NBER WORKING PAPER SERIES

MORTALITY EFFECTS OF HEALTHCARE SUPPLY SHOCKS: EVIDENCE USING LINKED DEATHS AND ELECTRONIC HEALTH RECORDS

Engy Ziedan
Kosali I. Simon
Coady Wing

Working Paper 30553
http://www.nber.org/papers/w30553

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2022

We would like to thank Diane Alexander, Sumedha Gupta, Alex Hollingsworth, Amanda Kowalski, Timothy Moore, Jim Poterba, Julian Reif, Adrienne Sabety, Daniel Sacks, Jon Skinner and Christopher Whaley for helpful comments. We would like to thank Claire Cravero, Quinn Johns, Alyssa Chen, Aleah Peffer, Isabel Francisco from the Covid-19 Research Database/ Datavant for facilitating our access to the Connected Death Index data and providing a secure computing environment to link the EMR and deaths records. We would like to thank Matt Shober, Rose Baumgardner, Bob Suhendra and Cliff Cavanaugh from Healthjump for answering our many questions about the data. We would also like to thank Danny Chang, Vishal Singh, Vamsi Bushan, Naba Sahoo and Madeline Yozwiak for excellent research assistance. This research was made possible by partial funding from the National Institutes of Health Contract No. 75N95D20F4000 The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Mortality Effects of Healthcare Supply Shocks: Evidence Using Linked Deaths and Electronic Health Records
Engy Ziedan, Kosali I. Simon, and Coady Wing
NBER Working Paper No. 30553
October 2022
JEL No. I0,I11

ABSTRACT

The contraction in health care consumption at the start of the pandemic provides insight into central economic questions of waste and productivity in the U.S. health care system. Using linked mortality and Electronic Medical Records, we compare people who had outpatient appointments scheduled for dates in 30 day periods immediately before and after the Covid-19 emergency declaration. Appointment cancellation rates were 77% higher for people with appointments in the shutdown period. Intent to treat estimates imply that having a scheduled appointment date right after the emergency declaration increased one-year mortality rates by 4 deaths per 10,000. Instrumental variable estimates suggest that a cancelled appointment increased one-year mortality by 29.7 deaths per 10,000 among compliers, implying that a 10% increase in health care appointments reduces mortality rates by 2.9%. The mortality effects are rooted in two mechanisms: a complier sub-population with high marginal benefits from care, and a cascade of delayed or missed follow-up care that lasted for about 3 months. Healthcare spending accounted for 19.7% of U.S. GDP in 2021, and controlling health spending is a major policy objective. Our results quantify health tradeoffs from cutting every-day non-emergency visits, illustrating the importance of cost-control efforts that differentiate between medical care with the largest and smallest benefits for patient health.

Engy Ziedan
Department of Economics
School of Liberal Arts
Tulane University
206 Tilton Hall
6823 St. Charles Avenue
New Orleans, LA 70118
eziedan@tulane.edu

Kosali I. Simon
O’Neill School of Public and
Environmental Affairs
Indiana University
1315 East Tenth Street
Bloomington, IN 47405-1701
and NBER
simonkos@indiana.edu

Coady Wing
Indiana University
1315 E 10th St
Bloomington, IN 47405
cwing@indiana.edu

1 Introduction

The onset of the Covid-19 pandemic led to a major contraction in outpatient health care utilization, as governments, patients, and health care providers took drastic measures to mitigate transmission. Many non-urgent health care appointments that were scheduled to occur during the first few weeks of the pandemic were cancelled. In this paper, we argue that the cancellation shock provides insight into central questions about waste and productivity in the health care system, and into the health consequences of disruptions to the regular operation of the health care system. Health care utilization varies substantially across geographic regions in the U.S. But previous work shows that population health outcomes are no better in areas with high utilization (Fisher et al., 2003; Baicker and Chandra, 2004; Wennberg et al., 2008). One interpretation is that much of the health care people consume is so-called "flat of the curve" medicine: care that is expensive but does little to improve patient health.

Measuring the trade-offs associated with reduced health care consumption is difficult because the health sector produces a vast range of services for heterogeneous patients. Crude comparisons of the subsequent health of people who consume different amounts of health care are likely confounded because baseline health status may affect health care consumption. A number of studies with strong causal research designs demonstrate the health benefits of more intensive treatment for specific acute conditions, such as heart attacks, strokes, and accidents. But these studies are not informative about the benefits of non-urgent health care services. About 20% of total health spending in the U.S. occurs in outpatient settings, where people receive well-care check-ups, diagnostic tests, minor and major surgeries, pre- and post-operation care, mental health counseling, and more. It is possible that some of these appointments may be on the "flat of the curve" while others are producing substantial health benefits. Broad cuts in health care consumption might be quite damaging if a substantial share of the foregone services were health improving. Under a strong version of the flat of the curve hypothesis – one where the flat part of the benefit curve is long and applies to much of the care people consume – large reductions in health care utilization would not lead to large changes in health outcomes.

Using linked mortality and Electronic Medical Records (EMR) data, we test the "flat of the curve" hypothesis by comparing two cohorts of patients. The February-March (control) cohort consists of people who were scheduled to have outpatient appointments between 13 February 2020 and 12 March 2020, which is a one-month window of dates immediately before the pandemic emergency declaration on 13 March 2020. The March-April (treatment) cohort consists of people who were scheduled to have an outpatient appointment on a date between 13 March 2020 and 12 April 2020, the one-month period immediately following the emergency declaration. The patients in the two groups were comparable at baseline along a range of demographic, health, and geographic characteristics. We tracked both groups over the 12 months following their scheduled appointment date. During that year, both groups experienced the same economic, social, and epidemiological conditions, making it unlikely that scheduled appointment dates are correlated with environmental conditions that might also have affected health.

Although both groups lived through the same pandemic risks, the March-April group experienced an unanticipated negative shock to health care consumption because of appointment cancellations induced by the emergency declaration. The appointment cancellation rate in the March-April cohort was nearly 34%, compared to 19% in the February-March cohort. Put differently, conditional on having scheduled an appointment, the appointment *completion rate* shifted from 81% in the February-March group to 66% in the March-April group. Thus, the shutdown created a sudden 18.5 percent contraction in regular health care use, affecting patients who were planning to consume a broad range of health services and procedures.

The contraction in health care use caused substantial damage to the health of the March-April scheduling cohort. One year after the focal appointment, the February-March control group experienced 83 deaths per $10,000$ appointments. In the March-April group – which experienced high cancellation rates – there were 87 deaths per $10,000$ appointments. The intent to treat (ITT) comparison implies that having an appointment scheduled during March-April rather than February-March generated about 4 extra deaths per 10,000 appointments, a 5 percent increase in the baseline mortality rate. These effects are not driven by Covid-19 deaths: we find that the March-April and February-March cohorts had approximately the same number of deaths among individuals with any recorded indication of a Covid-19 diagnosis.

The March-April shock led to $1,469$ extra cancellations per $10,000$ appointments. Scaling the ITT mortality effect by this first stage implies that an appointment cancellation increased one year mortality rates by about 30 per $10,000$ among compliers. The mortality-health care relationship in the outpatient care sector is relatively inelastic. Converting the effect estimates into percentage terms, we find that a 10% increase in health care appointments reduces mortality rates by slightly less than 3%. Nevertheless, cancelling a large number of appointments does induce a lot of mortality. The average effect among compliers implies that there was roughly 1 additional death for every 333 cancellations. Using a Years of Potential Life Lost (YPLL) approach, the mortality cost of $10,000$ cancelled appointments is about \$107 million, assuming life expectancy of 75 years and a Value of a Statistical Life Year (VSLY) of $\$150,000$. The average outpatient visit would have to cost more than $\$6,653$ before its cancellation outweighs the losses from increased mortality[1]. These results suggest that the average marginal effect of outpatient health services consumed in the U.S. has an economically important effect on patient health and well-being.

To understand the mortality results further, we examine two possible mechanisms. First, the compliers in our instrumental variable design may have a particularly high marginal benefit of health care. We find some evidence to support this claim. About 36% of our study sample is over age 65. In contrast, about 47% of the compliers are over 65. When we allow the ITT effect to differ by age group, we find that the cancellation shock had much larger mortality consequences for older patients. We also find that compliers had higher baseline comorbidity scores and that ITT effects are larger for comorbid patients. Taken together, these results suggest that compliers experience higher benefits of medical care. It is possible that patients and providers cancelled care in an effort to protect patients with higher (expected) Covid-19 risk. Unfortunately, these patients with more precarious health may be also the same patients who benefit most from medical care.

Second, the initial cancellation in our first stage could be the beginning of a longer cascade of missed and delayed care, causing our mortality results to arise from an accumulation of reduced health care use rather than from the missed focal visit alone. We find that missing the pre-scheduled focal appointment tipped off a period of low health care use in the March-April cohort that lasted about 1-3 months. For example, during the first month following the focal appointment, the March-April group had 35% fewer laboratory tests and 17% fewer follow-up appointments. These results suggest that delayed diagnosis and treatment may partly explain why the cancellation shock led to a mortality effect.

In addition to shedding light on the trade-offs involved in pursuing a leaner health care system, the cancellation shock can be interpreted as part of a mass triage that society used to cope with a public health threat. In that vein, our study measures the health consequences of disaster management in the early pandemic, and offers some perspective on the difficulties of navigating health care disruptions that may arise in other situations such as weather events, staff shortages, or power outages. Back of the envelope calculations suggest that the cancellation shock explains 1/4 to 1/3 of the non-Covid excess deaths that occurred between March 2020 and March 2021.

Our results suggest that the non-urgent health services people consume in outpatient

[1]In our data, in 2019, total charges up to 3 months after the visit date were on average \$313 which is considerably less than $\$6,653$

settings should not be considered *flat of the curve* medicine. Our study differs from most previous work in that we focus on outpatient medical care, and not health care for acute conditions such as heart attacks, strokes, accidents, injuries, and intensive care. One line of prior work leverages natural experiments in which similar patients are treated more vs less intensively along some dimension. For example, Doyle et al. (2015) exploits ambulance referral patterns to show that high spending hospitals produce lower mortality rates for heart attack patients. Similarly, Doyle (2007) finds that travellers who experience heart-related emergencies have lower mortality rates when treated at high spending hospitals. Almond et al. (2010) use a birth weight regression discontinuity design to show that expensive treatment in neonatal intensive care units reduces mortality rates among very low birth weight infants. Rose (2020) takes advantage of Medicare financing rules that exogenously lowers the out of pocket costs of skilled nursing facility (SNF) services among some post-acute hospital patients. The results show that recovering in a SNF reduces 30 day hospital readmission rates among people discharged from a hospital inpatient stay. Other work on screening and targeting in health care is concerned with efficient use of laboratory tests and other technologies (Einav et al., 2020; Lakdawalla et al., 2010; Welch and Black, 2010). Each of these quasi-experimental studies is designed to wrangle out the causal effects of a particular type of health care on downstream health outcomes.

Another strand of research exploits exogenous variation in health insurance coverage or generosity to study the causal effects of insurance coverage on health (Card et al., 2009; Simon et al., 2017; Miller et al., 2021; Newhouse et al., 1993; Goldin et al., 2021). These studies may shed light on the health benefit of health care across a wider range of services, a feature that is similar to our approach. A fundamental difference is that these studies measure the impact of *health insurance*, which may be somewhat different from the impact of health care in nuanced ways. Changes in insurance coverage or design may affect more than a person's health care consumption. For example, health insurance may provide financial stability (Finkelstein et al., 2012; Barcellos and Jacobson, 2015; Gross and Notowidigdo, 2011) and could affect labor force participation (Baicker et al., 2014; Garthwaite et al., 2014). These factors may also affect health outcomes. In addition, the specific contractions or expansions in health care consumption in insurance studies may be rooted in plan parameters (deductibles, coinsurance, narrow provider networks, capitation, etc) that are designed to discourage low value health care utilization that will have a negligible effect on health outcomes.

Our study complements these existing lines of research and makes several contributions to the economics literature. First, we study plausibly exogenous variation the use of non-urgent health services provided in the outpatient settings. This is distinct from earlier studies that emphasize health care for acute and emergent health conditions, and from studies based around health insurance variation. In addition, our study provides some insight into the health effects of efforts to manage the pandemic emergency by distributing health care using a form of *triage* or *non-price rationing* – cancelling appointments deemed non-essential. This approach may yield different results than efforts to reduce health expenditures through insurance plan designs that discourage moral hazard.

We also innovate by using new data that allows near real-time tracking of mortality combined with healthcare records of a generalizable population. We leverage a large Electronic Medical Records (EMR) database called Healthjump, which covers approximately 70 million patient records from outpatient facilities located across the country. The Healthjump data allows us to exploit our study design by constructing cohorts based on appointment scheduling data, and to determine whether scheduled appointments were cancelled. This is a key advantage that is typically not feasible with other data sources, such as insurance claims or surveys. To study mortality, we exploit a novel linkage between individual EMR records with date of death information from the Connected Death Index (CDI) using encrypted privacy tokens. In the past, researchers have relied on identifiable records from the Social Security Death Master File (DMF) when linking private sector health care data to mortality status. However, the quality of the DMF has diminished since 2011, when the SSA stopped reporting death information gleaned from state government sources. In recent years, the DMF has captured only $11 - 18\%$ of deaths, limiting its value for research purposes (Datavant, 2020b). The Connected Death Index (CDI) uses obituary data from online newspapers, funeral homes, and memorials to supplement DMF, capturing 82% of all US deaths in 2020 (Levin et al., 2019). The discovery and linkage of these data provides a powerful way to measure the downstream health outcomes of the patients in EMRs, even if they have no subsequent contact with any of the health care providers in the system.

The remainder of our paper explains the study and results in more detail. In Section 2, we describe the research design and identifying assumptions underlying our study. Section 3 introduces the data sources and explains the inclusion criteria used to construct the study population. We present the results and key sensitivity analysis in Section 4. Section 5 discusses our study in the context of related work on the returns to medical care and offers conclusions.

2 Research Design

Identifying the causal effects of cancelled outpatient appointments is difficult because cancellations are not usually randomly assigned. Sicker people may be less likely to cancel appointments than healthier people, for example. During the pandemic, the opposite may also be true if sicker patients cancelled at higher rates in an effort to avoid exposure to Covid-19. As a result, simply comparing people with cancelled vs maintained appointments may confound the causal effects of health care with pre-existing differences in health status. To mitigate confounding, we examine a natural experiment created by disruptions in health care delivery during the early weeks of the Covid-19 pandemic.

To clarify the design, let $j = 1...J$ index the population of scheduled outpatient appointments that are contained in the EMR booking data. Each appointment belongs to a specific patient and is scheduled to occur on a particular date. There are $i = 1...N$ unique patients in the appointment population, and we let $i(j)$ be a function that maps the appointment index into the relevant patient index. Similarly, $t(j)$ gives the calendar date that appointment j was scheduled to take place. Our main study population includes all appointments with 13 February 2020 $\leq t(j) \leq$ 13 April 2020.

We partition the study population into two scheduling cohorts: February-March and March-April. Specifically, let $z_{ij} = 1[t(j) \geq 13 \text{ March } 2020]$ be a binary variable set to 1 if appointment j was scheduled to occur between 13 March 2020 and 13 April 2020, and set to 0 if it was scheduled to occur between 13 February 2020 and 12 March 2020. The March-April group – with $z_{ij} = 1$ – is the treatment cohort, consisting of patient-appointments scheduled to occur in the first month of the emergency declaration on 13 March 2020.[2] The February-March group – with $z_{ij} = 0$ – is the control cohort, consisting of appointments scheduled for the month before the emergency declaration.

Let c_{ij} be a binary cancellation indicator set to 1 if the appointment was cancelled and set to 0 if the patient actually had a healthcare encounter on date $t(j)$. Next, let q_{ij}^f measure the quantity of outpatient health care that patient i consumed over the f month period following the focal appointment date, $t(j)$. In our empirical work, we examine several measures of utilization, including follow up appointments, laboratory tests, and total health care charges. m_{ij}^f represents patient $i's$ mortality status f months after the scheduled appointment. For

[2] The state of emergency declaration was only 2 days after the World Health Organization declared the coronavirus a pandemic. In early March, there was no inclination that worldwide shutdowns were about to begin. For example, on March 6th 2020 the front page of the New York Times confirmed that Bernie Sanders and Joe Biden will be the two candidates on stage in an upcoming debate. The sudden decline in human mobility after March 13th has also been confirmed by numerous papers on mobility trends in the United States (Gupta et al., 2020).

example, m_{ij}^{12} is a binary variable indicating whether the patient is deceased 12 months after the focal appointment date. m_{ij}^{6} would measure mortality status 6 months after the appointment date. The mortality measures are defined cumulatively so that $m_{ij}^{f} \leq m_{ij}^{f+1}$.

To define the causal effects of the pandemic emergency on appointment cancellations, we use $c_{ij}(z)$ to represent the potential cancellation outcome under alternative initial scheduling. That is $c_{ij}(0)$ indicates whether the patient-appointment would have been cancelled if it had been scheduled for a date in the February-March window, and $c_{ij}(1)$ indicates whether same patient-appointment would have been cancelled if it had instead been scheduled for a date in the March-April window. The unit specific causal effect of the pandemic shock on cancellation is $c_{ij}(1) - c_{ij}(0)$.

Finally, let $q_{ij}^{f}(c, z)$ and $m_{ij}^{f}(c, z)$ represent person $i's$ potential health care utilization and potential mortality outcomes, during the period following appointment j. Here, we use $c = [0, 1]$ and $z = [0, 1]$ to index the four possible combinations of scheduling cohort assignments and appointment cancellations outcomes. For example, $m_{ij}^{f}(1, 0)$ represents the potential health outcome that a person would experience if she were assigned to the March-April scheduling cohort and did not experience a cancelled appointment. $m_{ij}^{f}(1, 1)$ represents the same person's health outcome if she was instead assigned to the March-April cohort but did experience an appointment cancellation.

Throughout, we assume that responses to the scheduling instrument and to cancellation are individualistic, ruling out spillover effects. In addition, we maintain the four instrumental variable assumptions listed below (Angrist et al., 1996).

Assumption 1 *Cancellation rates are different in the March-April and February-March cohorts, such that* $Pr(c_{ij} = 1 | z_{ij} = 1]) \neq Pr(c_{ij} = 1 | z_{ij} = 0)$.

Assumption 2 *Membership in the March-April scheduling cohort is statistically independent of potential outcomes so that* $z_{ijt} \perp\!\!\!\perp (m_{ij}^{f}(c, z), q_{ij}^{f}(c, z), c_{ij}(z))$.

Assumption 3 *The March-April scheduling cohort satisfies an exclusion restriction, meaning that it has no direct causal effect on post-treatment mortality or utilization:*

$$m_{ij}^{f}(c, z) = m_{ij}^{f}(c)$$
$$q_{ij}^{f}(c, z) = q_{ij}^{f}(c)$$

Assumption 4 *Having an appointment scheduled in the March-April window weakly increases cancellation rates, implying* $c_{ij}(1) \geq c_{ij}(0)$ *for all* $j = 1...J$.

Assumption 1 is the first stage (relevance) condition, which is testable by comparing cancellation rates across the scheduling cohorts. We find a substantial difference in cancellation rates in the data. One concern is that cancellation rates are always higher for March-April appointments than February-March appointments, and that the cancellation rates we observe in 2020 are not generated by the pandemic shock but by recurring patterns related to hospital staffing, vacation plans, and weather conditions. To assess this concern, we also examine data from February-March and March-April scheduling cohorts in 2019 and compare the difference in cancellation rates. We find no evidence of any seasonal patterns that might account for the first stage in 2020.

Assumption 2 implies that – among appointments scheduled in the two month period near the start of the pandemic –people are as good as randomly assigned to appointments in the February-March and March-April cohorts. This *independence assumption* is plausible because the appointments are typically scheduled several weeks in advance. As a result, the appointments in the March-April cohort were made before patients took the pandemic into consideration when scheduling health care appointments. The independence assumption would likely fail if we expanded our study population to include people with scheduled appointments in May and June of 2020. It is very likely that people whose scheduled appointment dates fell in that window would had called in their health care appointments after the pandemic set in. To partially validate the independence assumption, we use historical medical records and patient demographic information to construct balancing tests comparing the February-March and March-April scheduling cohorts. Lack of balance would undermine the independence assumption. We find that the two groups are very comparable on baseline characteristics.

Assumption 3 rules out the possibility that appointment scheduling has a direct causal effect on downstream outcomes. We think the exclusion restriction is credible in our study because both the February-March and March-April scheduling cohorts contain patients who experienced the same basic public health conditions of the Covid-19 pandemic. The key difference is that the February-March cohort was better able to receive scheduled health services due to pre-pandemic scheduling variation. The exclusion restriction might well fail if we allowed the time period of the study to start much earlier than mid-February. For example, if we included people with appointments scheduled in – say – September of 2019 as a comparison group, then that group's health outcomes (measured several months downstream from the appointment date) would mostly have been realized prior to the epidemic.

In this case, the March-April treatment cohort would still have higher cancellation rates than a September 2019 control cohort, but the treatment cohort would have experienced different epidemiological conditions that might have generated higher mortality even in the absence of cancellation. Thus, the exclusion restriction has credibility when the treatment and comparison groups are defined so that both groups face similar epidemiological conditions during the follow up period, but it might fail in more general settings when the treatment and control cohorts are defined too far apart in appointment time. We explore possible violations of the exclusion restriction in more detail later in the paper.

Assumption 4 is a monotonicity condition that is easiest to interpret using the complier-group terminology developed by Angrist et al. (1996). In our framework, *always takers* consist of person-appointments with $c_{ij}(0) = c_{ij}(1) = 1$. These appointments are cancelled regardless of scheduling dates, perhaps because the person had recovered from illness. *Never takers* have $c_{ij}(0) = c_{ij}(1) = 0$. These appointments are kept (not cancelled) regardless of scheduling dates, perhaps because the appointment was considered so critical that it could not be skipped even during the early pandemic period. Compliers have $c_{ij}(1) - c_{ij}(0) = 1$, implying the appointment would have been kept if not for the pandemic emergency. Finally, appointments with $c_{ij}(1) - c_{ij}(0) = -1$ are defiers, which means the appointment would be kept because of the pandemic and cancelled in its absence. The monotonicity assumption implies that there are no defiers. It is difficult to validate the monotonicity assumption, but it seems reasonable to view the pandemic emergency as nudging people towards cancellation.

The difference in mean realized outcomes after f months in the March-April and February-March group gives a reduced form *Intent To Treat* (ITT) effect of the cancellation shock. The mortality ITT is:

$$
\begin{aligned}
ITT_m^f &= E[m_{ij}^f | z_{ij} = 1] - E[m_{ij}^f | z_{ij} = 0] \\
&= E[m_{ij}^f(1) - m_{ij}^f(0) | c_{ij}(1) - c_{ij}(0) = 1] Pr(c_{ij}(1) - c_{ij}(0) = 1) \\
&= E[\Delta_{ij}^f | c_{ij}(1) - c_{ij}(0) = 1] Pr(c_{ij}(1) - c_{ij}(0) = 1).
\end{aligned}
$$

The ITT_m^f represents the effect of the pandemic-induced contraction of the health care sector of the economy on downstream mortality.[3] The first stage difference in cancellation rates in the two scheduling groups recovers the complier share:

[3] A similar contrast – ITT_q^f – represents the reduced form effect of the contraction on downstream utilization.

$$F = E[c_{ij}|z_{ij} = 1] - E[c_{ij}|z_{ij} = 0]$$
$$= Pr(c_{ij}(1) - c_{ij}(0) = 1)$$

The Wald Ratio gives the standard complier average treatment effect (CATE):

$$\frac{ITT_h}{F} = \frac{E[h_{ij}|z_{ij} = 1] - E[h_{ij}|z_{ij} = 0]}{E[c_{ij}|z_{ij} = 1] - E[c_{ij}|z_{ij} = 0]}$$

$$= E[\Delta_{ij}|c_{ij}(1) - c_{ij}(0) = 1]$$

The CATE represents the average causal health effect of cancelled appointments among those who were induced to cancel because of the contraction in health care during the early pandemic emergency. We estimate the ITT and first stage relationships using simple OLS regressions and we estimate the CATE parameter using two stage least squares. Some patients were scheduled for more than one appointment during our study window. To avoid selection bias, we retain all of these patient-appointments, which means that some patients contribute more than one observation to the analysis. Accordingly, we estimate standard errors using a heteroskedasticity and cluster robust variance matrix, which allows for dependence across observations on the same patient.

We probe the credibility of key assumptions and test the sensitivity of our results in several ways. We test the independence assumption by comparing pre-appointment characteristics among members of the March-April cohort and members of the February-March cohort. Specifically, we examine differences in age, gender, race/ethnicity, geographic location, and baseline health (Charlson Scores and predicted mortality risk). Seasonality in mortality or health care utilization is an important alternative explanation for our results. To assess these concerns, we conduct placebo tests in which we repeat our analysis using data from 2019. In that analysis, we use appointments scheduled to occur between 13 March 2019 and 13 April 2019 as a placebo treatment group and appointments scheduled to occur between 13 February 2019 to 12 March 2019 as the corresponding placebo control group. We compute the same first stage and ITT comparisons for these 2019 groups as we do for 2020, providing insight into the prevailing seasonality in scheduling, health care use, and health outcomes that could be alternative explanations for our results.

3 Data

3.1 Electronic Health Records – Healthjump

Healthjump is a platform for managing the collection, storage, and movement of clinical and financial data between EHRs, applications and healthcare organizations. Healthjump's clients include physician practices, clinics, Accountable Care Organizations (ACOs), some hospitals and healthcare technology providers. They contract with Healthjump to solve interoperability challenges that arise because patients receive care at providers that use different EMR vendors such as Cerner, Epic, and NextGen. Healthjump's data management platform extracts and standardizes EMR data from multiple systems into a common data model. It remits the data to clients on a daily basis, so the database is updated in nearly real time.

As of November 2021, the Healthjump EMR contains records for almost 79 million unique patients, including patients in every state. Patient encounters with participating providers are captured regardless of payment, insurance coverage, or changes in insurance carriers.[4] The encounters in the Healthjump data mostly take place in outpatient settings because the participating providers are mainly physician practice groups and ACOs rather than hospitals. This is a disadvantage of the Healthjump EMR data compared with insurance claims databases, which include inpatient and ER. On the other hand, claims data do not include uncompensated care or care paid by other sources, which are included in Healthjump. Measures of total health care use based on Healthjump will differ from total utilization to the extent that patients receive services from non-participating outpatient providers or in inpatient or ER settings. This limitation is unlikely to affect the internal validity of our study design, as focal appointment cancellation rates will be measured similarly for control and treatment groups, and mortality rates are measured from outside sources. However, readers should keep in mind that our analysis assumes that February-March and March-April cohorts receive the same amount of care from non-Healthjump participating providers, so our estimates would correctly measure any post-treatment differences in utilization between the groups.

The Healthjump database includes files for appointments, encounters, procedures, diagnoses, lab orders, lab results, vital statistics, charges, and medications. Each patient pos-

[4] Ziedan et al. (2020) compared the Healthjump sample with the National Ambulatory Care Survey (NAMCS) as a benchmark. They find that Healthjump has national coverage but is not perfectly representative. Patients from LA, MS, AK, and NC are somewhat overrepresented; those from NY, NH, and NE are underrepresented, and Healthjump patients were slightly older than the NAMCS patients.

sesses a unique identifier, allowing tracking across providers. The appointments file contains information on each patient's scheduled appointment dates, appointment duration, and the provider identifier. We use the appointment file to extract a subset of appointments scheduled to occur on specific dates, and to determine whether the appointments were cancelled or maintained.[5]

It is important to define what we mean by outpatient care. Generally, any healthcare delivered without the need to stay in a hospital is referred to as ambulatory care or outpatient care. That includes diagnostic tests, treatments, or office visits. This care generally occurs in emergency rooms, outpatient clinics, outpatient clinics within hospitals or urgent care clinics. In our setup, since we study appointments that are scheduleable, we exclude emergency or urgent care from this list, and refer to the resulting set as outpatient care.

3.2 Mortality – Connected Death Index

Our mortality data comes from the Connected Death Index (CDI) through a technology company called Datavant. Datavant provides a privacy preserving record linkage technology that connects third party sources from online obituary data gleaned from newspapers, funeral homes, and memorials for individuals in the U.S. and Canada; it also uses information from the Social Security Administration (SSA) Death Master File (DMF) when it is available. The CDI contains approximately 100 million historical deaths, from 1899 onward with roughly 50 thousand deaths added every week. The use of obituary data for tracking mortality has become an important way to track deaths as other sources have declined in quality and coverage. In 2011, the SSA made changes to the DMF in response to new regulations that prohibits SSA from disclosing death records obtained directly from states and not from federal agencies. These changes resulted in the removal of approximately 4 million historical records, and the DMF now captures approximately 1 million (approximately 40%) fewer records of deceased per year (Rothwell, 2022). That also means the DMF now is more likely to miss individuals not on Medicare and those who do not receive SSA benefits. The decline in the DMF's capture rate created a market where technology firms scrape online obituaries for identifiable death information that is later sold to companies such as pharmaceutical firms seeking identifiable data on mortality. The tools used to collect obituary data and connect it in the CDI have improved over time. In recent years,the CDI has captured about

[5]The appointments file also contains the date and time the appointment first appeared in the EMR system. This date is *either* the date the visit itself was booked *or* the date the healthcare provider first integrated with Healthjump, whichever comes last. Thus, for a subset of Healthjump client providers, we can determine when the appointment date was originally booked, which provides some insight into advance booking times during our study period.

82-85% of new deaths in the US; see Data Appendix Figure 2.

The CDI has two main limitations for our study. First, it does not include the cause of death. In some of our analysis, we use the EMR data to distinguish *suspected* Covid and non-Covid deaths, based on classifying decedents who had a Covid-19 or related diagnosis in the three months prior to the month of death as suspected Covid deaths.[6] Second, the CDI undercounts recent deaths. A validation study by the Society of Actuaries, (Du et al., 2020) found that the CDI captures about 86% of all deaths in 2018, and that the capture rate is somewhat higher for older decedents. The CDI undercount is not likely an important threat to the internal validity of our study because there is no reason to expect undercount rates to differ systematically between the February-March and March-April cohorts.

3.3 Record Linkage

The version of the CDI we obtained includes several Datavant "tokens", which are produced by proprietary algorithms developed by Datavant to facilitate secure record linkages. In essence, Datavant assigns each individual in the population a collection of up to 17 encrypted tokens that are generated from different pieces of underlying information, such as SSN, date of birth, last name, first initial, soundex(last name), etc. Each death record includes multiple tokens to enable linkage with other databases on the basis of a mixture of underlying information. Some of the tokens involve a deterministic match on underlying information. Other tokens involve a probabilistic or *fuzzy* match on underlying information.

Datavant's algorithms were used to contruct the same set of 17 tokens for each patient in the Healthjump EMR. We used these tokens to link the CDI and Healthjump records in a hierarchical fashion that prioritized matches on SSN and on non-SSN tokens that minimize the risk of incorrectly classifying living people as deceased.[7] Data Appendix A provides more information on our approach to linking the Healthjump EMR and CDI mortality data.

We studied the quality of the linkage in three ways. First, for patients who were linked on SSN, we checked whether linking on deterministic terms from other tokens would yield different results. When both deterministic and probabilistic terms are available, we found that the death status does not change if we match on one vs the other. Second, using the entire sample of patients who were identified as deceased, we looked for patients who appear to receive health services (new prescriptions, laboratory tests, procedures, or visits) after

[6] Data Appendix D lists the ICD codes we use, obtained from Sacks et al. (2020)

[7] Datavant tokens have previously been used by Datavant's clients, which are mostly pharmaceutical firms. Datavant has studied the accuracy of linkages based on each of the 17 tokens, and has ordered the tokens from most to least accurate. Datavant (2020a) describes the tokens and their accuracy.

their apparent month of death. We did not find any examples of this kind of post-death utilization, suggesting that the false positives – living people who are incorrectly coded as deceased – are not a major problem in our data. However, our match likely does include false negatives – people who are deceased but are not linked to any CDI death record. Data Appendix B provides more details on these two quality checks. Third, we compare measures of match quality between the February-March and March-April cohorts to assess the possibility that measurement error in mortality may differ across the two arms of the study (Black et al., 2017). Appendix Table 1 shows the number of patient records matched on SSN vs matched on probabilistic terms in each focal group (March-April vs February-March) in 2019 and 2020. Panel A shows that within the standalone EMR data, the share of patients for whom SSN is available is consistent across the two cohorts. It is also consistent when we compare 2019 to 2020 records. Similarly, Panel B indicates that the share of deaths with SSN vs without does not significantly vary across the cohorts or across years. Thus, we do not believe the accuracy of the link between the EMR and CDI data meaningfully differed for scheduled visits in February-March vs March-April.

3.4 Analytic Samples

To satisfy privacy requirements, we were only able to link the CDI mortality records to a restricted version of the Healthjump EMR database, which had a narrower set of fields compared with the full EMR. The linked file excludes the provider id, precise laboratory test information, and patient's birth year. In the linked data, we measure age in bins (0-17, 18-25, 26-35, 46-54, 55-64, and 65+) and cannot determine the specialty of the provider the patient was supposed to visit at the focal appointment. Thus, our analysis involves two analytic datasets: (i) the linked EMR-mortality data, and (ii) the EMR-only Healthjump database, which does not include mortality information but contains additional variables.

After linking the CDI and Healthjump records, we apply a series of inclusion criteria to form our analysis samples. At the outset, we retained all records of appointments scheduled to occur between 13 February 2020 and 13 April 2020, resulting in $3,279,173$ appointments in the February-March window and $2,718,869$ appointments in the March-April window. We de-duplicated by treating multiple appointments for the same patient on the same day as a single appointment, leaving $3,136,851$ February-March appointments and $2,545,067$ March-April appointments. We linked the appointments with patient information and excluded all appointments for patients with missing covariates (age, race, gender, state, or baseline Charlson score). The final linked EMR-mortality sample used in our main analysis includes 3,127,205 patient-appointments in the February-March cohort, and 2,538,436 patient appointments for the March-April cohort. We repeated these steps to construct placebo February-March and March-April cohorts from 2019.

Because the linked EMR-mortality-data had a less complete set of covariates, we also conduct some analysis on an EMR-only sample where we can make use of some of the information that was suppressed in the linked data. We construct the February-March and March-April cohorts in the EMR-only data using the same procedures described above. The sample size of the linked EMR-mortality sample and the EMR-only sample are slightly different because the two datasets were frozen at different times by our research team. We present evidence to verify that the first stage and the covariates in both datasets are very similar, suggesting that conclusions from our analysis of additional variables in the unlinked EMR-only data likely applies to the linked sample as well.

4 Results

4.1 Covariate Balance

The natural experiment in our study involves a comparison of downstream utilization and mortality in the February-March vs March-April scheduling cohorts, exploiting the fact that March-April appointments were cancelled at higher rates. One of the central assumptions in the design is that membership in the March-April cohort is independent of potential outcomes and baseline characteristics. Table 1 shows covariate balance in the February-March vs March-April groups in the linked mortality sample. The first column shows the composition of the patients in the February-March cohort, and the second column shows the mean difference between the March-April cohort and the February-March cohort. The third column shows the Cohen's D statistic, defined here as the mean difference divided by the standard deviation of the covariate in the February-March group. A rule of thumb in the matching literature suggests that Cohen's D statistics less than 0.15 or 0.10 are well balanced. The February-March and March-April groups are well balanced by this standard, with Cohen's D statistics ranging from 0.001 to 0.07 standard deviations.

The distribution of Charlson Comorbidity Scores – an index designed to predict mortality risk – is similar in the two groups. In both scheduling cohorts, about 80% of people had Charlson scores of 0, about 6% had a Charlson score of 1, and about 13% had Charlson scores of 2 or more. We used data from 2019 to fit regressions of one-year mortality on a flexible specification of demographic covariates and indicators for the diagnoses that enter the Charlson Index. We used the coefficients from these 2019 models to predict the one year mortality risk for each member of our 2020 sample. The final row of Table 1 shows that the average predicted mortality score is very similar in the two cohorts. Figure 1 shows kernel density plots of predicted mortality in the February-March vs March-April cohorts, illustrating that the distribution of ex ante mortality risk is very similar in the two groups. Appendix Table 2 shows that the covariates are also well balanced in the EMR-only sample. Taken together, these results support the independence assumption.

4.2 The Pandemic Cancellation Shock

We measure the first stage effect of the cancellation shock using a linear regression of the appointment cancellation variable on a dummy variable indicating that the appointment was scheduled for the March-April, estimating standard errors using a cluster robust variance matrix that allows for heteroskedasticity and for dependence at the person level because some patients contribute more than one appointment to the sample. The first column of Table 2 shows estimated coefficients from the first stage regression in the linked EMR-mortality sample. The intercept shows that the cancellation rate in the February-March control group was about 19%. The coefficient on the March-April indicator shows cancellation rates were 14.7 percentage points higher in the March-April group. Thus, the first stage has the expected sign and the effect is precisely estimated with a standard error of only 0.041 percentage points. The F-statistic on the instrument is $128,548$, much larger than standard rules of thumb used to judge concerns about weak instrument bias. The first stage is also substantively large. The cancellation rate was 19% in the February-March control group and 33.7% in the March-April treatment group, a 77% increase in relative terms. Viewed the other way around, in the control group about 81% of the scheduled visits resulted in actual health care utilization compared to only 66.3% of the scheduled visits in the treatment group. By that measure, the Covid-19 emergency led to an 18% contraction in the consumption of planned health care services.

The second column in Table 2 shows the cancellation first stage based on the EMR-only sample. In these data, the baseline cancellation rate in the February-March cohort is 18.4% and is 16.4 percentage points higher in the March-April cohort, showing that the first stage is very similar in the EMR-mortality and EMR-only samples. However, in the EMR-only sample, we are able to examine alternative types of healthcare use affected by the cancellation shock. Columns 3 (linear) and 4 (Poisson) show that the March-April group consumed .57 fewer lab tests on the scheduled appointment day than the February-March group, a 44%

reduction in lab tests. Columns 5 (linear) and 6 (Poisson) show impacts of the cancellation shock on the service charges associated with each scheduled appointment. For visits that occurred, those involving more intensive medical care tend to have higher charges. (Charges are set to zero for cancelled appointments.) The linear model shows that average charges were $57 lower in March-April group. From a base of $154 in the February-March group, the cancellation shock reduced charges by about 37%; the Poisson model implies the same proportional effect. These alternative metrics illustrate that the cancellation shock produced a genuine reduction in health care utilization. People missed their scheduled appointments, had fewer lab tests, and accumulated fewer charges.

Our study design assumes that there no persistent seasonal pattern in health care cancellation rates that differentiates the February-March and March-April cohorts. This assumption would be violated if appointments booked in advance for early Spring months are always cancelled at higher rates due to things like vacations or weather patterns. The first column of Panel B of Table 2 shows estimated first stage regressions from a placebo study population constructed using appointment cohorts from 2019. The cancellation rate was 19.7% in the 2019 February-March cohort, which is nearly identical to the cancellation rate in the 2020 February-March cohort. However, the coefficient on the March-April indicator in the 2019 sample is only $-.004$ (95% CI: $-.003$ $-.005$), a precisely estimated zero. These results suggest that our 2020 first stage is not driven by seasonal patterns in cancellation rates.

4.3 Effects of Cancelled Care on Mortality

Figure 2 shows accumulated mortality in the February-March and March-April scheduling cohorts in 2020 (Panel a) and 2019 (Panel b). The horizontal axis measures time in 30 day blocks (months) centered at the focal appointment date in each group. Period -5 occurs 5 months before the focal appointment, period 0 contains the focal appointment, and period 12 occurs one year after the focal appointment. The vertical axis shows the cumulative number of deaths per $10,000$ patient appointments at each follow up period. In essence, we follow the February-March and March-April scheduling cohorts over time and keep track of the fraction of the original population that has died at each follow-up period. As expected, the cumulative mortality rate is close to zero for both groups 5 months before the focal appointment.[8] In both years, the two scheduling cohorts accumulate deaths at the same rate as the focal appointment approaches, providing further evidence that the two groups

[8]Since many appointments are booked in advance, some people have already died by the time of the focal appointment so that the cumulative mortality rate is above zero by the date of the focal appointment.

experienced comparable mortality risk at baseline.

In the 2020 sample, the scheduling cohorts diverge after the focal appointment. The March-April cohort accumulates more deaths per $10,000$ patient-appointments in every follow up period than the February-March comparison group. One year after the focal appointment, there were about 87 deaths per $10,000$ in the March-April cohort compared to about 83 deaths per $10,000$ in the February-March group. The ITT comparison implies that having an appointment scheduled during the March-April window increased the one year mortality rate by $87 - 83 \approx 4$ deaths per $10,000$, a 5% increase in the baseline mortality rate. Panel b shows that the 2019 scheduling cohorts accumulated mortality at the same pace for the entire follow up period. If the two groups had accumulated mortality differently in 2019, this placebo test would have cast doubt on the exclusion restriction in our 2020 analysis, most likely indicating some form of seasonal mortality patterns.

Table 3 shows regression estimates of the effects of the cancellation shock on mortality at 6 months and 12 months follow up using membership in the March-April cohort as an instrumental variable for a cancelled appointment. The ITT estimates in Panel A show that the March-April group had about 3.4 (95% CI: 2.0-4.8) extra deaths per $10,000$ after 6 months and about 4.4 (95% CI: 2.6-6.2) extra deaths per $10,000$ after 12 months. The IV estimates imply that, among compliers, a cancelled appointment increased mortality risk by about 23 (95% CI: 14-32) per $10,000$ after 6 months and by about 30 (95% CI: 18-41) per $10,000$ after 12 months. Over a 12 month follow up period, the compliers experienced about 1 extra death for every $\frac{1}{30/10{,}000} \approx 333$ cancelled appointments. A 10% increase in maintained health care appointments leads to a $\frac{4.4/83}{-14.7/81} \times 10 \approx -2.9\%$ decline in mortality rates, implying that mortality-health care relationship is negative but inelastic.

Panel B shows corresponding results for the 2019 placebo sample. The placebo ITT effects are quantitatively small and not statistically different from zero at 6 or 12 months. Since the first stage effect is precisely estimated and nearly zero, these results support the exclusion restriction. As Figure 2 shows, overall mortality was higher in the 2020 sample than in the 2019 sample. The intercept from the 2019 ITT mortality models shows that the 2019 February-March cohort had accumulated about 70 deaths per $10,000$ in the year following the appointment date. In contrast, the 2020 February-March cohort had accumulated 83 deaths per $10,000$ after one year. The difference in the two intercepts suggests the pandemic generated about 13 extra deaths per $10,000$ without including the effects of cancelled care. In comparison, the ITT cancellation effect of 4.3 extra deaths per $10,000$ is about $1/3$ of the implied pandemic effect.

The exclusion restriction could also fail if the March-April group faced worse pandemic conditions than the February-March group. Then differential pandemic exposure rather than health care utilization might explain the mortality differences. Although it is possible that cancelled care could induce Covid-19 mortality if depleted health stocks make people more susceptible, if the *differential pandemic exposure* threat actually happened, we would plausibly expect the March-April mortality effect to disproportionately involve deaths caused by Covid-19. To test this implication, we measure *suspected* Covid mortality using an indicator that a person died *and* had a Covid-19 related diagnosis in the three month period before death. Table 4 shows ITT and IV estimates of the effect of cancellation on Covid and non-Covid mortality after a 12 month follow up. The ITT estimates imply that the cancellation shock led to 1.1 (95% CI: 0.5-1.7) additional suspected Covid-19 deaths per $10,000$ compared with about 3.3 (95% CI: 1.7-4.8) extra non-Covid deaths per $10,000$. Thus, the overall mortality effects are driven primarily by non-Covid deaths, casting doubt on the possibility that the March-April group experienced more severe pandemic conditions than the February-March group.

4.4 Mechanisms

The IV estimates imply that cancelled appointments increased mortality risk among compliers by almost 30 deaths per $10,000$. We estimate that in the absence of cancellations, the one year mortality rate among compliers would have been about 37 per $10,000$.[9] Thus, among compliers, cancellations increased mortality rates by $\frac{30}{37} \approx 81\%$. We explore two mechanisms that might explain this large effect: high marginal benefit compliers, and cascade of care effects.

Compliers Have High Marginal Benefits of Care

The effect of cancellations on mortality among compliers may be high because the compliers turn out to be a group with high marginal benefits of medical care. In essence, the compliers may have traded Covid-19 risk reduction for the risks of under-utilization of high value medical care. Table 5 shows estimates of baseline covariate means for the complier subpopulation compared with the February-March cohort.

In the EMR-only sample – which has more detailed age information – 6.2% of compliers are over 85 years old, compared with 4% of the February-March sample. In addition, about 13% of the February-March cohort had Charlson comorbidity scores of 2 or more compared

[9] The instrumental variable estimate of the untreated mortality rate among compliers in our data is: $\widehat{Pr}(m_{ij}^{12}(0)|c_{ij}(1) - c_{ij}(0) = 1) \times 10,000 \approx 37$.

with 24% of compliers. These estimates show that compliers were older and sicker than the overall study population. Figure 3 shows estimates of ITT effects in selected sub-populations; Figure 4 shows corresponding IV effects. The mortality effects were larger among older and sicker patients. The ITT estimates imply that the cancellation shock increased one-year mortality rates by 2.6 deaths per $10,000$ among people under 65 and by 10.7 deaths per $10,000$ among people over 65. The IV estimates – which account for differences in the first stage by age group – imply that a cancelled appointment increased mortality rates by 22.8 deaths per $10,000$ among people under 65, and by 48.5 deaths per 10,000 among people 65+. Figure 5 shows estimates in more detailed age groups. The first stage estimates are largest among age 0-17 (14 percentage points), then begin to decline and rise again among age 55-64 (13 percentage points) and age 65 plus (20 percentage points). The IV estimates imply that cancellations had the largest mortality effects among people over 65.

Stratifying on comorbidities rather than age gives similar results. The IV effects imply that a cancellation increased the mortality risk by 9.6 deaths per $10,000$ in the low comorbidity group compared with 50.1 per $10,000$ among people with at least 2 comorbidities. Panel b in figures 3 and 4 show ITT and IV estimates for people with specific conditions: Cancer, Heart Disease, Alzheimer/Dementia, Diabetes, and Chronic Pulmonary Disease. The IV cancellation mortality effects are larger for people who have a Cancer diagnosis or an Alzheimer/Dementia diagnosis. For example, the IV effects imply that a cancelled appointment increased mortality risk by about 293.3 deaths per $10,000$ for Cancer patients and by about 848.9 deaths per $10,000$ for Alzheimer/Dementia patients. We might expect these effects to be larger for racially minoritized populations, given inequities in early Covid-19 outcomes (Alsan et al., 2021). However, we found little evidence that mortality effects differed substantially by race or by gender; see figures 3 and 4.

Taken together, the results support the theory that the complier sub-population is a group with above average marginal benefits of medical care. It makes sense to ask what health services the compliers were going to consume. This would provide insight into what types of appointments are "life saving". Unfortunately, the appointment data does not include information on the specific procedures, tests, or prescriptions that might be planned for the visit. We only learn this kind of information for visits that actually happen. However, in the EMR-only sample, we do observe the specialty of the physician the patient was planning to visit. Appendix Table 3 shows estimates of the proportion of compliers that were planning to see each type of provider. Compliers were more likely to be seeing a specialist than the overall sample: 64% of the compliers were scheduled to see a specialist compared with 56% of the full February-March group. Specialist care may be more consequential for mortality than primary care over a one-year time horizon. Cardiology, Obstetrics and Gynecology, and Physical Therapy appointments were over represented in the complier population, possibly indicating that they may have higher value added than other visits.[10] A large share of the specialist gap in complier cancellations came from Ophthalmology appointments, which were 3.8% of the February-March sample but 12% of the complier group. However, it seems less plausible that the mortality consequences of cancelled eye care appointments are driving our mortality results.

Another possibility is that cancellation effects are concentrated in parts of the country where the pandemic was particularly severe in the early 2020 period. For example, it might be the case that cancelled care was more common in places with severe early outbreaks, or that the severity of the early epidemic reduced health care quality independent of the cancellations – a violation of the exclusion restriction. Figure 6 shows how cancellation rates varied in the March-April and February-March groups across states with different early epidemic severity. The horizontal axis records the cumulative number of confirmed Covid-19 cases per $100,000$ up to April 13 2020. Cancellations rates are shown for the February-March and March-April cohorts in each state. In the first month after the pandemic was declared, NY, NJ, LA, MA, MI, CT, PA, WA, and IL were Covid-19 "hot spots". But the cancellation shock was not concentrated in these hot spot states. March-April cancellations were nearly doubled compared to February-March in most states across the country. Table 6 shows that the reduced form ITT effects of the cancellation shock on 12 month mortality rates were very similar in states with more vs less severe initial outbreaks. The cancellation shock increased mortality by about 4.8 deaths per $10,000$ in low initial severity states, and by 4.4 deaths per $10,000$ in high initial severity states. This suggests that the mortality effects of the cancellation shock are not driven by geographic variation or local health care systems that were overwhelmed by the pandemic.

Cascade of Cancelled Care

A second explanation is that many outpatient appointments are part of a chain of care that together generate improved health outcomes. For example, an initial appointment may generate lab tests, which support a diagnosis leading to a prescription or surgical procedure. Cancelling one appointment could disrupt the chain of events, producing a larger health effect than any appointment in isolation (Kaestner and Sasso, 2015). In this sense, our

[10] Our data includes telehealth visits and we do not distinguish between telehealth and face to face visits because distinguishing telehealth from face to face visits accurately can be challenging. In the early pandemic months, CMS urged providers to carry on with telehealth visits without additional documentation and offered a pay parity for both types of visits (CMS et al., 2020)

cancellation first stage understates the full reduction in health care utilization that patients experienced. To shed light on these cascades, we estimate the effects of cancellations on health care visits, labs, and charges consumed as of each month of the follow-up year.

Figure 7 shows the cumulative number of visits (not counting the index visit) attended by the scheduling cohorts following the index visit. In 2020, the March-April group accumulates fewer health care appointments during the first three months after the index visit, showing that the initial cancellation does induce reduce health care utilization for a longer period of time. Figures 8 and 9 show a similar pattern for the accumulation of laboratory tests and total charges during the follow-up year. We find with no corresponding patterns in the placebo sample from 2019. Table 7 shows point estimates of the ITT and IV effects of the cancellation shock on subsequent utilization measures after 1 month and after 6 months. The ITT effects show that one month after the focal visit, appointments were down 17%, laboratory tests were down 35%, and charges were down 32% in the March-April group. The initial slow down in utilization recovers and reverses course by 6 months, which could indicate that the initial shortfall has started to reduce health or simply that the group is catching up on postponed care.

The results suggest that the initial cancellations tipped off a cascade of low health care utilization that lasted about 1-3 months. As discussed earlier, the measures of utilization we examine may not fully capture each person's health care consumption. We do not measure any care delivered by non-participating providers or in hospital inpatient and emergency departments. With this caveat, one view is that the total mortality effects of the cancellation shock reflect a larger reduction in health care utilization than indicated by our cancellation first stage. If that is true then the denominator in our IV estimates is too small and the the mortality effect of medical care "per unit of medical care" is smaller.

4.5 Sensitivity Analysis and Robustness Checks

Special Seasonality and Calendar Time

Our research design follows the appointment scheduling cohorts over the 12 months following the index appointment. As a result, the follow-up periods are arranged in *event time* rather than *calendar time*. In practice, event time and calendar time are nearly aligned: the 12 month follow-up occurs in March 2021 for the March-April group, and in February 2021 for the February-March group. When we repeat the basic analysis using data from 2019, we found no evidence that pre-existing seasonal patterns explain the 2020 results. However, the 2019 comparison may not be a valid test if seasonal mortality changed in important ways during the Covid-19 pandemic. An alternative approach is to compare cumulative mortality in both groups in the same calendar months during the follow-up year. The problem with this approach is that the control group will always have a longer exposure window than the treatment group. Even if the two groups have identical mortality rates, the group that is observed for a longer follow up will accumulate more deaths.

To avoid this issue, we compare "pro-rated" mortality rates in March-April and February-March at each calendar month. Let N_{MA} and N_{FM} be the number of patient-appointments in the March-April and February-March cohorts at baseline. Then the total number of deaths in the March-April cohort as of March 2021 is $M_{MA}^{March} = \sum_{i=1}^{N_{MA}} m_{ij}^{12}$, since March 2021 is 12 follow up periods after the focal appointment in the design. Adjusting for the size of the group at risk gives mortality rate per 10,000 patient years as of March 2021: $\tilde{M_{MA}^{March}} = \frac{M_{MA}^{March}}{N_{MA}} \times 10,000$. In comparison, the total number of deaths in the February-March cohort as of March 2021 $M_{FM}^{March} = \sum_{i=1}^{N_{FM}} m_{ij}^{13}$, which is the number of deaths over a 13 months exposure period. The pro-rated one-year mortality rate is: $\tilde{M_{MA}^{March}} = \frac{\frac{13}{12} \times M_{MA}^{March}}{N_{MA}} \times 10,000$.[11]

Estimates of ITT and IV effects from these pro-rated calendar time aligned comparisons are in Table 8. The results imply that by March 2021 the March-April cohort had accumulated about 6.0 (95% CI: 4.2-7.8) additional deaths per 10,000 in comparison with the February-March cohort in March 2021. Likewise, the IV estimate implies that a cancellation generated 12.9 (95% CI: 1.3-24.5) extra deaths per 10,000 by March, 2021. These results are qualitatively similar to our main analysis, and suggest that special seasonality is not an important source of bias in our work.

Covariate Adjustment and Violations of Independence Assumption

Our analysis relies on the independence assumption, which implies that the March-April and February-March cohorts do not systematically differ on baseline characteristics that are associated with downstream mortality risk. We use data on a set of *observed* baseline covariates to assess this assumption using balancing tests. These tests suggest that the groups are well-balanced in both years. In addition, the placebo analysis of cumulative mortality patterns in the 2019 sample suggests that people with higher mortality risk do not systematically schedule appointments in the March-April window.

Nevertheless, the differences in baseline covariates in the 2020 March-April and February-March cohorts is not precisely zero. It is possible that small differences in covariates could contribute to differences in mortality. Although it deviates somewhat from our study design,

[11] Similar exposure weight procedures apply to comparisons in other calendar months.

a natural approach is to adjust for observable covariates using a matching or regression adjustment procedure. We kept this approach simple to avoid concerns about p-hacking. Thus Table 9 shows estimates of ITT and IV effects after 6 months and 12 months that adjust for baseline covariates using linear regression models that control for the characteristics we included in the balance table. The qualitative results from our main analysis are not sensitive to adjusting for covariates: the first stage regressions continue to show that the March-April cohort experienced higher cancellation rates than the February-March cohort. Likewise, the reduced form results continue to show that the March-April group experienced higher mortality rates after 6 months and 12 months of follow up time. The magnitude of ITT point estimates is smaller after adjusting for covariates. The simple ITT (without covariates) implies that the cancellation shock increased mortality rates by about 4.4 (95% CI: 2.6-6.2) extra deaths per $10,000$ after 12 months. In comparison, the covariate adjusted ITT analysis implies that the cancellation shock increased mortality rates by 1.9 (95% CI: 0.14-3.6) deaths per $10,000$ after 12 months. The confidence intervals from the two models overlap; however, the point estimates from the covariate adjusted models are smaller and might offer a more conservative estimate of the effect of the cancellation shock on downstream mortality.

Pre-booked Sample

The intuitive appeal of our study design partly flows from the idea that people book outpatient appointments weeks or even months in advance. This means that most of the appointments in our design were made before the pandemic had started to affect people's plans. But how realistic is this account of outpatient health care? Surely some appointments are made on very short notice. And – at some point – new bookings must have started to fall off in response to the pandemic. For about 50% of the appointments in our sample, we are able to observe not only the appointment date but also the *date when the appointment was booked*.[12] To measure pre-booking patterns, we limited our main sample to the subset of appointments with an observed booking date. This yields $1,491,303$ appointments from February-March and $1,290,452$ from March-April. Then we compute the difference between the appointment date and the booking date, which measures how many days in advance the appointment was booked. Figure 10 shows the cumulative distribution of advance booking times in the two cohorts. In both cohorts, more than 50% of appointments were scheduled at least 30 days

[12] Specifically, booking dates are available for all appointments booked after the date when a provider first participates in the Healthjump system. Essentially, once Healthjump integrates with a provider EMR system, it pulls records retrospectively and continues to pull EMR data on a daily basis. EMR encounters prior to the integration date – those pulled retrospectively – are assigned the integration date itself as the booking date.

in advance, supporting the idea that the bulk of outpatient appointments in our study were booked before pandemic concerns were salient.

In addition, we replicated our mortality analysis in a special sub-sample of appointments that are known to have been booked prior to the pandemic emergency declaration.[13] Appendix Table 4 shows that the characteristics of the pre-booked sample and covariate balance are nearly identical to those in our main analysis (see Table 1). Table 10 presents first stage, ITT, and IV estimates of the effects of cancelled care on mortality at six months and one year in the pre-booked sample. The first stage remains strong in the pre-booked sample and is actually somewhat larger than in the main sample. The ITT and IV estimates in the pre-booked sample are very similar to our main estimates. The ITT estimate in the pre-booked sample implies that the cancellation shock increased one year mortality rates by 4 deaths per $10,000$ appointments compared with 4.4 deaths per $10,000$ in the full sample. Likewise, the IV analysis implies that a cancellation increased mortality rates by 22 deaths per $10,000$ among compliers in the pre-booked sample compared with 30 deaths per $10,000$ in the main sample. The results of this sensitivity analysis suggests that there is little reason to be concerned with strong selection on last minute appointment scheduling in our main analysis.

Summary of Sensitivity Analysis Results

The results presented indicate that missed outpatient care had a significant effect on downstream mortality. However, a key assumption underlying our results is that those with visits scheduled for March-April vs February-March are comparable. We presented evidence that the two samples were balanced in terms of baseline covariates and then examined the sensitivity of our conclusions to changing key features of our analysis. Figure 11 describes the ITT point estimates from all analyses graphically. Our main ITT estimates are in the first row of the figure. The second row adds adds a vector of controls for state fixed effects, demographic variables, and baseline health conditions. In the third row, we show ITT estimates based on the sub-sample where we are sure that visits were pre-booked in advance. The fourth row shows ITT comparisons computed at a common terminal month (March 2021), pro-rated to account for differential follow up time. The fifth row gives ITT estimates of the effects

[13] We constructed this pre-booked sample in several steps to accommodate the sample size and maintain the study design. First, we expanded the definition of the March-April group to cover the period from 13 March 2020 to 30 April 2020, which is a slightly longer period than our main analysis and allows for a larger sample size. Second, we limited our sample to appointments with an observed booking date. Third, we limited the March-April appointments to those with booking dates before 13 March 2020, which is the pandemic emergency. Similarly we limited February-March appointments to those booked before 13 February 2020. These steps lead to a pre-booked sample size of 1.6 million observations.

of the shock on suspected non-Covid mortality. Finally, in the sixth and seventh rows, we show ITT effects in states with a high vs low early Covid-19 burden. The collection of ITT point estimates is quite consistent across these sensitivity analysis. The confidence intervals exclude zero in each case, but the confidence intervals overlaps across estimates. In total, the sensitivity results shown here indicate that the assumptions underlying our main specification appear supportive of a causal interpretation in the relationship between cancelled care and mortality.

5 Discussion

Taking advantage of a natural experiment produced by sudden cancellations at the start of the pandemic, our empirical work makes two substantive contributions to the economics literature. First, we show that non-acute outpatient health services have a substantial causal effect on health. Cutting back on these kinds of "everyday" health services increased mortality rates in statistically detectable and economically meaningful ways, pushing against the idea that outpatient care is often flat of the curve medicine. Second, our study quantifies an important unintended consequence of the early pandemic shutdown, showing that efforts to reduce mobility and infection risk led to disruptions in health care and loss of life that weigh against benefits of pandemic policy.

Our analysis is built on an innovation in data linkage. We linked mortality information from obituary data with electronic medical records at the person level. The new mortality data we use – the Connected Death Index – improves on traditional mortality sources, which are often released with a significant lag and/or severely under count deaths. The mortality record linkage worked well in our study, and it is potentially useful for other research databases such as on employment, education, housing, pensions, and clinical trials. Very recently, these data have been linked to voter registration records to examine excess mortality by political party affiliation (Wallace et al., 2022).

In the natural experiment we examined, a 10% increase in outpatient utilization achieved through maintained appointments reduced the one year risk of mortality by 2.9%. The literature on *flat of the curve medicine* shows the benefits of a narrow set of urgent and acute care services (Doyle, 2007; Almond et al., 2010; Doyle, 2011), and that geographical variation in health care spending/utilization does not generate improved health outcomes (Fisher et al., 2003; Baicker and Chandra, 2004; Wennberg et al., 2008). Our work finds that health care has an important effect on mortality even in non-urgent and non-acute situations. A natural question is how our estimates of the impact of cancelled care on patient outcomes compare to the treatment effect estimates published in the literature. The most relevant comparisons involve studies on the health effects of outpatient care, the health effects of health insurance coverage, and the health effects of natural disasters and other system level disruptions.

There are not many studies of the causal health effects of outpatient care itself. One recent example is Bradley et al. (2017), which reports results from an RCT in which low-income uninsured people in Virginia were randomly assigned the offer of cash payments if the patient completed at least one primary care appointment. Using the incentive offers as instruments for PCP visits, Bradley et al. (2017) found that PCP visits reduce emergency room visits but did not affect overall health spending. In essence, the study creates exogenous substitution to PCP visits, a form of outpatient care. The cancellation instrument we examined in the present paper induced exogenous changes in health care consumption in a different way. The cancellation shock reduced consumption of primary care visits as well as other types of outpatient care. In our study window, it was not easy to substitute cancelled care for other types of care, thus the health effects we are able to observe were likely caused by total reductions in care rather than substitution away from emergency departments or other forms of care.

The literature on the health effects of insurance coverage is larger. Two recent studies seem most directly connected to ours. First, Goldin et al. (2021) study a large scale RCT in which the IRS sent informational letters to people who had paid a tax penalty because they lacked health insurance in the previous year. The letter informed people that they had been penalized in 2015 and provided information about how they could obtain low cost insurance from the Exchange or Medicaid for 2017. People were randomly selected to receive a letter or no letter. Goldin et al. (2021) use tax data to measure health insurance coverage and Social Security data to measure mortality. They found that – among people who had no health insurance coverage in the baseline year – the informational letter increased insurance coverage by about 0.23 months in the full sample and by about 0.36 months among adults ages 45 to 64, leading to reduced mortality. The ITT effect implies that receiving a letter reduced the one-year mortality rate from 1.01% to 0.94%, which is about 1 fewer death for every $1,587$ people who were sent a letter. Goldin et al. (2021) interpret these reduced form results as health insurance reducing middle aged mortality rates through three possible channels: increasing timely medical treatment for acute conditions (heart attack or stroke), expanding access to non-acute care which might lead to earlier diagnosis and treatment, and by reducing stress through financial security and peace of mind. However, they cannot investigate these possibilities empirically because they do not have access to health care data.

Unlike Goldin et al. (2021) the variation we examine comes from a direct contraction in the availability of health services rather than an indirect effect from insurance coverage, and our study does measure health care utilization as well as mortality.

A second recent insurance coverage study that is related to our work is Miller et al. (2021), which examines the mortality effects of expanding Medicaid to uninsured low-income adults who were ages 55-64 in 2014. Their analysis implies that one year of Medicaid coverage reduced mortality rates among compliers by about 0.35 percentage points. To interpret the magnitude of the effect, Miller et al. (2021) suggest that an upper bound estimate of the untreated (no Medicaid) mortality rate among compliers was about 2.95%. This suggests that Medicaid enrollment reduced mortality among compliers by *at least* 11.9% relative to the counterfactual. When they treat the mortality rate in the full sample as a lower bound on untreated complier mortality, Miller et al. (2021) suggest that Medicaid enrollment could have reduced complier mortality rates by lower bound estimate of complier mortality by around 21.5%. Like the study by Goldin et al. (2021), the Medicaid expansions studied by Miller et al. (2021) may arise through multiple channels, including timely treatment of acute conditions, better access to non-acute care, and the financial benefits of insurance. However, Miller et al. (2021) present some evidence based on cause of death information. Health care amenable causes of death fell more than other causes of death, suggesting that medical care played a role. Because our study does not rest on variation in insurance coverage, the estimates are more directly linked with health care consumption. In addition, the mortality and EMR data we use in the paper allows a more through investigation into possible mechanisms. We show fairly clearly that the complier group in our sample consisted of people with characteristics that suggest they may benefit substantially from medical care. And – indeed – the treatment effects among the compliers in our sample are large in percentage terms. However, we also show that some of the effect was likely driven by a cascade of cancellations and disruptions in care, rather than by a single appointment.

Our estimates can also be compared to estimates from studies of other sudden disruptions in healthcare delivery. Deryugina et al. (2020) examined Medicare hospital elective visits and charges from 1997-2013 in counties impacted by hurricanes compared to nearby unaffected counties. On average a hurricane reduces elective surgeries by 7% in the month it makes landfall. The cancelled and delayed elective surgeries did not generate substantial changes in mortality effect, perhaps suggesting that a significant share of these surgeries were over-utilized at baseline. Gruber and Kleiner (2012) examined the effects of nurses' strikes in hospitals on patient outcomes in New York State. They found hospital mortality rose by 18.3 percent and 30-day readmission increased by 5.7 percent for patients admitted during a strike. Like our study, both Deryugina and Molitor (2018) and Gruber and Kleiner (2012) examine disruptions to health care that are considered elective or deferrable. Finally, our work relates to Qian (2022) who examined infant outcomes for newborns delivered during the early pandemic shutdown period and found they were less likely to receive care and more likely to have worse outcomes (birthweight and hospital readmission; their data do not allow them to examine mortality).

Recent work by Ruhm (2021) and Cronin and Evans (2021) estimates excess deaths during the pandemic and suggests that non-Covid deaths account for 10% to 28% of all excess deaths in 2020.[14]. Ruhm (2021) estimates that there were about $97,000$ excess non-Covid deaths in 2020. Two back of the envelope calculations suggest that the cancelled and disrupted care we study in this paper account for 1/3 to 1/4 of these non-Covid excess deaths.

The first exercise combines point estimates from our study with information about the number of outpatient visits that were likely affected. Specifically, the ITT estimates from our study imply that the cancellation shock increased (suspected) non-Covid mortality rates by about 3.3 deaths per 10,000 appointments. Across the country, there are approximately $71,698,833$ ambulatory care visits per month (Santo and Okeyode, 2021).This suggests that the cancellation shock would have generated about $3.3 \times \frac{71{,}698{,}833}{10{,}000} \approx 23,660$ additional deaths over the following year. Viewed this way, the cancellation shock accounts for about $\frac{23{,}660}{97{,}000} \times 100 \approx 24\%$ of all non-Covid excess deaths. The remaining non-Covid excess deaths might be attributed to other channels such as suicides, drug overdose, and alcohol misuse (Case and Deaton, 2021), pandemic isolation, crime and pandemic recession effects (Mulligan and Arnott, 2022; Ruhm, 2021).

The second estimate comes directly from our linked mortality-EMR data. Comparing the February-March cohorts from 2020 and 2019 gives an rough estimate of the excess mortality in 2020 that is not driven by the cancellation shock. The 2019 February-March cohort experienced 70 deaths per $10,000$ after one year compared with 83 deaths per $10,000$ in the 2020 February-March cohort. A simple interpretation is that the pandemic generated about 13 extra deaths per $10,000$, without including the effects of cancelled care that applied to the March-April cohort. The overall ITT cancellation effect in our study implies that the cancellation shock increased mortality rates by 4.3 extra deaths per $10,000$. This implies that the cancellation shock represents about $\frac{4.3}{13} \approx 33\%$ of the implied pandemic effect. These two

[14]Ruhm (2021) estimates that Covid-19 deaths account for 71.6% - 87.7% of all excess deaths for the period of March 2020 to February 2021. Cronin and Evans (2021) estimates that Covid-19 deaths account for 77.5% - 90.2% of all excess deaths in 2020

rough calculations imply the cancellation shock accounts for 1/3 to 1/4 of all non-covid19 excess deaths.

To assess the economic significance of the mortality consequences of the cancellation shock, we used a Years of Potential Life Lost (YPLL) framework – which assumes that deaths before age 75 are premature – to construct estimates of the VSL-costs per 10,000 visits; See Data Appendix E for details. Our calculations imply that the cancellation shock led to 443.5 lost years of life for every 10,000 cancelled appointments. A typical estimate from the economics literature puts the Value of Statistical Life Year (VSLY) at around $\$150,000$ (Cutler and Zeckhauser, 2000; Aldy and Viscusi, 2008). This implies that the early pandemic shutdown caused mortality losses pf about \$66.5 million per $10,000$ appointment cancellations. Given these survival implications, the break even price of an outpatient appointment is approximately $\$6,653$. Using our most conservative IV estimate instead suggests that the average outpatient appointment would be worth up to $\$3,000$. In our EMR data, the average of total charges for outpatient services over the three month period following the focal appointment was approximately \$300 in the 2019 February-March cohort. By this measure, the mortality reduction benefits of outpatient care are easily worth the costs.

The results of our study suggest that non-acute health care services provided in outpatient settings produce valuable improvements in health, at least as measured by mortality. That is, broad reductions in outpatient health care consumption would have substantial mortality consequences. These findings do not necessarily imply that flat of the curve medicine is uncommon in the outpatient sector, but they suggest that efforts to slow the growth of healthcare spending by reducing outpatient care should be designed to shield high-value services. This may be harder than it sounds in practice, considering that the cancellation shock we examine originated in efforts to cancel *non-essential* care during the pandemic. Future research that identifies outpatient services *and* patients that produce negligible health benefits would be valuable both for cost-containment purposes and for efforts to cope with future disruptions.

References

Aldy, J. E. and W. K. Viscusi (2008). Adjusting the value of a statistical life for age and cohort effects. *The Review of Economics and Statistics 90*(3), 573–581.

Almond, D., J. J. Doyle, A. E. Kowalski, and H. Williams (2010). Estimating marginal returns to medical care: Evidence from at-risk newborns. *The quarterly journal of economics 125*(2), 591–634.

Alsan, M., A. Chandra, and K. Simon (2021). The great unequalizer: Initial health effects of covid-19 in the united states. *Journal of Economic Perspectives 35*(3), 25–46.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association 91*(434), 444–455.

Baicker, K. and A. Chandra (2004). Medicare spending, the physician workforce, and beneficiaries' quality of care: Areas with a high concentration of specialists also show higher spending and less use of high-quality, effective care. *Health Affairs 23*(Suppl1), W4–184.

Baicker, K., A. Finkelstein, J. Song, and S. Taubman (2014). The impact of medicaid on labor market activity and program participation: evidence from the oregon health insurance experiment. *American Economic Review 104*(5), 322–28.

Barcellos, S. H. and M. Jacobson (2015). The effects of medicare on medical expenditure risk and financial strain. *American Economic Journal: Economic Policy 7*(4), 41–70.

Black, D. A., Y.-C. Hsu, S. G. Sanders, L. S. Schofield, and L. J. Taylor (2017). The methuselah effect: The pernicious impact of unreported deaths on old-age mortality estimates. *Demography 54*(6), 2001–2024.

Bradley, C. J., D. Neumark, and L. S. Walker (2017). The effect of primary care visits on health care utilization: Findings from a randomized controlled trial. Technical report, National Bureau of Economic Research.

Card, D., C. Dobkin, and N. Maestas (2009). Does medicare save lives? *The quarterly journal of economics 124*(2), 597–636.

Case, A. and A. Deaton (2021). *Deaths of Despair and the Future of Capitalism*. Princeton University Press.

CMS et al. (2020). Medicare telemedicine health care provider fact sheet.

Cronin, C. J. and W. N. Evans (2021). Excess mortality from covid and non-covid causes in minority populations. *Proceedings of the National Academy of Sciences 118*(39), e2101386118.

Cutler, D. M. and M. McClellan (2001). Is technological change in medicine worth it? *Health affairs 20*(5), 11–29.

Cutler, D. M. and R. J. Zeckhauser (2000). The anatomy of health insurance. In *Handbook of health economics*, Volume 1, pp. 563–643. Elsevier.

Datavant (2020a, 5). Deep dive on token selection. https://covid19researchdatabase.org/wp-content/uploads/2020/05/Token-Selection-Deep-Dive-updated-May-2020-1.pdf.

Datavant (2020b). Mortality data in healthcare analytics. https://datavant.com/resources/whitepapers/mortality-data-in-healthcare-analytics/.

Deryugina, T., J. Gruber, and A. Sabety (2020). Natural disasters and elective medical services: How big is the bounce-back? Technical report, National Bureau of Economic Research.

Deryugina, T. and D. Molitor (2018). Does when you die depend on where you live? evidence from hurricane katrina. Technical report, National Bureau of Economic Research.

Doyle, J. J. (2007). Returns to local-area health care spending: using health shocks to patients far from home.

Doyle, J. J. (2011). Returns to local-area health care spending: evidence from health shocks to patients far from home. *American Economic Journal: Applied Economics 3*(3), 221–43.

Doyle, J. J., J. A. Graves, J. Gruber, and S. A. Kleiner (2015). Measuring returns to hospital care: Evidence from ambulance referral patterns. *Journal of Political Economy 123*(1), 170–214.

Du, M., J. Hou, B. Hsu, Y. Tang, Y. Wang, A. Weishaus, and L. Xu (2020). Incorporating datavant's death index in mortality analysis.

Einav, L., A. Finkelstein, T. Oostrom, A. Ostriker, and H. Williams (2020). Screening and selection: The case of mammograms. *American Economic Review 110*(12), 3836–70.

Finkelstein, A., S. Taubman, B. Wright, M. Bernstein, J. Gruber, J. P. Newhouse, H. Allen, K. Baicker, and O. H. S. Group (2012). The oregon health insurance experiment: evidence from the first year. *The Quarterly journal of economics 127*(3), 1057–1106.

Fisher, E. S., D. E. Wennberg, T. A. Stukel, D. J. Gottlieb, F. L. Lucas, and E. L. Pinder (2003). The implications of regional variations in medicare spending. part 1: the content, quality, and accessibility of care. *Annals of internal medicine 138*(4), 273–287.

for Disease Control (CDC, C. et al. (1986). Premature mortality in the united states: public health issues in the use of years of potential life lost. *MMWR supplements 35*(2), 1S–11S.

Gardner, J. W. and J. S. Sanborn (1990). Years of potential life lost (ypll)—what does it measure? *Epidemiology*, 322–329.

Garthwaite, C., T. Gross, and M. J. Notowidigdo (2014). Public health insurance, labor supply, and employment lock. *The Quarterly Journal of Economics 129*(2), 653–696.

Goldin, J., I. Z. Lurie, and J. McCubbin (2021). Health insurance and mortality: Experimental evidence from taxpayer outreach. *The Quarterly Journal of Economics 136*(1), 1–49.

Gross, T. and M. J. Notowidigdo (2011). Health insurance and the consumer bankruptcy decision: Evidence from expansions of medicaid. *Journal of public Economics 95*(7-8), 767–778.

Gruber, J. and S. A. Kleiner (2012). Do strikes kill? evidence from new york state. *American Economic Journal: Economic Policy 4*(1), 127–57.

Gupta, S., K. I. Simon, and C. Wing (2020). Mandated and voluntary social distancing during the covid-19 epidemic: A review.

Kaestner, R. and A. T. L. Sasso (2015). Does seeing the doctor more often keep you out of the hospital? *Journal of health economics 39*, 259–272.

Lakdawalla, D. N., E. C. Sun, A. B. Jena, C. M. Reyes, D. P. Goldman, and T. J. Philipson (2010). An economic evaluation of the war on cancer. *Journal of health economics 29*(3), 333–346.

Levin, M. A., H.-M. Lin, G. Prabhakar, P. J. McCormick, and N. N. Egorova (2019). Alive or dead: validity of the social security administration death master file after 2011. *Health services research 54*(1), 24–33.

Miller, S., N. Johnson, and L. R. Wherry (2021). Medicaid and mortality: new evidence from linked survey and administrative data. *The Quarterly Journal of Economics 136*(3), 1783–1829.

Mulligan, C. B. and R. D. Arnott (2022). Non-covid excess deaths, 2020-21: Collateral damage of policy choices? Technical report, National Bureau of Economic Research.

Murphy, K. M. and R. H. Topel (2006). The value of health and longevity. *Journal of political Economy 114*(5), 871–904.

Newhouse, J. P. et al. (1993). *Free for all?: lessons from the RAND health insurance experiment*, Volume 172. Harvard University Press Cambridge, MA.

Nordhaus, W. (2018). Projections and uncertainties about climate change in an era of minimal climate policies. *American economic journal: economic policy 10*(3), 333–60.

Qian, X. (2022). *Essays on Impacts of Childhood Experiences*. Ph. D. thesis, The Ohio State University.

Robinson, L. A., R. Sullivan, and J. F. Shogren (2021). Do the benefits of covid-19 policies exceed the costs? exploring uncertainties in the age–vsl relationship. *Risk Analysis 41*(5), 761–770.

Rose, L. (2020). The effects of skilled nursing facility care: regression discontinuity evidence from medicare. *American Journal of Health Economics 6*(1), 39–71.

Rothwell, C. (accessed 9/11/2022). Ssa death master file nchs national death index: How do they relate? *Social Security Administration Presentation https://www.cdc.gov/nchs/data/bsc/rothwell.pdf*.

Ruhm, C. J. (2021). Excess deaths in the united states during the first year of covid-19. Technical report, National Bureau of Economic Research.

Sacks, D. W., N. Menachemi, P. Embi, and C. Wing (2020). What can we learn about sars-cov-2 prevalence from testing and hospital data? *arXiv preprint arXiv:2008.00298*.

Santo, L. and T. Okeyode (2021). National ambulatory medical care survey: 2018 national summary tables. *National Center for Health Statistics*.

Simon, K., A. Soni, and J. Cawley (2017). The impact of health insurance on preventive care and health behaviors: evidence from the first two years of the aca medicaid expansions. *Journal of Policy Analysis and Management 36*(2), 390–417.

Wallace, J., P. Goldsmith-Pinkham, and J. Schwartz (2022). Excess death rates for republicans and democrats during the covid-19 pandemic. *arXiv preprint arXiv:2209.10751*.

Welch, H. G. and W. C. Black (2010). Overdiagnosis in cancer. *Journal of the National Cancer Institute 102*(9), 605–613.

Wennberg, J. E., E. S. Fisher, D. C. Goodman, J. S. Skinner, and K. K. Bronner (2008). Tracking the care of patients with severe chronic illness-the dartmouth atlas of health care 2008.

Ziedan, E., K. I. Simon, and C. Wing (2020). Effects of state covid-19 closure policy on non-covid-19 health care utilization. Technical report, National Bureau of Economic Research.

6 Tables and Figures

Table 1: Covariate Balance in the Linked Mortality Sample

	February-March Mean	Difference	Cohen's D
Age 0-17	0.10	-0.02	0.071
Age 18-25	0.06	0.00	0.005
Age 26-35	0.08	0.01	-0.02
Age 36-45	0.10	0.01	-0.02
Age 46-54	0.12	0.01	-0.02
Age 55-64	0.18	0.01	-0.02
Age 65+	0.36	-0.01	0.01
Black	0.09	0.01	-0.02
White	0.49	0.00	-0.001
Female	0.59	0.00	-0.001
Charlson Score = 0	0.80	-0.01	0.03
Charlson Score = 1	0.06	0.00	-0.01
Charlson Score $\geq$ 2	0.13	0.01	-0.03
Predicted Mortality	0.007	0.0003	-0.02
N	5,665,641		

Notes - The unit of observation is the patient-appointment. Each row is estimated from a regression of baseline characteristics (7 age brackets, race, gender or baseline morbidity) on an indicator for the record being in the March-April cohort. The coefficient from the constant term is the February-March mean, the coefficient from the March-April indicator is the difference between the two cohorts. Given the large sample size we report the Cohen's d statistic to measure the standardised difference between two means instead of P-values.

Table 2: Effects of the Pandemic Disruption on Appointment Cancellations and Charges and Labs on the Visit Day

	Linked EMR Sample	Unlinked EMR Sample				
Panel A: 2020	*Missed Visit (Yes/No)*	*Missed Visit (Yes/No)*	*Labs (Count)*	*Labs (Poisson Model)*	*Total Charges*	*Total Charges (Poisson Model)*
March-April	0.147	0.164	-0.566	-0.574	-57.417	-0.465
	(0.000)	(0.000)	(0.006)	(0.001)	(0.804)	(0.000)
February-March	0.190	0.184	1.296	0.259	154.444	5.04
	(0.0003)	(0.000)	(0.004)	(0.001)	(0.541)	(0.000)
Number of Observations	5,665,641	5,625,882	5,625,882	5.625,882	5,625,882	5,625,882
	Linked EMR Sample	**Unlinked EMR Sample**				
Panel B: 2019	*Missed Visit (Yes/No)*	*Missed Visit (Yes/No)*	*Labs (Count)*	*Labs (Poisson Model)*	*Total Charges*	*Total Charges (Poisson Model)*
March-April	-0.004	-0.004	-0.006	-0.007	1.624	0.01
	(0.0004)	(0.000)	(0.005)	(0.001)	(0.805)	(0.000)
February-March	0.197	0.160	0.858	-0.153	156.565	5.053
	(0.0003)	(0.000)	(0.004)	(0.001)	(0.582)	(0.000)
Number of Observations	5,382,158	5,212,275	5,212,275	5,212,275	5,212,275	5,212,275

Notes - The unit of observation is the patient-appointment. Panel A reports first stage estimates from two samples. The linked EMR to mortality sample (column 1) and the unlinked EMR sample (columns 2 to 6). Column 1 in panel A reports the share of missed visits in the March-April cohort relative to the February-March cohort. Columns 2 to 6 report the share of missed visits, lab orders on the day of the visit, and total charges on the day of the visit for the two cohorts. Panel B, repeats this analysis for the year 2019. All regressions include no controls and standard errors are clustered at the patient level to allow for non-independence between patient records

Table 3: Effects of Cancelled Care on all Cause Mortality

		Six Month Mortality		**One Year Mortality**	
Panel A: 2020	First Stage	ITT	IV	ITT	IV
March Cohort	0.147	0.00034	–	0.00044	–
	(0.00041)	(0.00007)	–	(0.00009)	–
Cancelled Visit	–	–	0.0023	–	0.00297
	–	–	(0.00045)	–	(0.00059)
Intercept	0.190	0.00425	0.00382	0.0083	0.00775
	(0.00026)	(0.00006)	(0.00012)	(0.00016)	(0.00016)
Number of Observations	5,665,641	5,625,882	5,625,882	5.625,882	5,625,882
		Six Month Mortality		**One Year Mortality**	
Panel B: 2019	First Stage	ITT	IV	ITT	IV
March Cohort	-0.004	-0.00004		0.00004	
	(0.00036)	(0.00006)	–	(0.00008)	
Cancelled Visit	–	–	0.00942	–	-0.00916
	–	–	(0.01498)	–	(0.01935)
Intercept	0.197	0.00391	0.00205	0.00701	0.00882
	(0.00029)	(0.00007)	(0.00292)	(0.00009)	(0.00378)
Number of Observations	5,382,158	5,382,158	5,382,158	5,382,158	5,382,158

Notes - The unit of observation is the patient-appointment. Panel A reports the 2020 estimates and there only appointments scheduled for February-March and March-April 2020 enter the regression. Panel B reports the 2019 estimates and there only appointments scheduled for February-March and March-April 2019 enter the regressions. In both panels, column 1 reports the first stage estimate from the linked deaths EMR data. columns 2 and 3 report the ITT and IV estimates at 6 months after the visit date. Columns 4 and 5 report the ITT and IV estimates at 12 months after the visit date. All regressions include no controls. Standard errors are clustered at the patient level to allow for non-independence between patient records.

Table 4: Effects of Cancelled Care on Covid and Non-Covid Mortality

	First Stage	Covid Mortality ITT	Covid Mortality IV	Non-Covid Mortality ITT	Non-Covid Mortality IV
March Cohort	0.14671	0.00011	—	0.00033	—-
	(0.00041)	(0.00003)		(0.00008)	
Cancelled Visit	—	—	0.00072	—	0.00226
	—	—	(0.00022)	—	(0.00055)
Intercept	0.1896	0.00093	0.00079	0.00738	0.00695
	(0.00026)	(0.00003)	(0.00006)	(0.00008)	(0.00015)
N			5,665,641		

Notes - The unit of observation is the patient-appointment. Only appointments in February-March and March-April 2020 enter the regressions. Column 1 reports the first stage estimate from the linked deaths EMR data. Columns 2 and 3 report the ITT and IV estimates at 12 months for covid19 likely deaths. Columns 4 and 5 report the ITT and IV estimates at 12 months for covid19 unlikely deaths. For ICD codes that define these groupings see the Data Appendix. All regressions include no controls and standard errors are clustered at the patient level to allow for non-independence between patient records.

Table 5: Complier Characteristics

	Linked EMR data		Unlinked EMR data	
	February-March	Compliers	February-March	Compliers
Age 0-17	0.100	0.06	0.111	0.07
Age 18-25	0.06	0.03	0.05	0.03
Age 26-35	0.08	0.06	0.08	0.06
Age 36-45	0.100	0.08	0.102	0.08
Age 46-54	0.120	0.10	0.120	0.10
Age 55-64	0.180	0.18	0.178	0.18
Age 65+	0.360	0.48	0.360	0.47
Age 65-74 (from unlinked data)	—	—	0.195	0.25
Age 75-84 (from unlinked data)	—	—	0.118	0.16
Age 85+ (from unlinked data)	—	—	0.043	0.062
Black	0.09	0.06	0.09	0.06
White	0.49	0.42	0.50	0.43
Female	0.59	0.59	0.57	0.58
Charlson-0	0.80	0.76	0.80	0.76
Charlson-1	0.06	0.042	0.06	0.042
Charslon-2+	0.13	0.238	0.13	0.238
Predicted Mortality	0.0069	0.00719	—	—
N	5,665,641		5,625,882	

Notes- The unit of observation is the patient-appointment. Only appointments in February-March and March-April 2020 enter the regressions. Column 1 reports the mean in the February-March cohort and column 2 reports the estimated mean among the compliers. All coeffecients except Age 65-74, Age 75-84 and Age 85+ are from the linked mortality EMR sample. To obtain estimates for more granular age brackets we used the unlinked EMR data. N= 5,665,641 in the linked data and N = 5,625,882 in the unlinked data.

Table 6: Effects of Cancelled Care on All Cause Mortality by Early Pandemic Severity

	Six Month Mortality per 10,000		One Year Mortality per 10,000	
	ITT	IV	ITT	IV
Pooled Model (All States)	3.4	23	4.3	29.7
	(0.7)	(4.5)	(0.9)	(5.9)
High Vs Low Pandemic Severity				
Low Early Covid	3.7	26.6	4.8	34.9
	(0.8)	(5.8)	(1.1)	(7.7)
High Early Covid	2.59	14.5	4.4	19.2
	(1.0)	(5.84)	(1.4)	(7.7)
N		5,665,641		

Notes - The unit of observation is the patient-appointment. The top panel reports our main ITT and IV for all-cause mortality at 6 months and 12 months after the visit date. The bottom panel reports the same estimates but from a regression with an added interaction term for states with high early covid-19 cases per 100k between March and April 2020. Coefficients and standard errors for the low early covid group are obtained using linear combination of coefficients post estimation. All reported coefficients are converted to deaths per 10,000 (we multiply the coefficient estimate by 10,000).

Table 7: Effects of Cancelled Care on Downstream Outpatient Visits, Laboratory Tests, and Total Charges

	Downstream Outpatient Visits		**Laboratory Tests**		**Total Charges**	
Panel A: 2020	March Cohort	Cancelled Visit	March Cohort	Cancelled Visit	March Cohort	Cancelled Visit
One Month Post ITT	-0.26738	–	-0.717	–	-87.02	–
	(0.00114)	–	(0.006)	–	(1.190)	–
One Month Post IV	–	-0.939	–	-5.943	–	-727.02
	–	(0.0090)	–	(0.059)	–	(10.05)
Six Months Post ITT	0.09821	–	-0.566	–	29.90	–
	(0.00295)	–	(0.013)	–	(2.269)	–
Six Months Post IV	–	0.585	–	0.554	–	249.87
	–	(0.0180)	–	(0.109)	–	(18.97)
Number of Observations	5,491,077	5,491,077	5,491,077	5,491,077	5,491,077	5,491,077
	Downstream Outpatient Visits		**Laboratory Tests**		**Total Charges**	
Panel A: 2019	March Cohort	Cancelled Visit	March Cohort	Cancelled Visit	March Cohort	Cancelled Visit
One Month Post ITT	-0.03706	–	-0.031	–	-3.769	–
	(0.0012)	–	(0.008)	–	(1.362)	–
One Month Post IV	–	14.15117	–	11.973	–	1444.71
	–	(1.555)	–	(3.418)	–	(545.04)
Six Months Post ITT	-0.11569	–	-0.105	–	0.243	–
	(0.0031)	–	(0.014)	–	(2.315)	–
Six Months Post IV	–	44.17030	–	40.364	–	-93.269
	–	(4.627)		(6.688)	–	(887.61)
Number of Observations	5,097,076	5,097,076	5,097,076	5,097,076	5,097,076	5,097,076

Notes- The unit of observation is the patient-appointment. Panel A reports estimates for the 2020 March-April and February-March scheduled visits. Panel B reports estimates for the 2019 March-April and February-March scheduled visits. In each panel, columns 1 and 2 report the ITT and IV estimates for number of outpatient visits (not conditioning on visits to the same physician associated with the focal visit) after 1 and 6 months of the visit date. Columns 3 and 4 report the ITT and IV estimates for the number of laboratory tests ordered (not conditioning on labs ordered by the same physician associated with the focal visit) after 1 and 6 months of the visit date. Columns 5 and 6 report the ITT and IV estimates for total charges billed (not conditioning on labs ordered by the same physician associated with the focal visit) after 1 and 6 months of the visit date. All regressions include no controls and standard errors are clustered at the patient level to allow for non-independence between patient records. For magnitude calculations we report the constant term from the ITT regression which is the the baseline (February-March) mean. The baseline mean number of visits 1 and 6 months after the visit date is 1.5 and 3.7 in 2020. Its 1.6 and 4.1 in 2019. The baseline mean number of labs ordered 1 and 6 months after the visit date is 2.1 and 5.2 in 2020. It is 2.2 and 1.04 in 2019. The baseline mean total charges 1 and 6 after the visit date is $270 and $696 in 2020. It is $313 and $844 in 2019.

Table 8: Effects of Cancelled Care on All Cause Mortality at Specific Calendar Months

		September 2020 Mortality		March 2021 Mortality	
	First Stage	ITT	IV	ITT	IV
March Cohort	0.149	0.0005	–	0.0006	–
	(0.00039)	(0.00006)	–	(0.00009)	–
Cancelled Visit	–	–	0.00357	–	0.00405
	–	–	(0.00044)	–	(0.00059)
N			5,665,641		

Notes- The unit of observation is the patient-appointment. Only appointments in February-March and March-April 2020 enter the regressions. Column 1 reports the first stage estimate from the linked deaths EMR data. Columns 2 and 3 report the ITT and IV estimates at September 2021 (6th month after the visit date). Columns 4 and 5 report the ITT and IV estimates at March 2021 (the 12th month after the visit date). All regressions include no controls and standard errors are clustered at the patient level to allow for non-independence between patient records.

Table 9: Effects of Cancelled Care on All Cause Mortality Model with Controls

		Six Month Mortality		**One Year Mortality**	
Panel A: 2020	First Stage	ITT	IV	ITT	IV
March Cohort	0.149	0.0002	–	0.00019	–
	(0.00039)	(0.00007)	–	(0.00009)	–
Cancelled Visit	–	–	0.00132	–	0.00129
	–	–	(0.00045)	–	(0.00067)
Number of Observations	5,665,641	5,665,641	5,665,641	5,665,641	5,665,641
		Six Month Mortality		**One Year Mortality**	
Panel B: 2019	First Stage	ITT	IV	ITT	IV
March Cohort	-0.005	-0.00005	–	0.00001	–
	(0.0004)	(0.00006)	–	(0.00008)	–
Cancelled Visit	–	–	0.011	–	-0.002
	–	–	(0.022)	–	(0.014)
Number of Observations	5,382,158	5,382,158	5,382,158	5,382,158	5,382,158

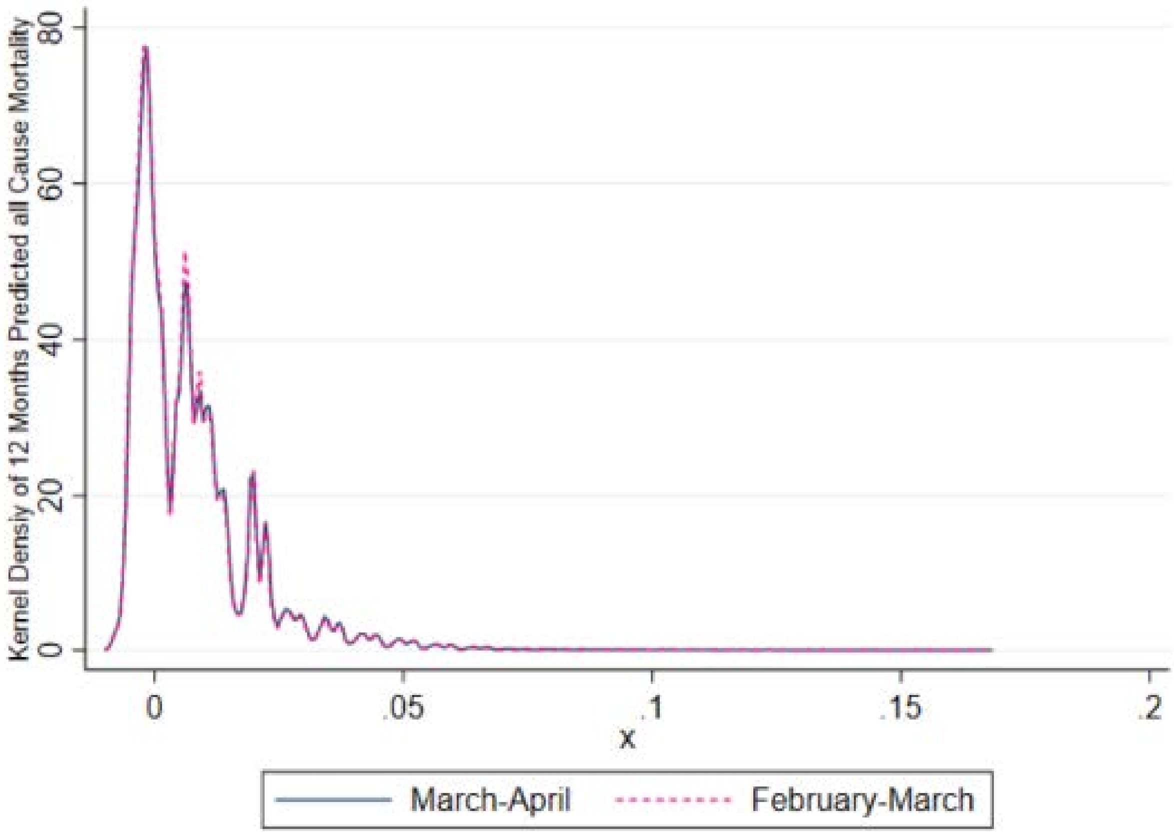
Notes - The unit of observation is the patient-appointment. Panel A reports the 2020 estimates and there only appointments scheduled for February-March and March-April 2020 enter the regression. Panel B reports the 2019 estimates and there only appointments scheduled for February-March and March-April 2019 enter the regressions. In both panels, column 1 reports the first stage estimate from the linked deaths EMR data. Columns 2 and 3 report the ITT and IV estimates at 6 months after the visit date. Columns 4 and 5 report the ITT and IV estimates at 12 months after the visit date. All regressions include controls for age (7 age brackets: 0-17, 18-25, 26-35, 36-45, 46-54, 65+), race (black, white, other/missing), female, three indicator terms for charlson score equal 0, equal 1 and at or above 2, and state fixed effects. Standard errors are clustered at the patient level to allow for non-independence between patient records.

Table 10: Effects of Cancelled Care on All Cause Mortality Among Appointments Booked Prior to 13th of each Focal Period

		Six Month Mortality		One Year Mortality	
	First Stage	ITT	IV	ITT	IV
March Cohort	0.186	0.00015	–	0.0004	–
	(0.0009)	(0.00008)	–	(0.00013)	–
Cancelled Visit	–	–	0.0009	–	0.00216
	–	–	(0.0005)	–	(0.0007)
Intercept	0.224	0.0023	0.0002	0.0052	0.0047
	(0.0007)	(0.00008)	(0.00016)	(0.00013)	(0.00026)
N			1,612,887		

Notes- The unit of observation is the patient-appointment. Only appointments in February-March and March-April 2020 enter the regressions. We select visits booked before February 13th in the February- March group and visits booked before March 13th in the March-April group. Column 1 reports the first stage estimate from the linked deaths EMR data. Columns 2 and 3 report the ITT and IV estimates at 6 months after the visit date. Columns 4 and 5 report the ITT and Iv estimates at 12 months after the visit date. All regressions include no controls and standard errors are clustered at the patient level to allow for non-independence between patient records.

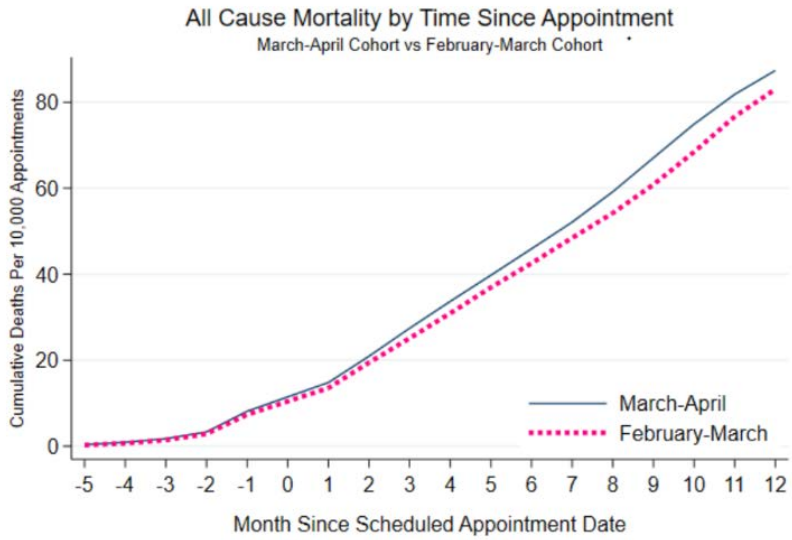
Figure 1: Predicted Mortality for March-April vs February-March 2020 Patients



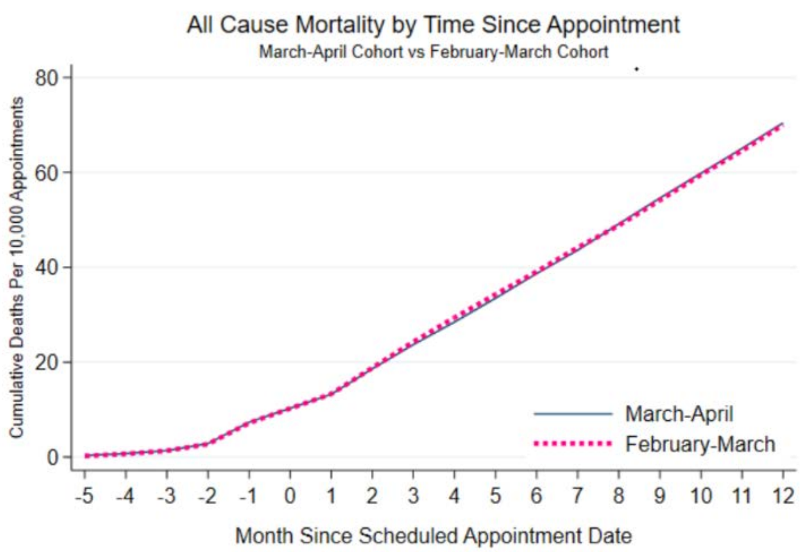
Notes - Distribution of predicted mortality. Figure shows distribution of predicted mortality after 1 year of visit date for the 2020 cohorts (N= 5,665,641). Prediction model is based on 2019 data (N= 5,382,158). Predictors include: The age in brackets available to us in the linked data (0-17, 18-25, 26-35, 46-54, 55-64,65+), race (white, black, other/missing), female, the charlson comorbidity score and state fixed effects.

Figure 2: Cumulative All-Cause Mortality Before and After the Scheduled Visit Date

(a) February-March vs March-April 2020

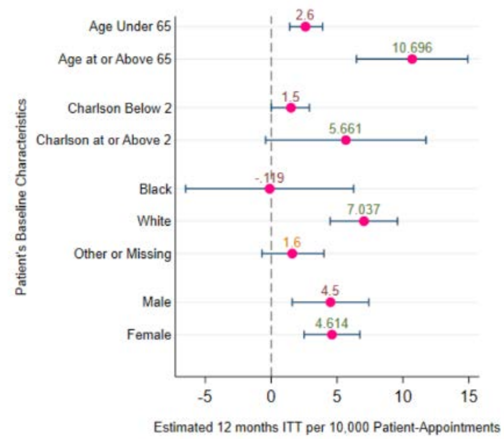


(b) February-March vs March-April 2019

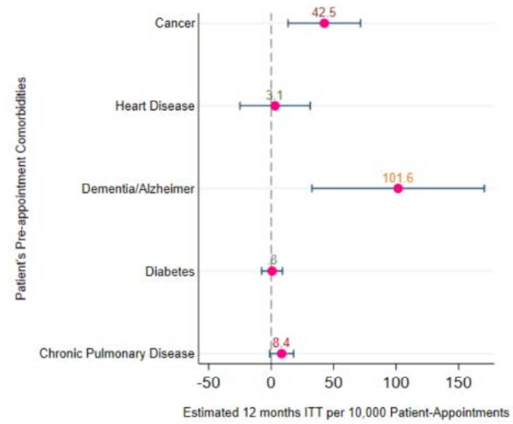


Notes: The unit of observation is the patient-appointment. In each figure, we compare the February-March cohort to the March-April cohort. The first figure included the 2020 cohorts and plots trends in cumulative mortality before the visit date by 5-months and after the visit date by 12-months. Patients deceased before their scheduled visit are counted in pre-visit mortality. The second figure follows accordingly and plots the 2019 cohorts of patients.

Figure 3: ITT Estimates of 12 Month Mortality Effect Across Patient Groups

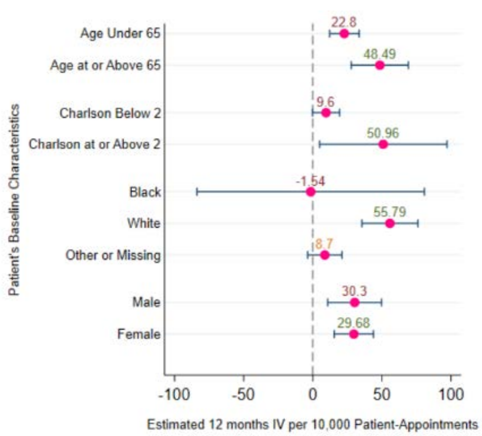


(a) By Baseline Patient Characteristics

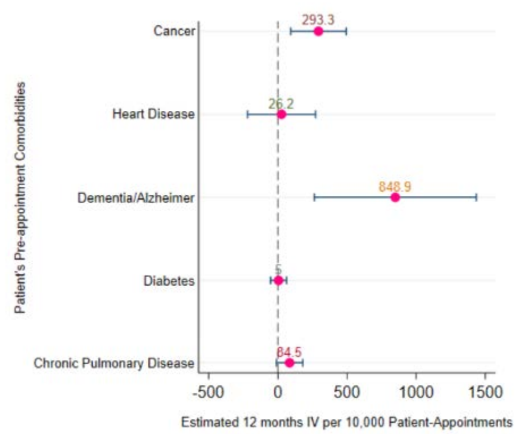


(b) By Chronic Conditions Diagnosed Prior to the Appointment

Figure 4: IV Estimates of 12 Month Mortality Effect Across Patient Groups



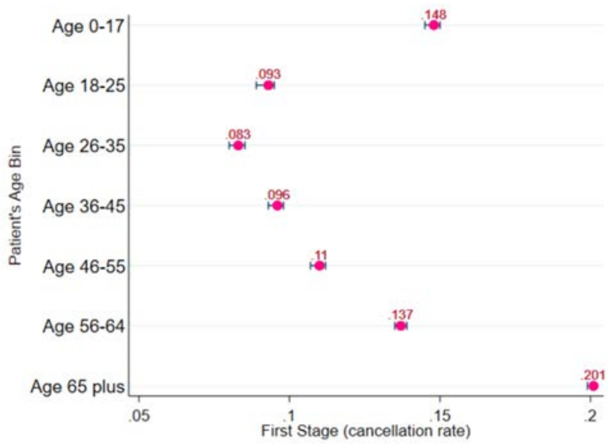
(a) Baseline Patient Characteristics



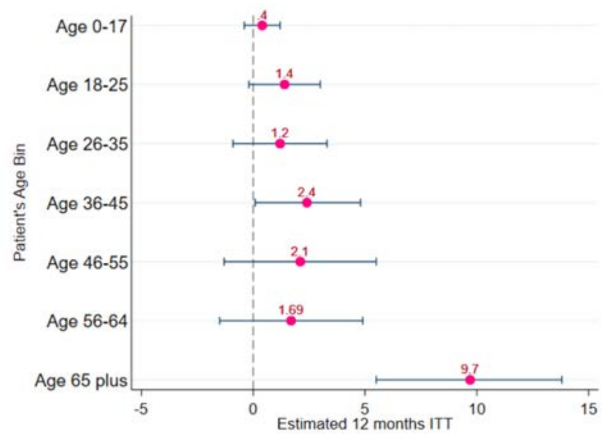
(b) Chronic Conditions Diagnosed Prior to the Appointment

Notes: Figure 4 Panel A reports the 12 month ITT estimates from four separate regressions, one for each demographic trait (age, charlson score, race and gender). Figure 4 Panel B reports the ITT estimates from five separate regressions, one for each baseline disease. Figure 5 panels A and B present the corresponding IV estimates. In all regressions, the unit of observation is the patient-appointment. All regressions include no controls and standard errors are clustered at the patient level to allow for non-independence between patient records.

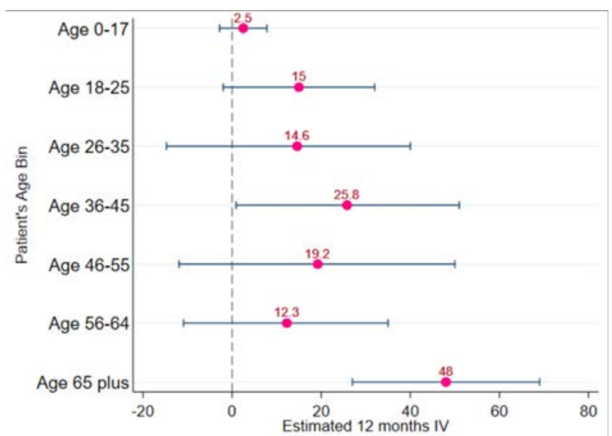
Figure 5: IV, ITT and First Stage Estimates of 12 Month Mortality Effect Across Patient Age Groups



(a) First Stage Estimates by Age Group of Patient



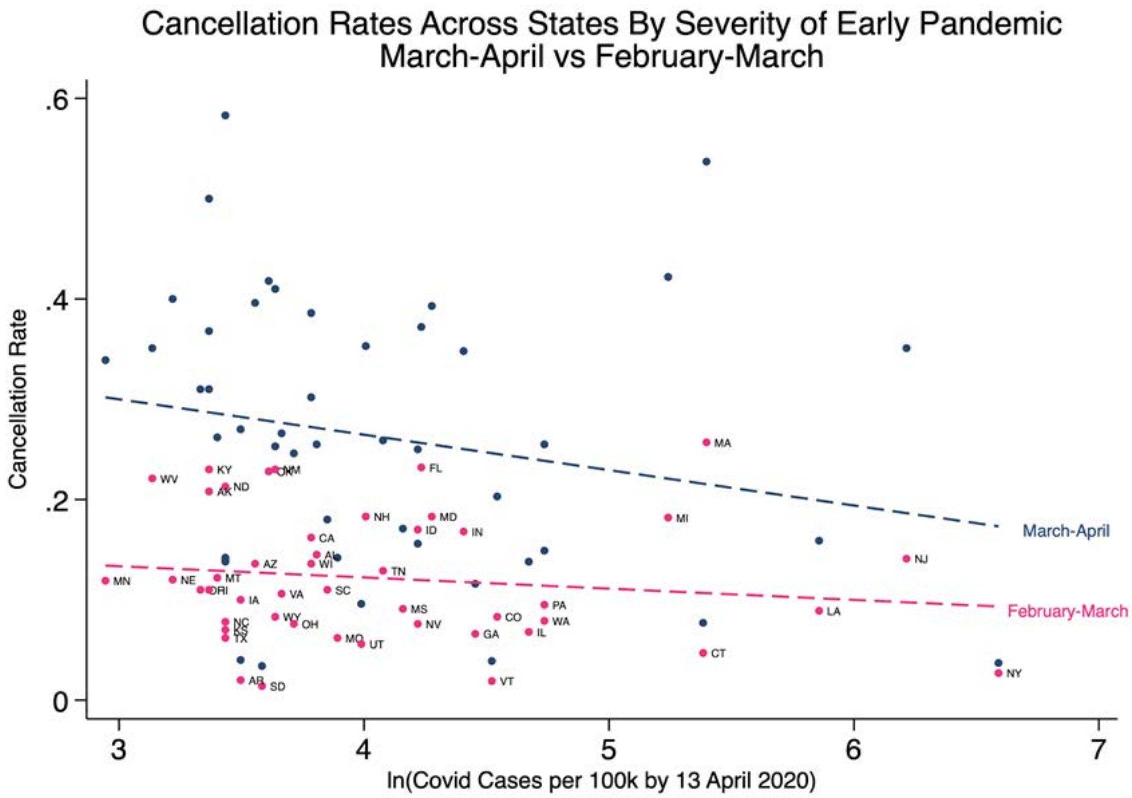
(b) ITT Estimates by Age Group of Patient



(c) IV Estimates by Age Group of Patient

Notes: Figure 6 Panel A reports the first stage estimates for each age group. Figure 6 Panel B reports the ITT estimates from a single regression with interaction terms for each age group. We use linear combination (lincom) to combine the coefficient on the indicator for March-April and the coefficient on the age group specific interaction (eg: March-April X Age18 to 25) and obtain ITT estimates for each age group. Figure 6 panel C shows the IV estimates for 12 months downstream mortality. In all regressions, the unit of observation is the patient-appointment. All regressions include no controls and standard errors are clustered at the patient level to allow for non-independence between patient records.

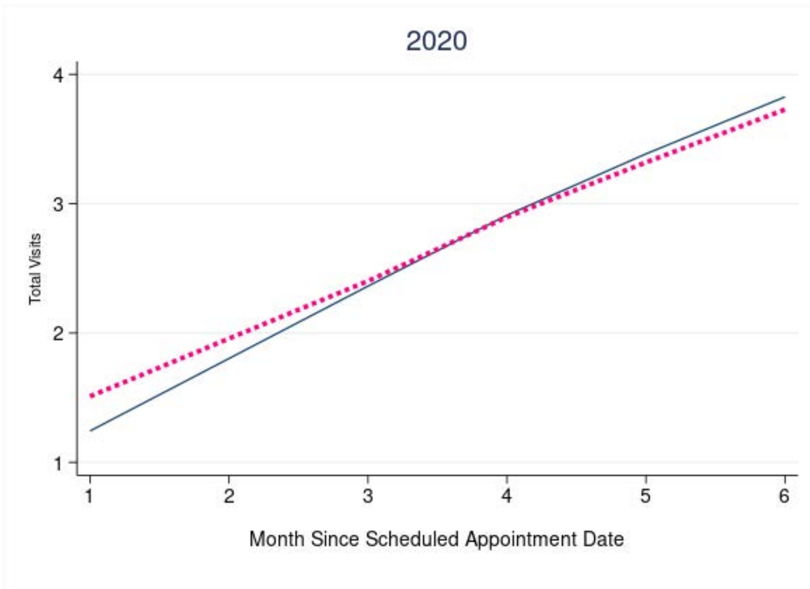
Figure 6: Cancellation Rates and Total Cases Per 100k Across State up to April 13th 2020



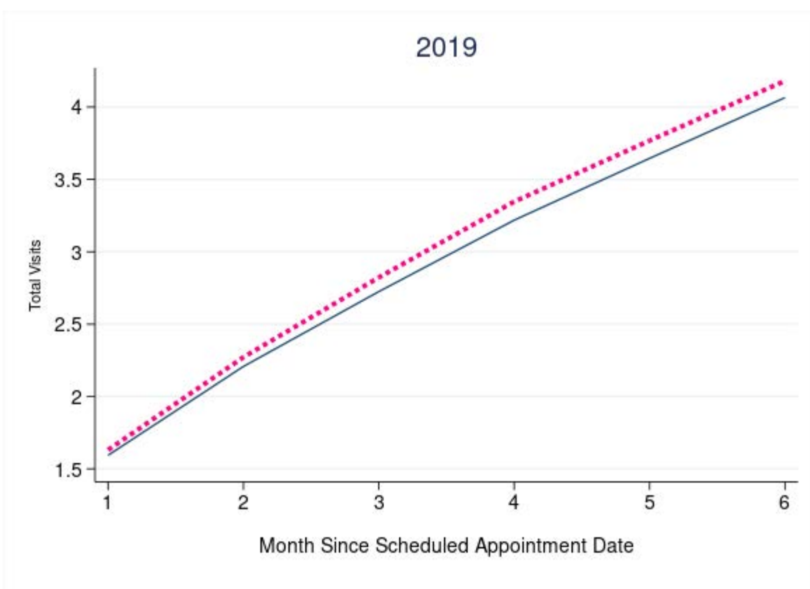
Notes: The unit of observation is the state. The x-axis plots log cumulative cases per 100,000 in a state up to April 13th 2020. The Y-axis plots the cancellation rate for the March-April cohort relative to the February-March cohort. To obtain the state cancellation rate we estimate the first stage regression on individual state samples.

Figure 7: Cumulative Number of Visits Attended after the Focal Visit

(a) February-March vs March-April 2020



(b) February-March vs March-April 2019

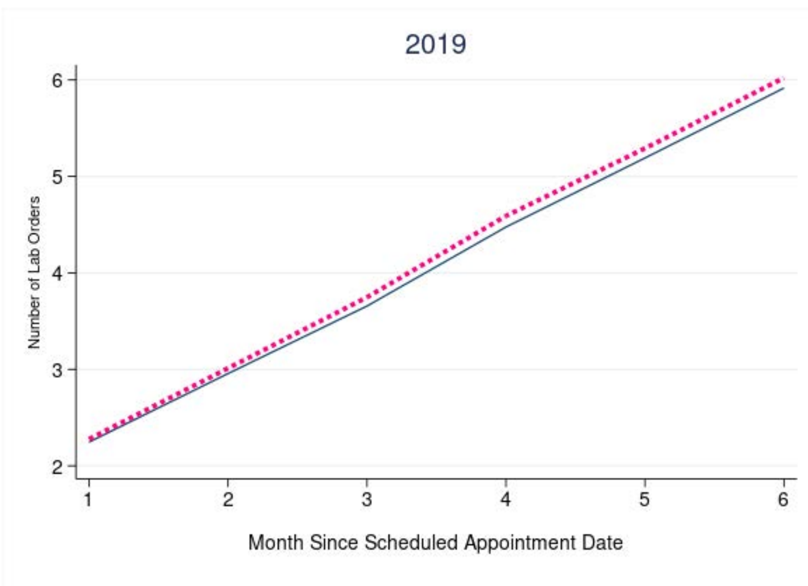


Notes: The unit of observation is the patient-appointment. The top figure plots the average cumulative number of visits attended by the February-March cohort vs the March-April cohort in the 6 month proceeding the focal visit. The count excludes the focal visit. The bottom plots the same but now with 2019 data

Figure 8: Cumulative Number of Laboratory Tests Ordered after the Focal Visit

(a) February-March vs March-April 2020

(b) February-March vs March-April 2019

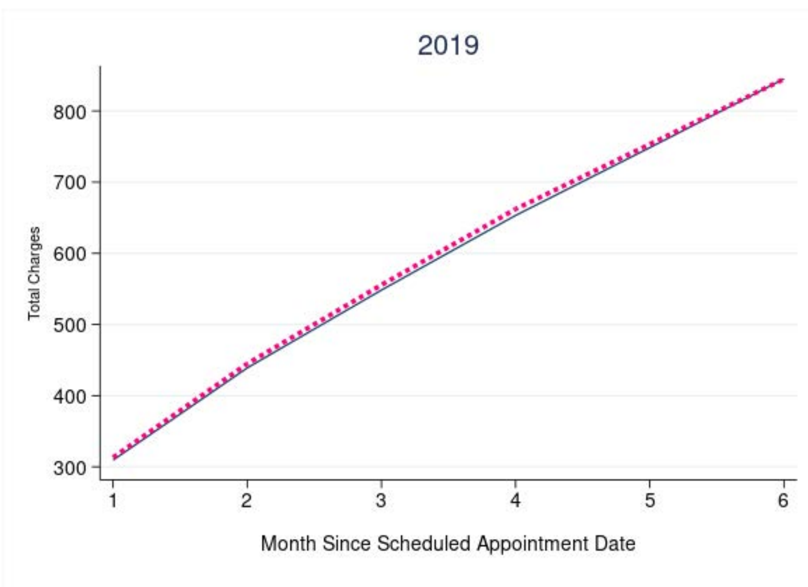


Notes: The unit of observation is the patient-appointment. The top figure plots the average cumulative number of laboratory tests ordered for the February-March cohort vs the March-April cohort in the 6 months proceeding the focal visit. The count excludes labs ordered on the focal visit date. The bottom plots the same but now with 2019 data.

Figure 9: Cumulative Total Outpatient Charges in US Dollars after the Focal Visit

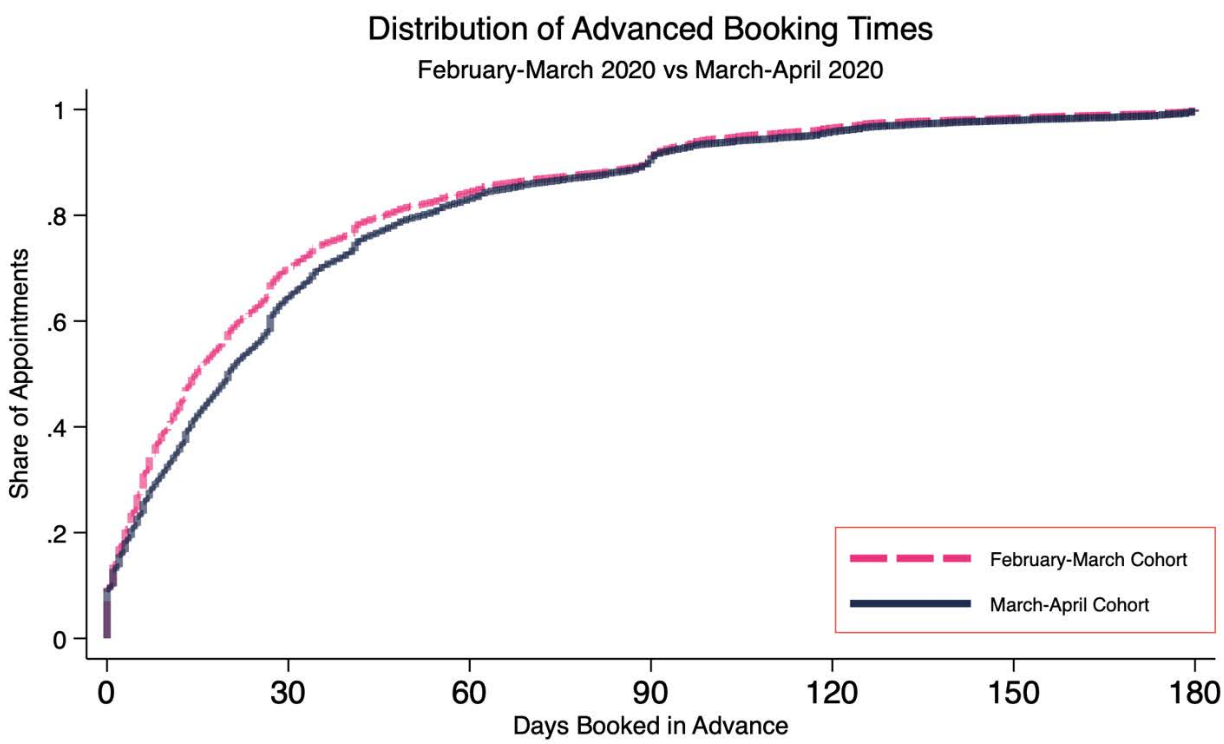
(a) February-March vs March-April 2020

(b) February-March vs March-April 2019



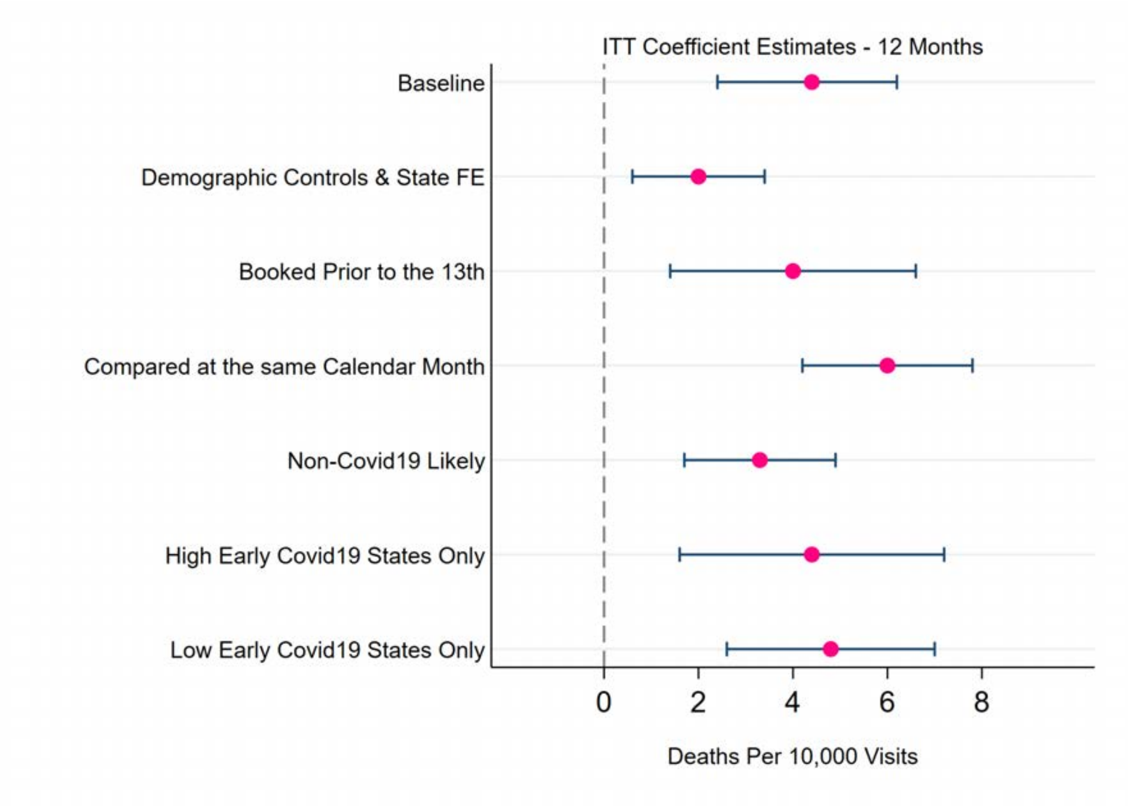
Notes: The unit of observation is the patient-appointment. The top figure plots the average cumulative total outpatient charges for the February-March cohort vs the March-April cohort in the 6 months proceeding the focal visit. The count excludes the focal visit. The bottom plots the same but now with 2019 data.

Figure 10: Cumulative Distribution of Lag in Days between a Visit and its Booking Date



Notes - The sample of appointments used in the plot include only appointments for which the booking date is observed(see Data section). Each line reflects the cumulative distribution function for the number of days a visit is booked in advance. Since most visits are booked less than 180 days ahead we truncate the data at 180 days for clarity.

Figure 11: Effects of Cancelled Care on Mortality - Alternative Specifications



Notes: The figure plots the 12 month ITT coefficient for the 2020 sample for the model that has no controls, adds controls, conditions on a sample of visits booked prior to the 13th, compares the two cohorts in the same calendar month (March 2021), examines only differences in Non-Covid19 Likely deaths, conditions on only high Covid19 burden in March-April (the early pandemic period) and conditions only on low Covid19 burden in March-April. Each coefficient is from a separate regression reported in the main tables.

Appendix Table 1 : Sample Size of Electronic Medical Records and Deaths Data

Number of Patients per Cohort by SSN Presence on EMR Record				
A: Patients	*With SSN Present*	*Without SSN Present*	*Total Number of Patients*	
Feb 13th to March 12th-2019	1,209,114	532,722	1,741,836	
March 13th to April 12th-2019	1,311,274	584,092	1,895,366	
Feb 13th to March 12th-2020	1,378,431	673,298	2,051,729	
March 13th to April 12th-2020	1,159,558	540,807	1,700,365	
Deaths Records from CDI				
B: Deaths	*With SSN Present (Deaths Master File)*	*Without SSN Present (Obituaries)*	*Total Deaths in CDI*	*CDC Reported Deaths*
CY 2019 Deaths	716,565	1,715,760	2,432,325	2,854,838
CY 2020 Deaths	781,882	1,970,905	2,752,787	3,358,814
CY 2021 Deaths	360,091	913,186	1,273,277	3,458,697
Linked EMR Patients with Deaths in CDI				
C: Linked Patients to Deaths	*With SSN Present (Deaths Master File)*	*Without SSN Present (Obituaries)*	*Total Number of Patients Linked to Deaths*	
Feb 13th to March 12th-2019	21,940	6,062	28,002	
March 13th to April 12th-2019	22,993	6,487	29,480	
Feb 13th to March 12th-2020	15,819	5,191	21,010	
March 13th to April 12th-2020	13,406	4,140	17,546	

Notes: Panel A reports the unique number of patients in the HealthJump data by cohort. The table divides patients into those with a social security number (SSN) reported and those without SSN. Panel B reports the unique number of deaths in the Connected Death Index (CDI) data. Here, the table also divides patients into those with and without SSN in the CDI. Since the CDI only provides the month-year of death, we are unable to categorize deaths into day-month-year cohorts. We report annual deaths for 2019, 2020 and 2021. 2021 CDI deaths were obtained in June 2021, hence the count is incomplete. In Panel B, an additional column reports the number of deaths reported by the CDC in each year. For 2021, the estimated number of deaths is approximate and was reported by the CDC in April 2022. Panel C reports the number of patients in each appointment cohort who were found in the CDI data by SSN or by probabilistic terms (without SSN).

Appendix Table 2 : Covariate Balance 2020 - Unlinked EMR-Only Sample

	February-March Mean	Difference	Cohen's D
Age 0-17	0.111	-0.02	0.068
Age 18-25	0.05	0.00	-0.00
Age 26-35	0.08	0.005	-0.02
Age 36-45	0.102	0.005	-0.015
Age 46-54	0.120	0.005	-0.014
Age 55-64	0.178	0.007	-0.02
Age 65+	0.360	-0.001	0.002
Age 65-74	0.195	0.002	-0.004
Age 75-84	0.118	-0.002	0.004
Age 85+	0.043	-0.001	0.005
Black	0.09	0.005	-0.02
White	0.50	0.003	-0.005
Female	0.57	0.002	-0.004
Charlson Score = 0	0.80	-0.01	0.03
Charlson Score = 1	0.06	0.00	-0.01
Charlson Score $\geq$ 2	0.13	0.01	-0.03
N		5,625,882	

Notes - The unit of observation is the patient-appointment. Each row is estimated from a regression of baseline characteristics (10 age brackets, race, gender or baseline morbidity) on an indicator for the record being in the March-April cohort. The coefficient from the constant term is the February-March mean, the coefficient from the March-April indicator is the difference between the two cohorts. Given the large sample size we report the Cohen's d statistic to measure the standardised difference between two means instead of P-values.

Appendix Table 3 : Complier Physician Speciality

	February-March Average	Complier Average	Cum. % of Compliers
Family Practice	0.442	0.36	0.36
General Surgery	0.093	0.097	0.457
Ophthalmology	0.039	0.094	0.551
Physical Therapy	0.033	0.037	0.588
Orthopedic Surg/Sports Med	0.023	0.028	0.616
Pediatrics	0.037	0.026	0.642
Cardiology	0.011	0.022	0.664
Obstetrics and Gynecology	0.017	0.021	0.685
Neurology	0.015	0.019	0.704
Podiatry	0.010	0.014	0.718
Dentistry	0.009	0.013	0.731
otolaryngology	0.007	0.013	0.744
Endocrinology	0.009	0.012	0.756
Psych/Mental Health	0.008	0.009	0.765
Dermatology	0.007	0.008	0.773
Oncology/Hematology	0.003	0.004	0.777
Dietitian/Nutritionist	0.002	0.002	0.779
Allergy/Immun	0.001	0.002	0.781
Geriatric Medicine	0.001	0.002	0.783
Nephrology and Urology	0.001	0.001	0.784
Proctology	0.000	0.00	0.784
All Other	0.24	0.23	1.0
N	5,625,882		

Notes- The unit of observation is the patient-appointment. Only appointments in February-March and March-April 2020 enter the regressions. Column 1 reports the mean in the February-March cohort and column 2 reports the estimated mean among the compliers. The table is produced from the unlinked mortality EMR sample where we observe the specialty of the scheduled appointment's physician.

Appendix Table 4 : Covariate for Appointments Booked Prior to the 13th of March in March-April Group and Prior to the 13th of February in the February-March Group

	February-March Mean	Difference	Cohen's D
Age 0-17	0.06	0.0009	-0.004
Age 18-25	0.044	-0.002	0.010
Age 26-35	0.071	-0.004	0.014
Age 36-45	0.100	-0.004	0.014
Age 46-54	0.124	-0.0025	0.007
Age 55-64	0.199	-0.001	0.004
Age 65+	0.404	0.013	-0.026
Black	0.084	0.004	-0.014
White	0.550	-0.006	0.012
Female	0.591	0.0000	0.000
Charlson Score = 0	0.82	-0.015	0.040
Charlson Score = 1	0.06	0.003	-0.015
Charlson Score $\geq$ 2	0.12	0.011	-0.034
Predicted Mortality	0.0034	0.0002	0.035
N	1,612,887		

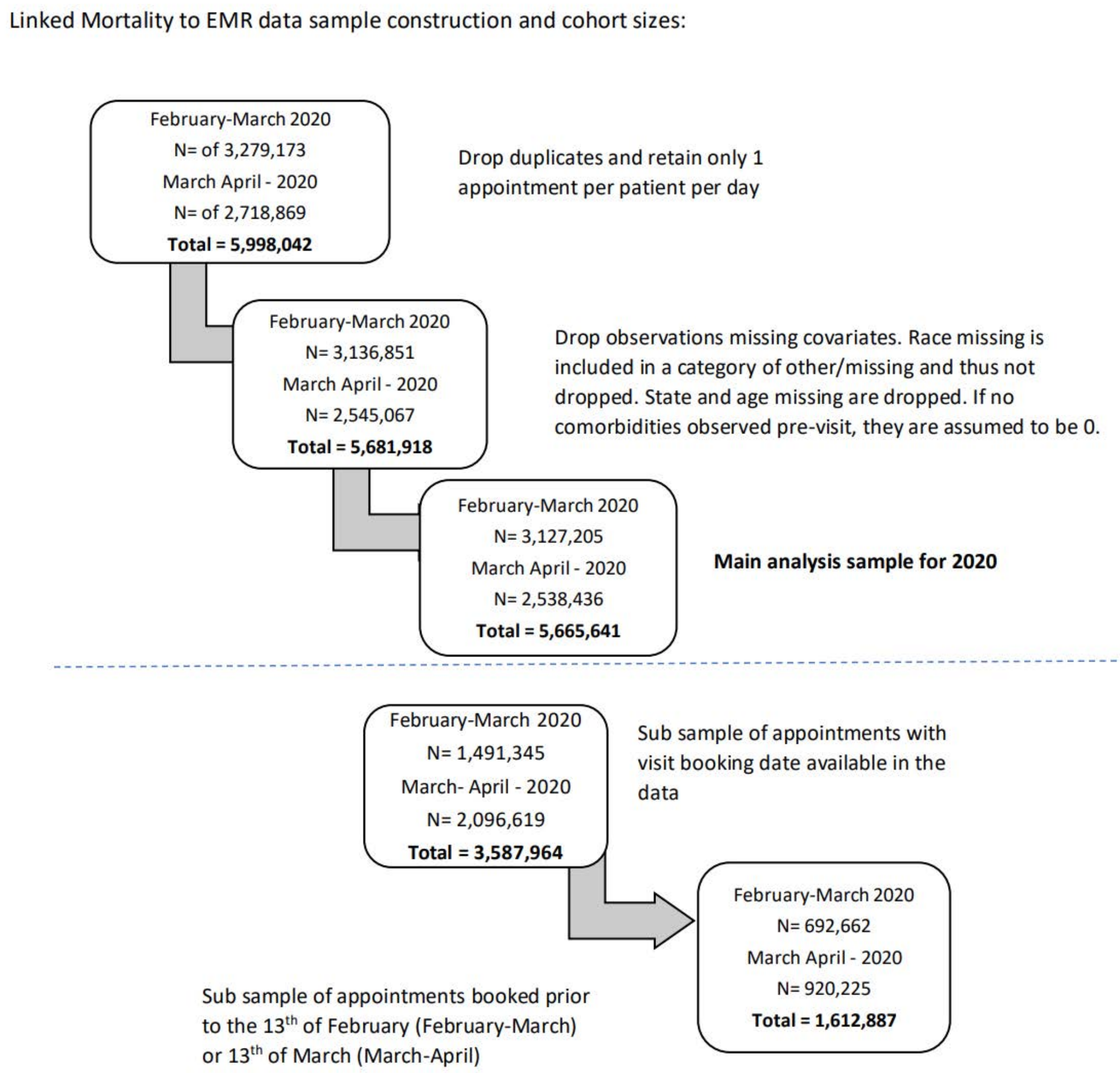
Notes - The unit of observation is the patient-appointment. From the sub sample of appointments for which we observe the booking date, we limit sample to appointments booked prior to the 13th of March in the March-April cohort and appointments booked prior to the 13th of February in the February-March cohort. Each row is estimated from a regression of baseline characteristics (7 age brackets, race, gender or baseline morbidity) on an indicator for the record being in the March-April cohort. The coefficient from the constant term is the February-March mean, the coefficient from the March-April indicator is the difference between the two cohorts. Given the large sample size we report the Cohen's d statistic to measure the standardised difference between two means instead of P-values.

Appendix Table 5 : Effects of Cancelled Care on All Cause Mortality Among Appointments Booked Pre-13th of each Focal Month -2019

		Six Month Mortality		One Year Mortality	
	First Stage	ITT	IV	ITT	IV
March Cohort	0.01	-0.00005	–	0.00025	–
	(0.00125)	(0.0001)	–	(0.00016)	–
Cancelled Visit	–	–	-0.00149	–	0.001
	–	–	(0.0028)	–	(0.0044)
Intercept	0.254	0.00153	0.0005	0.0034	0.00006
	(0.001)	(0.0001)	(0.0007)	(0.00015)	(0.0011)
N			700,336		

Notes- The unit of observation is the patient-appointment. Only appointments in February-March and March-April 2019 enter the regressions. We select visits booked before February 13th in the February- March group and visits booked before March 13th in the March-April group. Column 1 reports the first stage estimate from the linked deaths EMR data. Columns 2 and 3 report the ITT and IV estimates at 6 months after the visit date. Columns 4 and 5 report the ITT and Iv estimates at 12 months after the visit date. All regressions include no controls and standard errors are clustered at the patient level to allow for non-independence between patient records.

Appendix Figure 1 : 2020 Focal Appointments Sample Construction Steps



Mortality Effects of Healthcare Supply Shocks: Evidence Using Linked Deaths and Electronic Health: Online Data Appendix

A. Linking the CDI data to EMR Records

We use tokenized identifiers to link the CDI data to the EMR data from HealthJump. The Connected's Death Index (CDI) is a dataset containing information on deceased individuals in the United States and Canada. It contains over 100 million historical deaths, with roughly 50 thousand deaths added weekly. Within this data, delivered to us, are several Datavant tokens. The tokens are Datavant's proprietary technology where each individual in the population is assigned up to 17 encrypted tokens that can link individuals across datasets without releasing person identifiers to researchers or firms. Not all 17 Datavant tokens were present in the HealthJump data. Below we list the tokens we received in both the CDI data and the EMR data and how we chose what tokens to priotrize:

List of tokens in both the EMR and CDI data:

Token 1 - Datavant's standard token 1 (lastname + first initial + dob + gender)

Token 2 - soundex(lastname) + soundex(firstname) + dob + gender)

Token 4- last name + first name + gender + DOB

Token 5- SSN + DOB + gender ; only found on SSA records

Token 7- last name + 1st 3 characters of first name + gender + DOB

Token 16 - SSN + first name

The soundex variable allows for matching despite minor differences in spelling. Soundex enables homophones (eg: Jon and John) to be encoded to the same representation.

In order to understand which tokens we should prioritize, we relied on previous research by Datavant scientists who have documented in a white paper the ranking of these tokens from highest to lowest in accuracy for linking data [15]. Recognizing that there is a tradeoff between Type I (not linking deaths to EMR records when in fact patient is deceased) and Type II error (over linking deceased death records to the EMR data when patient is in fact alive), we chose to minimize Type II error. Thus we linked on the most specific tokens in the data. The below summarizes the rule we used to link the deaths data to the EMR records. We provide PSQL code for these steps for readers who may find this useful.

[15] https://covid19researchdatabase.org/wp-content/uploads/2020/05/Token-Selection-Deep-Dive-updated-May-2020-1.pdf last accessed: 8/28/2022

Rule 1 Prioritize deterministic tokens: we prioritized SSN + first name (Token 16). If a patient has SSN in the EMR data we search for her record in the CDI data based on SSN + first name.

Rule 2 Prioritize probabilistic tokens built from frequently occurring fields:

Frequently occurring fields include names, and date of birth. Less frequent fields include address and zip code. Thus, if the patient has no SSN + first name in the EMR data, we used Tokens 1,2 and 4 (defined above) together to search for her in the death data.

We note here that all patients in the CDI data and EMR data have Tokens 1,2 and 4. This means, for patients who have both deterministic and probabilistic tokens in the deaths data, we can compare the match rates between the EMR data and the CDI data when we use SSN + first name or when we use the probabilistic tokens. We describe this test and more on how we validate the quality of the linkage below.

B. Testing the Validity of the link between the CDI data and EMR records

We test the accuracy of the link between the CDI data and the EMR data in two ways:

First, we measured the number of patients redeemed deceased who still appear to continue care post death in the EMR data. We searched the procedures, medications, and encounter files (encounters include visits and ER admissions), for care dates after the death date reported in the CDI. The figure below gives an example of how we conducted this test. For each visit date we built a table where we capture the last treatment date (eg: last encounter date in the encounter file or last procedure date in the procedure file). We note here that deaths date in the data is at the mm/yy level and deaths are assumed on the 1st of every month. So we measure, follow-up care in months t+1 to avoid confounding valid treatment that occurs in the days before death in our test. We then calculated for what share of patients do we observe a treatment date in the month t+1 to t+12 after the death date. We found 0% of records flagged as deceased have a treatment date in months t+1 to t+12 after the death date. The share is 0% whether we link of deterministic terms or probabilistic terms.

Data Appendix Figure 1 : Assigning death record in the linked death-EMR Data the date of last procedure, medication order or healthcare encounter

100 records preview

- Encounter

```
SELECT * FROM "COVID19_COMMONPROFILE"."COVID19_COMMONPROFILE_00053"."ENCOUNTER_DATE_DIFF_RAW" LIMIT 100;
```

CLIENT_ID	PATIENT_ID	TREATMENT_DATE_EN	DEATH_DATE	DATE_DIFF_EN	EN_FLAG
ABM01	139	2020-06-09	2021-04-01	296	0
ABM01	139	2021-03-17	2021-04-01	15	0
ABM01	139	2020-08-28	2021-04-01	216	0
ABM01	139	2020-10-13	2021-04-01	170	0
ABM01	139	2019-10-03	2021-04-01	546	0
ABM01	139	2020-02-20	2021-04-01	406	0

- Procedure

```
SELECT * FROM "COVID19_COMMONPROFILE"."COVID19_COMMONPROFILE_00053"."PROCEDURE_DATE_DIFF_RAW" LIMIT 100;
```

CLIENT_ID	PATIENT_ID	TREATMENT_DATE_PR	DEATH_DATE	DATE_DIFF_PR	PR_FLAG
BWC01	10081211	2020-10-13	2020-11-01	19	0
BWC01	10081211	2019-07-15	2020-11-01	475	0
BWC01	10081211	2020-02-19	2020-11-01	256	0
BWC01	10081211	2019-10-07	2020-11-01	391	0
BWC01	10081211	2020-06-08	2020-11-01	146	0
BWC01	10081211	2019-05-13	2020-11-01	538	0

- Medication

```
SELECT * FROM "COVID19_COMMONPROFILE"."COVID19_COMMONPROFILE_00053"."MEDICATION_DATE_DIFF_RAW" LIMIT 100;
```

CLIENT_ID	PATIENT_ID	TREATMENT_DATE_ME	DEATH_DATE	DATE_DIFF_ME	ME_FLAG
CVMG01	952007	2021-01-25	2021-03-01	35	0
CVMG01	952007	2020-01-03	2021-03-01	423	0
CVMG01	952007	2020-06-29	2021-03-01	245	0
CVMG01	952007	2020-06-04	2021-03-01	270	0
CVMG01	952007	2020-11-13	2021-03-01	108	0
CVMG01	952007	2020-02-07	2021-03-01	388	0

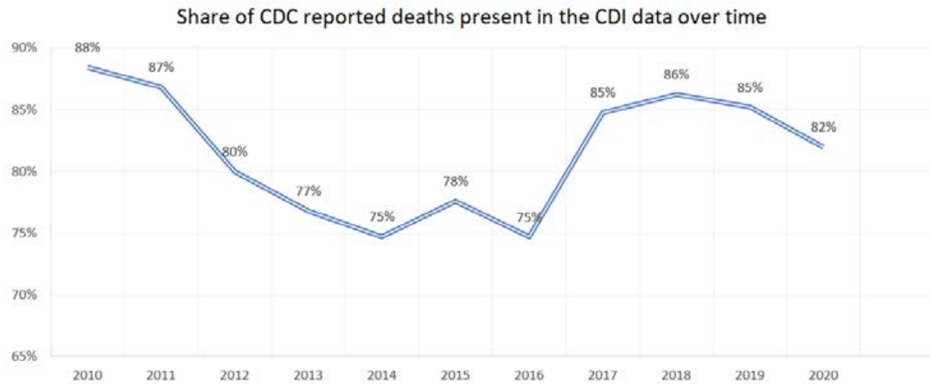
Second, we used the patients who have both SSN and probabilistic terms in both datasets - so they are deceased and can be linked to EMR data on the deterministic token SSN + first n ame. We asked what share of them would not link to the EMR data if we had used only the probabilistic tokens for those patients. We found that 100% of those patients would link to the EMR if we used only the probabilistic tokens meaning the probabilistic tokens are highly accurate.

C. Distribution of Deaths in the CDI data

We searched the procedures, medications, and encounter files (encounters include visits and ER admissions), for care dates after the death date reported in the CDI.

The Figure below shows the share of CDC deaths present in the CDI data by year. In 2011, the CDI stops reporting deaths from states and only deaths from federal sources. Since the CDI is a combination of the Death Master File (DMF) and obituaries, the share of deaths captured decline post 2011 and begin to increase again in 2017 onward. In that time the CDI becomes a much more known data product among pharmaceutical firms known from conversations with Datavant team members and the Datavant mortality product website. In addition, the CDI data is also used to link mortality data to the National Covid Cohort Collaborative (N3C) which began as a result of the pandemic.

Data Appendix Figure 2 : Percent of CDC reported deaths captured by the CDI data



D. Defining Covid19 Likely and Covid19 unlikely Deaths

The following ICD10/ICD9 codes were used to identify Covid19 related deaths

B97.89,H66.9, H66.91, H66.92, H66.93, J00, J01.9, J01.90 J06.9, J09, J09.X, J09.X1 , J09.2, J09.X3 , J09.X9, J10 J10.00, J10.01, J10.08, J10.1, J10.2, J10.8, J10.81, J10.82 J10.83, J10.89, J11, J11.0, J11.08, J11.1, J11.2, J11.8 J11.81, J11.82, J11.83, J11.89, J12.89, J12.9, J18, J18.1 J18.9, J20.9, J40, R05, R50.9, J12.89, J20.8,J22, J40 J80, J98.8, 095.5, R05, R06.02, R50.9, U07.1, Z03.818 Z11.58, Z20.828

E. Value of Statistical Life Lost Calculations

To measure the Value of Statistical Life (VSL) lost from missed care we use an approach based on Years of Potential Life Lost (YPLL). YPLL 75 measures premature mortality assuming end of life at age 75. Several state and federal agencies use the YPLL approach to calculate premature deaths (for Disease Control (CDC et al., 1986; Gardner and Sanborn, 1990)

We use the IV estimates obtained for each age group and the respective YPLL for each age group to recover the total YPLL. For each age group, we multiply the YPLL by the number of deceased individuals (from our estimates) to get the total YPLL for that age group. We then impose a dollar value on each year of potential life lost to obtain the total dollar value of years of life lost. Below we describe our process.

Column 1 of Data Appendix Table 1 reports the age groups we observe in the linked mortality - EMR data. The table differs in one small aspect from what we showed earlier in Table 1. We assumed that everyone who is above 65 is 65-74. Since the YPLL75 approach does not take into account years lost after 75; it equates to assuming no VSL was lost after 75. Column 2 reports the share of compliers in those groups. We care about the age distribution of the compliers and not the age distribution of the entire analysis sample because the compliers drive our IV estimates. Column 3 reports the mid age of each age group, we do this because we do not have patients' exact date of birth year, and so we assign the group the mid point within their age specific bracket. Column 4 is column 3 subtracted from 75 and it is the years of potential life lost if deceased. Column 5 is the regression 12 months IV estimates from our base model. Column 6 assumes we have 10,000 compliers and gives us the number of compliers per 10,000 visits cancelled. Example, since those age 0-17 are 11.1% of the compliers, there is 1110 of them in every 10,000 cancelled visits. Finally column 7 is produced using the following equation: YPLL per $\tilde{10000}$ Cancelled visits = $\frac{IV Estimate}{10000}$ × Number of compliers per 10,000 × YPLL per complier

From column 7, the total YPLL for the age 0-17 group for every 10,000 cancelled visits is 18.4 years. Similarly, for the age 65-74 group they lose 95 years for every 10,000 cancelled visits. Across all groups, per 10,000 cancelled visits the total YPLL is 443.54 years.

An obvious next step is dollarize these years to produce a VSL estimate. In economics, a vast literature has assessed the value of a statistical life (VSL). Across this literature there is not a single universally accepted number, instead ranges are often used. There are also different approaches to VSL and the three modal approaches often used are: an invariant population-average VSL; a constant value per statistical life-year (VSLY); and a VSL that follows an inverse-U pattern, peaking in middle age. In the inverse-U pattern VSL grows until the middle ages and then begins to decline (Robinson et al., 2021; Murphy and Topel, 2006). In the invariant population-average VSL approach researchers often assume VSL is \$7 to \$10 million (Nordhaus, 2018). We thought after a constant value per statistical life year (VSLY) in order to multiply each forgone life year by its assumed value. In the literature, estimates for VSLY range between 150,000 and 500,000 (Aldy and Viscusi, 2008; Cutler and McClellan, 2001). We chose the lowest value of 150,000- meaning we are being very conservative in defining the benefits of outpatient care.

Assuming a VSLY of 150,000 we find that for every 10,000 cancelled visits \$107,761,095 in VSL is lost. That is, the average visit has to cost no more than \$6,653.11 before its monetary costs outweigh its VSL monetary benefits.

This calculation above was done using our main estimate of 29.9 deaths per 10,000 cancelled visits. We ran several specifications (see figure 11). If we impose the smallest coefficient obtained of 10 deaths per 10,000 cancelled visits we find that the average visit has to cost \$2,773.63 before its monetary costs outweigh its VSL monetary benefits.

Data Appendix Table 1 : Value of Statistical Life Lost from Missed Care

Age	Share of compliers	Mid-Age	YPLL75 per complier	IV Estimate per 10,000	Number of compliers per 10,000 visits cancelled	Total YPLL per 10,000 visits cancelled
0-17	0.111	8.5	66.5	2.5	1110	18.45375
18-25	0.050	21.5	53.5	15	500	40.125
26-35	0.080	30.5	44.5	14.6	800	106.444
36-45	0.102	40.5	34.5	25.8	1020	90.7902
45-54	0.120	49.5	25.5	19.2	1200	58.752
55-64	0.178	59.5	15.5	12.3	1780	33.9357
65-74	0.360	69.5	5.5	48	3600	95.04
Total	1.00					443.54065

Notes - the figure shows how we use the YPLL75 to calculate YPLL for the compliers in our sample.