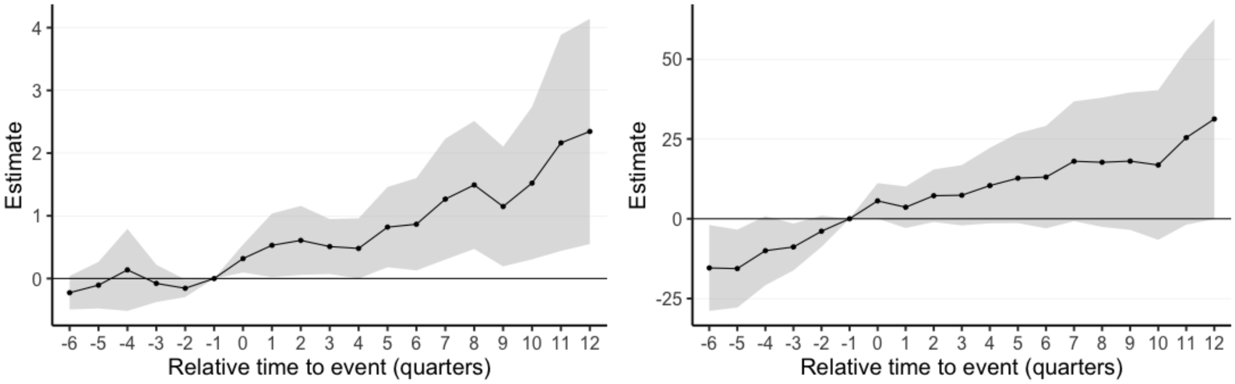
Figure 2.C.3: Event-study estimates for E&M revenue excluding ACP with specialty restrictions without control group



(a) Top 3 specialties by ACP propensity **(b)** Specialists in hospice and palliative care

Notes. Estimates of regression equation (2.1) only for adopters of specified specialties with adoption of ACP as the event. Instead of physician fixed effects the specification includes 5-digit zip code fixed effects. Top 3 specialties by propensity are: hospice and palliative care, geriatric medicine and hospitalist

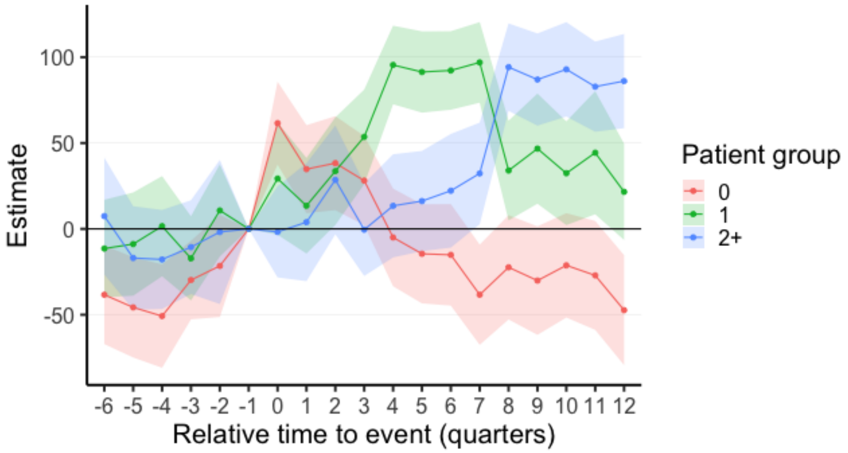
Figure 2.C.4: Event-study estimates for revenue from routine office visits



(a) Established patient office visit, 30-39 min **(b)** Annual wellness visit (subsequent)

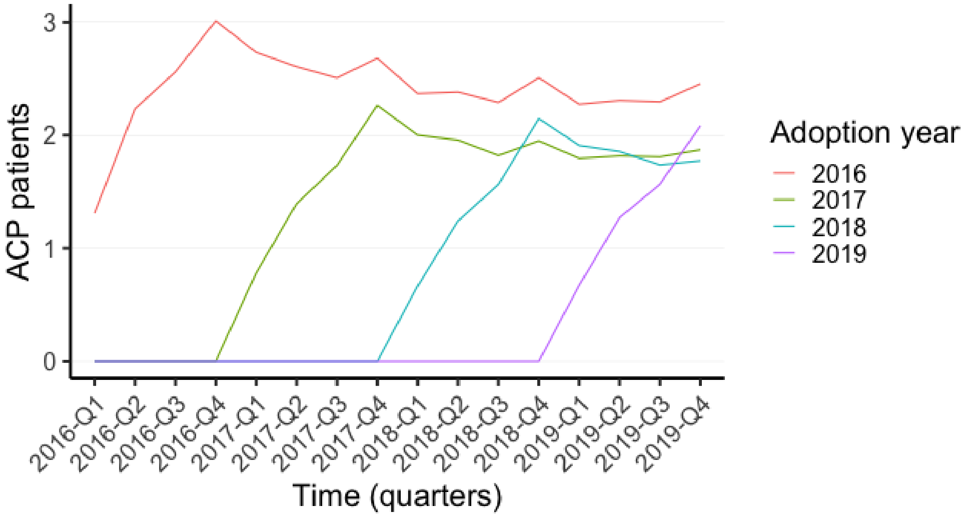
Notes. Estimates of regression equation (2.1) for physicians of top 3 specialties by ACP propensity with revenue from HCPCS code 99214 in panel (a) and G0439 in panel (b).

Figure 2.C.5: Event-study estimates for E&M revenue excluding ACP per patient by patient group



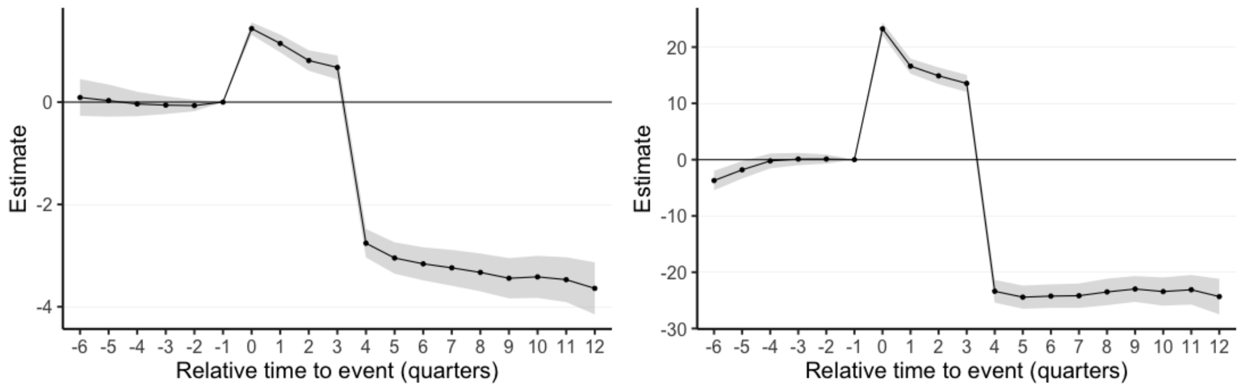
Notes. Estimates of modified regression equation (2.2) for physicians of top 3 specialties by ACP propensity. Patients are grouped by time between their ACP and the time of physician's ACP adoption in years (patients who never had ACP are the reference group).

Figure 2.C.6: Average number of patients receiving ACP services by year of ACP adoption



Notes. Quarterly averages across physicians of the number of unique beneficiaries who had an ACP procedure provided by the physician in question.

Figure 2.C.7: Event-study estimates of differences between "early" and "late ACP" patients for additional outcomes


(a) Number of patients

(b) Revenue per claim line

Notes. Estimates of regression equation (2.2) for physicians of top 3 specialties by ACP propensity. Claim lines correspond to specific procedures provided.

Bibliography

ABADIE, A. AND M. D. CATTANEO (2018): "Econometric methods for program evaluation," *Annual Review of Economics*, 10, 465–503.

ABALUCK, J., M. C. BRAVO, P. HULL, AND A. STARC (2021): "Mortality effects and choice across private health insurance plans," *The Quarterly Journal of Economics*.

ABALUCK, J. AND J. GRUBER (2011): "Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program," *American Economic Review*, 101, 1180–1210.

AGARWAL, S., X. GABAIX, J. C. DRISCOLL, AND D. LAIBSON (2009): "The age of reason: Financial decisions over the life cycle and implications for regulation," *Brookings Papers on Economic Activity*, 51–101.

AGHA, L. (2014): "The effects of health information technology on the costs and quality of medical care," *Journal of Health Economics*, 34, 19–30.

ALZHEIMER'S ASSOCIATION (2021): "2021 Alzheimer's disease facts and figures," *Alzheimer's & Dementia*, 17, 327–406.

AMA (2013): "Category I and III code change application for CPT code 99497," *American Medical Association.*

——— (2023a): "CPT overview and code approval," `https://www.ama-assn.org/practice-management/cpt/cpt-overview-and-code-approval#:~:text=CPT%C2%AE%20code%3F-,What%20is%20a%20CPT%C2%AE%20code%3F,reporting%2C%20increase%20accuracy%20and%20efficiency.`, Accessed: 2023-03-06.

——— (2023b): "Criteria for CPT Category I and Category III codes," `https://www.ama-assn.org/practice-management/cpt/criteria-cpt-category-i-and-category-iii-codes`, Accessed: 2023-03-06.

AMERIKS, J., A. CAPLIN, M. LEE, M. D. SHAPIRO, AND C. TONETTI (2022): "Cognitive decline, limited awareness, imperfect agency, and financial well-being," National Bureau of Economic Research Working Paper No. 29634.

BARNATO, A. E., A. J. O'MALLEY, J. S. SKINNER, AND J. D. BIRKMEYER (2019): "Use of advance care planning billing codes for hospitalized older adults at high risk of dying: A national observational study," *Journal of Hospital Medicine*, 14, 229–231.

BELANGER, E., L. LOOMER, J. M. TENO, S. L. MITCHELL, D. ADHIKARI, AND P. L. GOZALO (2019): "Early utilization patterns of the new Medicare procedure codes for advance care planning," *JAMA Internal Medicine*, 179, 829–830.

BORUSYAK, K., X. JARAVEL, AND J. SPIESS (2022): "Revisiting event study designs: Robust and efficient estimation," Working Paper.

BRINKMAN-STOPPELENBURG, A., J. A. RIETJENS, AND A. V. D. HEIDE (2014): "The effects of advance care planning on end-of-life care: A systematic review," *http://dx.doi.org/10.1177/0269216314526272*, 28, 1000–1025.

BYRNE, M. M. AND P. THOMPSON (2000): "Death and dignity: Terminal illness and the market for non-treatment," *Journal of Public Economics*, 76, 263–294.

CALLAWAY, B. AND P. H. SANT'ANNA (2021): "Difference-in-differences with multiple time periods," *Journal of Econometrics*, 225, 200–230.

CCW (2020): "27 CCW Chronic Conditions (1999-2020)," Tech. rep.

CHANDRA, A., C. COILE, AND C. MOMMAERTS (2020): "What can economics say about Alzheimer's disease?" National Bureau of Economic Research Working Paper No. 27760.

CHANDRA, A., E. FLACK, AND Z. OBERMEYER (2021): "The health costs of cost-sharing," National Bureau of Economic Research Working Paper No. 28439.

CHRISTELIS, D., T. JAPPELLI, AND M. PADULA (2008): "Cognitive abilities and portfolio choice," CFS Working Paper No. 2008/35.

CMS (2023): "HCPCS coding questions," `https://www.cms.gov/medicare/coding/medhcpcsgeninfo/hcpcs_coding_questions`, Accessed: 2023-03-06.

CMS (2023): "Physician Fee Schedule for year 2016," `https://www.cms.gov/medicare/physician-fee-schedule/search`, Accessed: 2023-03-10.

Curto, V., L. Einav, A. Finkelstein, J. Levin, and J. Bhattacharya (2019): "Health care spending and utilization in public and private medicare," *American Economic Journal: Applied Economics*, 11, 302–332.

Cutler, D., J. S. Skinner, A. D. Stern, and D. Wennberg (2019): "Physician beliefs and patient preferences: A new look at regional variation in health care spending," *American Economic Journal: Economic Policy*, 11, 192–221.

Cutler, D. M. (2015): "Payment Reform Is About to Become a Reality," *JAMA*, 313, 1606–1607.

Dafny, L. S. (2005): "How do hospitals respond to price changes?" *American Economic Review*, 95, 1525–1547.

Deb, A., J. D. Thornton, U. Sambamoorthi, and K. Innes (2017): "Direct and indirect cost of managing Alzheimer's disease and related dementias in the United States," *Expert Review of Pharmacoeconomics & Outcomes Research*, 17, 189–202.

Dobkin, C., A. Finkelstein, R. Kluender, and M. J. Notowidigdo (2018): "The economic consequences of hospital admissions," *American Economic Review*, 108, 308–352.

Dranove, D., C. Garthwaite, C. Heard, and B. Wu (2022): "The economics of medical procedure innovation," *Journal of Health Economics*, 81, 102549.

Einav, L., A. Finkelstein, T. Oostrom, A. Ostriker, and H. Williams (2020): "Screening and selection: The case of mammograms," *American Economic Review*, 110, 3836–3870.

Fagerlin, A. and C. E. Schneider (2004): "Enough. The failure of the living will." *The Hastings Center report*, 34, 30–42.

Fang, H. and Q. Gong (2017): "Detecting potential overbilling in Medicare reimbursement via hours worked," *American Economic Review*, 107, 562–591.

Finkelstein, A., M. Gentzkow, and H. Williams (2016): "Sources of geographic variation in health care: Evidence from patient migration," *The Quarterly Journal of Economics*, 131, 1681–1726.

Fisher, E. S., J. P. Bynum, and J. S. Skinner (2009): "Slowing the growth of health care costs – lessons from regional variation," *New England Journal of Medicine*, 360, 849–852.

Fisher, E. S., D. E. Wennberg, T. A. Stukel, D. J. Gottlieb, F. L. Lucas, and É. L. Pinder (2003): "The implications of regional variations in Medicare spending. Part 2: Health outcomes and satisfaction with care," *Annals of Internal Medicine*, 138, 288–298.

Gawande, A. (2016): "Quantity and quality of life," *JAMA*, 315, 267.

Gerteis, J., D. Izrael, D. Deitz, L. LeRoy, R. Ricciardi, T. Miller, and J. Basu (2014): "Multiple chronic conditions chartbook," *Rockville, MD: Agency for Healthcare Research and Quality*, 7–14.

Goldfeld, K. S., M. B. Hamel, and S. L. Mitchell (2013): "The cost-effectiveness of the decision to hospitalize nursing home residents with advanced dementia," *Journal of Pain and Symptom Management*, 46, 640–651.

Gorina, Y. and E. A. Kramarow (2011): "Identifying chronic conditions in Medicare claims data: Evaluating the Chronic Condition Data Warehouse algorithm," *Health Services Research*, 46, 1610.

HCCI (2020): "Health Care Cost and Utilization Report: 2020," Tech. rep.

Hoffman, G. J., D. T. Maust, M. Harris, J. Ha, and M. A. Davis (2022): "Medicare spending associated with a dementia diagnosis among older adults," *Journal of the American Geriatrics Society*, 70, 2592–2601.

Hurd, M. D., P. Martorell, A. Delavande, K. J. Mullen, and K. M. Langa (2013): "Monetary costs of dementia in the United States," *New England Journal of Medicine*, 368, 1326–1334.

Imbens, G. W. (2015): "Matching methods in practice: Three examples," *Journal of Human Resources*, 50, 373–419.

Keane, M. P. and S. Thorp (2016): "Complex decision making: the roles of cognitive limitations, cognitive decline and ageing," *Handbook of the Economics of Population Ageing*, 61, 0–70.

Kelley, A. S., K. McGarry, R. Gorges, and J. S. Skinner (2015): "The burden of health care costs for patients with dementia in the last 5 years of life," *Annals of Internal Medicine*, 163, 729–736.

KESSLER, D. P. AND M. B. MCCLELLAN (2004): "Advance directives and medical treatment at the end of life," *Journal of Health Economics*, 23, 111–127.

KFF (2022): "Medicare Advantage in 2022: Enrollment update and key trends," https://www.kff.org/medicare/issue-brief/medicare-advantage-in-2022-enrollment-update-and-key-trends/, Accessed: 2022-10-04.

——— (2023): "Total number of Medicare beneficiaries by type of coverage," https://www.kff.org/medicare/state-indicator/total-medicare-beneficiaries, Accessed: 2023-03-06.

KRUMHOLZ, H. M., S. L. T. NORMAND, AND Y. WANG (2019): "Twenty-year trends in outcomes for older adults with acute myocardial infarction in the United States," *JAMA Network Open*, 2, e191938–e191938.

KWOK, C. S., S. BENNETT, Z. AZAM, V. WELSH, R. POTLURI, Y. K. LOKE, AND C. D. MALLEN (2021): "Misdiagnosis of acute myocardial infarction: A systematic review of the literature," *Critical pathways in cardiology*, 20, 155–162.

LEE, B. K., J. LESSLER, AND E. A. STUART (2010): "Improving propensity score weighting using machine learning," *Statistics in Medicine*, 29, 337–346.

LI, J., C. ANDY, AND S. MITCHELL (2021): "Use of Medicare's New Reimbursement Codes for Cognitive Assessment and Care Planning, 2017-2018," *JAMA Network Open*, 4, e2125725–e2125725.

LIN, P. J., Y. ZHONG, H. M. FILLIT, E. CHEN, AND P. J. NEUMANN (2016): "Medicare expenditures of individuals with Alzheimer's disease and related dementias or mild cognitive impairment before and after diagnosis," *Journal of the American Geriatrics Society*, 64, 1549–1557.

LOCHNER, K. A. AND C. S. COX (2013): "Prevalence of multiple chronic conditions among Medicare beneficiaries, United States, 2010," *Preventing chronic disease*, 10, 1–10.

LUSTBADER, D., M. MUDRA, C. ROMANO, E. LUKOSKI, A. CHANG, J. MITTELBERGER, T. SCHERR, AND D. COOPER (2017): "The impact of a home-based palliative care program in an Accountable Care Organization," *Journal of Palliative Medicine*, 20, 23–28.

MEDICARE LEARNING NETWORK (2023): "MLN fact sheet: Advance care planning," `https://www.cms.gov/outreach-and-education/medicare-learning-network-mln/mlnproducts/downloads/advancecareplanning.pdf`, Accessed: 2023-03-07.

MITCHELL, S. L., J. M. TENO, D. K. KIELY, M. L. SHAFFER, R. N. JONES, H. G. PRIGERSON, L. VOLICER, J. L. GIVENS, AND M. B. HAMEL (2009): "The clinical course of advanced dementia," *New England Journal of Medicine*, 361, 1529–1538.

MORRISON, R. S. AND A. L. SIU (2000): "Survival in end-stage dementia following acute illness," *Journal of the American Medical Association*, 284, 47–52.

MULLAINATHAN, S. AND J. SPIESS (2017): "Machine learning: An applied econometric approach," *Journal of Economic Perspectives*, 31, 87–106.

National Institutes of Health (2022): "Estimates of funding for various research, condition, and disease categories (RCDC)," https://report.nih.gov/funding/categorical-spending#/, Accessed: 2022-10-04.

Ng, J. B., M. Turek, and A. M. Hakim (2013): "Heart disease as a risk factor for dementia," *Clinical Epidemiology*, 5, 135.

Nicholas, L. H., K. M. Langa, J. P. W. Bynum, and J. W. Hsu (2020): "Financial presentation of Alzheimer disease and related dementias," *JAMA Internal Medicine*.

Parekh, A. K., R. Kronick, and M. Tavenner (2014): "Optimizing health for persons with multiple chronic conditions," *JAMA*, 312, 1199–1200.

Reich, A. J., G. Jin, A. Gupta, D. Kim, S. Lipstiz, H. G. Prigerson, J. Tjia, K. Ladin, S. D. Halpern, Z. Cooper, and J. S. Weissman (2020): "Utilization of ACP CPT codes among high-need Medicare beneficiaries in 2017: A brief report," *PLoS One*, 15, 1–8.

Roe, M. T., J. C. Messenger, W. S. Weintraub, C. P. Cannon, G. C. Fonarow, D. Dai, A. Y. Chen, L. W. Klein, F. A. Masoudi, C. McKay, K. Hewitt, R. G. Brindis, E. D. Peterson, and J. S. Rumsfeld (2010): "Treatments, trends, and outcomes of acute myocardial infarction and percutaneous coronary intervention," *Journal of the American College of Cardiology*, 56, 254–263.

Sacarny, A. (2018): "Adoption and learning across hospitals: The case of a revenue-generating practice," *Journal of Health Economics*, 60, 142–164.

Schwartz, A. L., B. E. Landon, A. G. Elshaug, M. E. Chernew, and J. M. McWilliams (2014): “Measuring low-value care in Medicare,” *JAMA Internal Medicine*, 174, 1067–1076.

Song, Y., J. Skinner, J. Bynum, J. Sutherland, J. E. Wennberg, and E. S. Fisher (2010): “Regional variations in diagnostic practices,” *New England Journal of Medicine*, 363, 45–53.

Sun, L. and S. Abraham (2020): “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects,” *Journal of Econometrics*, 225, 175–199.

Teno, J., J. Lynn, A. F. Connors Jr., N. Wenger, R. S. Phillips, C. Alzola, D. P. Murphy, N. Desbiens, and W. A. Knaus (1997): “The Illusion of End-of-Life Resource Savings with Advance Directives,” *Journal of the American Geriatrics Society*, 45, 513–518.

Teno, J. M. (2000): “Advance directives for nursing home residents: Achieving compassionate, competent, cost-effective care,” *JAMA*, 283, 1481–1482.

Thorpe, K. E., L. L. Ogden, and K. Galactionova (2010): “Chronic conditions account for rise in medicare spending from 1987 to 2006,” *Health Affairs*, 29, 718–724.

Tolle, S. W. and J. M. Teno (2017): “Lessons from Oregon in embracing complexity in end-of-life care,” *New England Journal of Medicine*, 376, 1078–1082.

Tulsky, J. A. (2005): “Beyond advance directives: Importance of communication skills at the end of life,” *JAMA*, 294, 359–365.

Woloshin, S. and A. S. Kesselheim (2022): "What to know about the Alzheimer drug aducanumab (Aduhelm)," *JAMA Internal Medicine*, 182, 892–892.

Wright, M. S. (2018): "Change without change? Assessing Medicare reimbursement for advance care planning," *Hastings Center Report*, 48, 8–9.

Yadav, K. N., N. B. Gabler, E. Cooney, S. Kent, J. Kim, N. Herbst, A. Mante, S. D. Halpern, and K. R. Courtright (2017): "Approximately one in three us adults completes any type of advance directive for end-of-life care," *Health Affairs*, 36, 1244–1251.

Yu, W., A. Ravelo, T. H. Wagner, C. S. Phibbs, A. Bhandari, S. Chen, and P. G. Barnett (2003): "Prevalence and costs of chronic conditions in the VA health care system," *Medical Care Research and Review*, 60, 146S–167S.

Zeitoun, N. K. (2015): "New Medicare rule will reimburse physicians for advance care planning," *The Hospitalist*, Accessed: 2023-01-16.

Zeltzer, D., L. Einav, A. Finkelstein, T. Shir, S. Stemmer, and R. Balicer (2020): "Why is end-of-life spending so high? Evidence from cancer patients," National Bureau of Economic Research Working Paper No. 28162.