Ambulance Taxis *

Preliminary — Please Do Not Cite or Circulate

Paul Eliason[†], Riley League[‡], Jetson Leder-Luis[§],
Ryan C. McDevitt[¶], James W. Roberts[‖]

August 2021

We study the relative effectiveness of administrative regulations, criminal enforcement, and civil whistleblower lawsuits for combatting health care fraud. Between 2003 and 2017, Medicare spent $7.7 billion on regularly scheduled, non-emergency ambulance rides for patients traveling to and from dialysis facilities, with dozens of lawsuits alleging that Medicare reimbursed rides for patients who did not meet the requirements for receiving one. Using a novel data set and an identification strategy based on the staggered timing of regulations and lawsuits across the US, we find that a regulation requiring prior authorization for ambulance reimbursements reduced spending much more than criminal and civil lawsuits did. Despite the sharp drop in both ambulance transports and the companies that provide them following prior authorization, patients' health outcomes did not change, indicating that most rides were not medically necessary. Our results suggest that administrative actions have a much larger impact than targeted criminal enforcement, providing novel evidence that regulations may be more cost-effective than ex post ligation for preventing health care fraud.

*We thank Becky Staiger and participants of the ASHEcon 2021 Conference for their helpful feedback. We also thank Paul Kaufman for very helpful discussions about this industry. Evan Henley, Yicheng Liu, Mingrui Ma, Baiden Wright and Peck Yang provided excellent research assistance.

[†]Department of Economics, Brigham Young University
[‡]Department of Economics, Duke University
[§]Boston University
[¶]Duke University
[‖]Department of Economics, Duke University and NBER

1 Introduction

Fraud poses a serious problem for Medicare: it can both distort patient care and waste limited public resources. In 2019, improper payments made by Medicare — defined as "payments that did not meet statutory, regulatory, administrative, or other legally applicable requirements" — totaled $28.9 billion, or 7.3% of overall spending (CMS, 2020). In this paper, we assess the different approaches used to combat a particularly widespread and egregious type of fraudulent behavior, the unnecessary use of ambulances to transport patients between their homes and dialysis facilities, to better understand when regulations will be more effective than litigation at preventing Medicare fraud.

We study ambulance rides for dialysis patients because they provide an empirical setting well suited to an analysis of anti-fraud policy in health care. For those with kidney failure, dialysis replaces the life-sustaining function of kidneys by filtering wastes and toxins out of the blood, with approximately half a million patients in the US going to a dialysis facility three times per week for treatment. The frequent travel to and from facilities makes transportation a pressing concern for many of these patients, and although Medicare only reimburses transportation costs for those who demonstrate a medical need for assistance, unscrupulous ambulance companies often exploited a lax enforcement of the rules to provide fraudulent rides to ineligible patients, effectively serving as a very expensive taxi service. From 2003 to 2017, Medicare spent $7.7 billion on 37.5 million non-emergency ambulance rides for dialysis patients provided by over 3,000 firms across the US.

Although the billions of dollars at stake make a study of fraudulent ambulance rides worthwhile on its own, the relevance of our findings extends beyond the narrow setting of dialysis. This particular form of fraud represents a larger class of illicit activity in which providers receive payments for health care services that do not meet the regulatory standards for a valid reimbursement. Of the nearly $30 billion Medicare loses to improper payments each year, to say nothing of the losses in other federal and state programs, a lack of medical necessity has been a key factor in cases as varied as inpatient hospitalizations, physician-administered drugs, nursing homes, medical devices, and hospice care.

The US government uses an array of policies and mechanisms to prevent health care fraud. Both criminal and civil enforcement work through the court system, with the former potentially resulting in jail time and the latter imposing heavy penalties for those found guilty of fraud. Whereas the Federal Bureau of Investigation and Department of Justice expend substantial resources to investigate and litigate each of their cases individually, administrative regulations imposed by the Center for Medicare and Medicaid Services (CMS) require additional documentation, called prior authorization, from every provider seeking reimbursement.

For our empirical analysis, we combine Medicare dialysis claims data with a novel data set of

all criminal and civil enforcements of fraudulent ambulance rides to study the effects of litigation and regulation. Using the staggered rollout of Medicare's requirements for prior authorization as an identification strategy, we find that regulation is much more effective than litigation for reducing fraud: prior authorization caused an immediate and persistent drop in non-emergency ambulance rides of nearly 53%, whereas civil enforcement had a minimal effect and criminal enforcement resulted in only a gradual reduction in the upward trend in spending. We also show that prior authorization caused a substantial change in the ambulance market: the number of companies providing non-emergent rides fell 22% immediately following prior authorization, while those that remained became more specialized.

One possible tradeoff of reducing Medicare fraud through administrative procedures is that patients may find it more difficult to receive care. In this case, the sharp drop in ambulance rides following prior authorization could have made some patients more likely to miss dialysis sessions, increasing their risk of serious complications. Despite this possibility, we find that the regulatory change did not disrupt patients' regularly scheduled dialysis sessions or lead to worse downstream health outcomes, suggesting that prior authorization resulted in a more efficient use of Medicare's resources. We estimate that the federal government would have saved $4.8 billion if it had required prior authorization in 2003, when it first criminally prosecuted ambulance fraud, rather than waiting until 2014.

We conclude our paper by connecting our empirical results to prominent theories of enforcement and regulation to explain why litigation failed to reduce ambulance fraud while regulation succeeded. Most directly related are the models of Glaeser and Shleifer (2003) and Behrer et al. (2021) that consider the tradeoffs between regulation and litigation, though the idea that regulation may be a necessary complement to enforcement through the courts dates back at least a century (Wilson, 1913).

Building on these theoretical insights, a large empirical literature has established that criminal behavior responds to various types of enforcement, like increased policing (e.g., Levitt, 1997), with more-recent results also showing the importance of regulatory reforms for securing property rights (Behrer et al., 2021). This prior work notwithstanding, the relative effectiveness of regulatory, criminal, and civil enforcement remains an open empirical question. In our setting, the most relevant factors that shift the balance in favor of regulation include the reluctance of prosecutors to hold impoverished and seriously ill patients liable for fraud, the difficulty of recovering payments from fly-by-night firms, the diffuse nature of the harm, the need for specialization among regulators, and the "bright line rules" of prior authorization that make it easy to enforce.

Our empirical results also add to the literature on fraud and overbilling in Medicare. The seminal work of Silverman and Skinner (2004) and Dafny (2005) describes the incentives for hospitals to upcode inpatient care to receive larger reimbursements, with Fang and Gong (2017) continuing this thread by estimating the time intensity of outpatient procedures to identify providers

who bill for unrealistic hours. Similarly, Sanghavi et al. (2021) link emergency ambulance rides to hospital claims to identify "ghost rides" — rides that do not appear to be substantiated by a hospital visit — among all Medicare beneficiaries, estimating that they make up nearly 2% of ambulance transports nationwide; O'Malley et al. (2021) find in their recent work that home health care fraud diffuses faster in cities where firms have more patients in common. These studies have largely focused on the detection and incentives for fraud in specific contexts, however, which we extend by considering the mechanisms available to combat such illicit behavior.

Some recent evidence suggests that civil litigation by whistleblowers deters overbilling. Most notably, Howard and McCarthy (2021) show that whistleblowing deters the excessive use of implantable cardiac devices, while Leder-Luis (2019) finds that whistleblowing deters Medicare fraud in a series of case studies covering many different types of fraud and care. We complement this literature by considering the effect of criminal enforcement on Medicare fraud and the relative value of criminal versus civil enforcement in combatting overbilling.

In addition to the unnecessary ambulance rides we study in this paper, the dialysis industry has been subject to scrutiny for several other improper practices. As one example, Eliason et al. (2020) show that independent dialysis facilities acquired by large chains engage in behavior consistent with wasteful drug dumping and increase patients' doses of highly reimbursed drugs, practices detrimental to patients' health. The approach in Fang and Gong (2017) that uses hours worked to detect overbilling also shows that dialysis makes up a significant share of the physician services flagged as infeasible. This literature reflects the pervasive issue of overbilling in dialysis, although not all of it rises to the level of criminal fraud.

Our paper proceeds as follows. Section 2 describes the institutional details of dialysis and anti-fraud enforcement. Section 3 outlines the data used for our study. Section 4 presents our empirical analysis of the effects of prior authorization and litigation. Section 5 considers the effect of these enforcement actions on the industrial organization of the ambulance market. Section 6 shows the effects of prior authorization on patients' health outcomes. Section 7 documents the change in riders' characteristics following prior authorization. Section 8 places our empirical findings within the theoretical literature that considers the effectiveness of regulation and litigation. Section 9 concludes with our arguments for why regulatory actions are a cost-effective way to prevent health care fraud.

2 Background

Medicare's End Stage Renal Disease (ESRD) program covers patients needing dialysis, a procedure that filters the blood of those with kidney failure. Dialysis patients typically visit one of the nation's more than 7,000 dialysis facilities three times per week for three to four hours each session. Due to the frequent nature of these visits, patients spend a considerable amount of time

traveling to and from facilities. Many patients arrange for transportation on their own, either in a personal vehicle or on public transportation, but some with severe medical conditions require an ambulance. Medicare pays for transportation to and from dialysis only when an ambulance is medically necessary.

Ambulance companies must satisfy a number of requirements to receive Medicare reimbursements for dialysis transports. Federal regulations stipulate that ambulances must be staffed by at least two people, with at least one certified as an emergency medical technician (EMT), and the vehicles themselves must be specifically designed as ambulances.[1] To receive a reimbursement, participating providers first need a National Provider Identifier (NPI), and dialysis patients must be bedridden or need lifesaving procedures in transit for the ride to qualify as medically necessary.

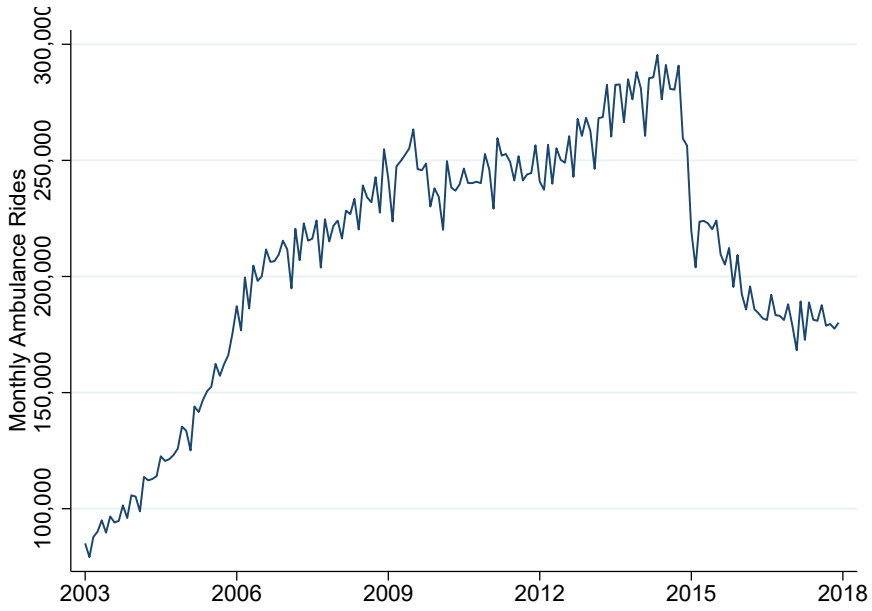
Medicare pays for ambulance rides through Part B, making patients responsible for a 20% copayment on top of their annual deductible. The payment rates set by Medicare consist of a base fee, which depends on the level of life support (e.g., whether the ride was emergent or in a helicopter) and a per-mile fee, for which ambulances receive a bonus if the pickup is in a rural location. Today, the base and mileage rates are $231.98 and $7.62, respectively, up from $209.65 and $6.74 in 2010. Medicare also adjusts rates by location, and at times they have been subject to bonuses if, for example, all of the mileage is in a rural area.

Fraud has become a major concern for Medicare's ambulance reimbursements as a whole, not just among dialysis patients. The Department of Health and Human Services Office of Inspector General (OIG) has published several reports about Medicare's ambulance benefit, concluding that it is often abused. For example, a 2006 OIG study, "Medicare Payment for Ambulance Transport," evaluated the appropriate use of the ambulance benefit and found that 20% of non-emergent transports were improper in that they did not meet Medicare's coverage requirements.

The issue is particularly acute in dialysis, where for many years ambulance companies transported patients who did not have a medical necessity under Medicare's criteria. The large reimbursements paid by Medicare provide a strong financial incentive for unscrupulous providers, especially if they transport non-emergent patients who do not require costly medical attention during the ride. And because dialysis patients must make regular visits to their facilities, they can be an especially lucrative target for those providing fraudulent rides. From 2007 to 2011, the volume of transports to and from dialysis facilities increased 20%, more than twice the rate of all other ambulance transports. In 2011, ambulance transports to and from dialysis facilities accounted for nearly $700 million in Medicare spending, or approximately 13% of Medicare's total expenditures on ambulance services (Centers for Medicare and Medicaid Services, 2020). Reflecting this, Figure 1 shows the initial growth and eventual decline of dialysis ambulance transports from 2003 to 2017.

[1]States may also impose their own regulations, such as the certificate of need laws currently in place in Arizona, Hawaii, Iowa, Kentucky, New Jersey, and New York. All states also license various levels of emergency medical service occupations and have different requirements for these licenses.

Figure 1: Non-Emergent Basic Life Support Dialysis Rides over Time



Notes: The sample includes non-emergent basic life support ambulance rides from a dialysis facility to a place of residence for ESRD patients from 2003–2017.

The US government has used several different approaches to prevent unnecessary ambulance rides for dialysis patients. Those who commit Medicare fraud can run afoul of criminal statutes, including the health care fraud statute (18 U.S.C. §1347) and the anti-kickback statute (42 U.S.C. §1320a-7a(a)(5)), with the crimes investigated by the Federal Bureau of Investigation and prosecuted by Department of Justice district offices nationwide. The US compounds its enforcement with laws against conspiracy, racketeering, organized crime, and lying to investigators. Beginning in 2003, the Department of Justice has pursued 38 criminal lawsuits against ambulance company operators who engaged in criminal fraud to provide dialysis ambulance transports. Along with knowingly billing the government for medically unnecessary care, allegations in these cases include paying kickbacks to patients to induce them to ride, paying referral bonuses to patients who recruited others to participate in the scheme, and concealing or manipulating documentation to justify the ongoing use of ambulances.

In addition to criminal statutes, federal health care fraud violates the False Claims Act, a civil statute that imposes monetary penalties of triple damages on firms that overbill federal health care programs. The False Claims Act contains a qui tam whistleblower provision, wherein individuals with knowledge and evidence of fraud can file their own lawsuits on behalf of the US government against those who bill fraudulently, in exchange for 15–30% of the recovered funds. The Department of Justice can also bring civil lawsuits against fraudsters directly. We identify

19 civil lawsuits, initiated as early as 1996, alleging the unnecessary transport of dialysis patients by ambulance companies.

Medicare administrators also thwart overbilling and fraud by making regulatory changes. Beginning in 2014, Medicare imposed prior authorization requirements through Medicare Administrative Contractors (MACs), the companies that process Medicare claims, stipulating that they would not pay a claim for a non-emergency dialysis ambulance ride without first documenting a medical necessity. Providers could receive authorization before the ride by submitting documentation or filing a claim for rides already completed, but they would not be paid without documentation. In 2014, Medicare first rolled out prior authorization in New Jersey, South Carolina, and Pennsylvania — states that had some of the heaviest use of dialysis ambulance rides — and extended the regulation in 2016 to include Delaware, DC, Maryland, North Carolina, Virginia, and West Virginia. Plans to expand prior authorization nationwide were postponed in 2020 due to the Covid-19 pandemic.

3 Data & Descriptive Statistics

We use a 100% sample of claims data for the entire universe of patients diagnosed with ESRD and enrolled in Medicare between 2003 and 2017. These data consist of patient- and facility-level information compiled by the United States Renal Data System (USRDS).[2] The patient-level data allow us to observe demographics (e.g., sex, race, BMI, cause of ESRD, payer, comorbidities, ZIP Code, and a facility identifier) and complete ESRD treatment histories, while facility-level data have information on location and ownership. Importantly, our data allow us to observe each ambulance trip to and from a dialysis facility billed to Medicare. For firms that provide non-emergency ambulance rides, we also have data on their other claims for Medicare ESRD beneficiaries, such as emergency hospital transports.[3]

Table 1 provides summary statistics for patient characteristics, ridership, and health outcomes for those who receive any non-emergent ride to a dialysis facility as well as ESRD patients who never receive such a ride, split across months with and without rides. Riders are older, more likely to be women, more likely to be Black, and more likely to have diabetes. Patients who use ambulances for non-emergency transportation to dialysis facilities take on average 20 rides each month, with a lifetime average of 561 rides. Given that dialysis patients receive roughly 12 treatments per month, these statistics imply that those using ambulances receive a ride to and from their facility for nearly nine out of ten sessions.

[2]USRDS combines data from a variety of sources, including Medicare claims, annual facility surveys, and dialysis treatment histories, to create the most comprehensive data set for studying the US dialysis industry. For a more thorough description of USRDS, please see the *Researcher's Guide to the USRDS System* at USRDS.org.

[3]USRDS began recording identifiers for ambulance companies in 2012, so our firm-level analyses use data from 2012 to 2017.

Table 1: Summary Statistics of Patient-Month Level Data

	Patient Rider Status			
	Never-Rider	Rider, Non-Riding Month	Rider, Riding Month	Overall
Patient Characteristics				
Age (Years)	62.06	67.46	69.30	63.03
Months with ESRD	56.32	57.20	53.89	56.32
Black	0.376	0.417	0.451	0.385
Male	0.560	0.496	0.457	0.548
Diabetic	0.525	0.626	0.661	0.543
Drug User	0.014	0.011	0.008	0.013
Smoker	0.065	0.056	0.045	0.063
Drinker	0.014	0.013	0.011	0.013
Uninsured at Incidence	0.128	0.086	0.061	0.120
Employed at Incidence	0.180	0.097	0.066	0.165
Ridership				
Non-Emergent Dialysis Rides	0.00	0.00	19.53	0.87
Emergent Rides	0.101	0.183	0.408	0.125
Total Lifetime Rides	0.0	116.1	560.7	39.0
Continuing to Ride Next Month	.	.	0.838	0.838
Health Outcomes				
Dialysis Sessions	12.20	12.03	11.28	12.14
All-Cause Hosp.	0.111	0.154	0.250	0.122
Fluid Hosp.	0.011	0.016	0.020	0.012
Mortality	0.009	0.006	0.034	0.010
Patient-Months	16,054,292	2,317,247	857,390	19,228,929

Notes: Data are from 2011–2017. Patient characteristics except age and dialysis tenure are at incidence of ESRD. All ridership variables other than emergent rides are based on non-emergent basic life support rides between a dialysis facility and a patient's home. The probability of continuing to ride is the conditional probability of riding in the next month given the patient rides in this month. Fluid hospitalizations are those for which the primary diagnosis indicates excess fluids, an indication of insufficient dialysis.

We supplement these data with information from the criminal and civil enforcement of fraud. Using publicly available press releases from the Department of Justice, corroborated for completeness by internet searches, we identify 56 lawsuits in 23 different judicial districts against 98 ambulance companies or individuals for unnecessary ambulance transports related to dialysis. For each of these lawsuits, we collect court records from the Public Access to Court Electronic Records (PACER) system, which include specific fraud allegations and data on the lawsuit's timing and location of enforcement. For context, Table 2 provides descriptive statistics for different types of ambulance rides, broken down by status (i.e., emergent or non-emergent) and type of firm (e.g., indicted or non-emergent only).

4 Empirical Analysis

The data described above allow us to study the effectiveness of regulations relative to litigation in combating ambulance fraud. As discussed in Section 2, Medicare regulations requiring prior authorization stipulate that ambulance companies obtain approval for each patient receiving

Table 2: Summary Statistics for Selected Ride Types

	Total Rides	Total Payments	Firms Involved*
All Non-Emergent Dialysis Rides	37,501,752	$7,733,452,800	3,081
Rides by Non-Emergent-Only Firms*	921,419	$190,515,568	262
Rides by Indicted Firms*	642,032	$139,586,464	42
All Emergent Rides	5,986,533	$2,082,876,032	10,532

Notes: Unless explicitly identified as an emergent ride, rides are non-emergent basic life support rides between a dialysis facility and a patient's home observed in the USRDS data. These data include rides from 2003–2017. Non-emergent-only firms are those that are never observed giving an emergency ambulance ride in the USRDS data.

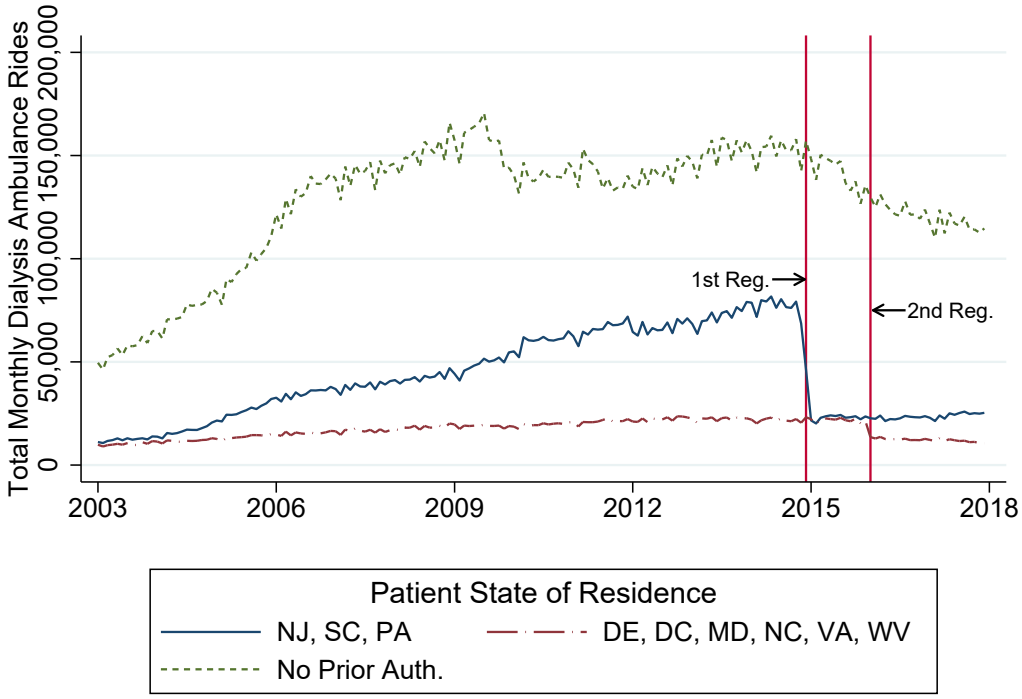
*Firm identifiers are available from 2012–2017, and figures reported in this row or column use only data from this period.

repetitive, non-emergent ambulance transports, which must be renewed periodically.[4] The policy was piloted on December 15, 2014, in New Jersey, Pennsylvania, and South Carolina, the top three states in per-beneficiary dialysis ambulance expenditures for Medicare at the time. On January 1, 2016, Medicare expanded prior authorization to Delaware, DC, Maryland, North Carolina, Virginia, and West Virginia. As shown in Figure 2, rides for patients residing in Pennsylvania, New Jersey, and South Carolina drop sharply after Medicare first imposed prior authorization, followed by another drop corresponding to the states included in the second wave of regulation.

Authorities also use legal actions like criminal enforcement and civil lawsuits to deter fraud, and a case study of the Pennsylvania East District helps motivate our research strategy for identifying the effects of such litigation. Figure 3 shows the growth in Medicare spending on non-emergent dialysis rides between 2003 and 2010 in this district. Following the pronounced spike in rides, authorities brought multiple cases against suspected fraudulent individuals and firms. The left-most vertical red line marks the indictment date of the first criminal case, and the subsequent drop in payments suggests that criminal prosecution reduced the number and cost of ambulance rides in the district. The sharp and immediate drop in rides following prior authorization, however, implies that regulation may have an even stronger effect than litigation. Below, our empirical analysis will use variation in the timing of criminal indictments and civil cases brought in each district to test whether the trend in ambulance expenditures systematically changed following these proceedings and whether they were as effective as prior authorization.

[4]Medicare considers "three or more round trips during a ten-day period, or at least one round trip per week for at least three weeks" to be repetitive transports. Prior authorization is required for the fourth ride in a 30-day period.

Figure 2: Rides by Prior Authorization Regulation



Notes: The sample includes non-emergent basic life support ambulance rides from a dialysis facility to a place of residence for dialysis patients from 2003–2017. State is determined by the transported patient's residence. First vertical line marks the implementation of prior authorization in NJ, SC, and PA, and the second marks DE, DC, MD, NC, VA, and WV.

4.1 Methodology

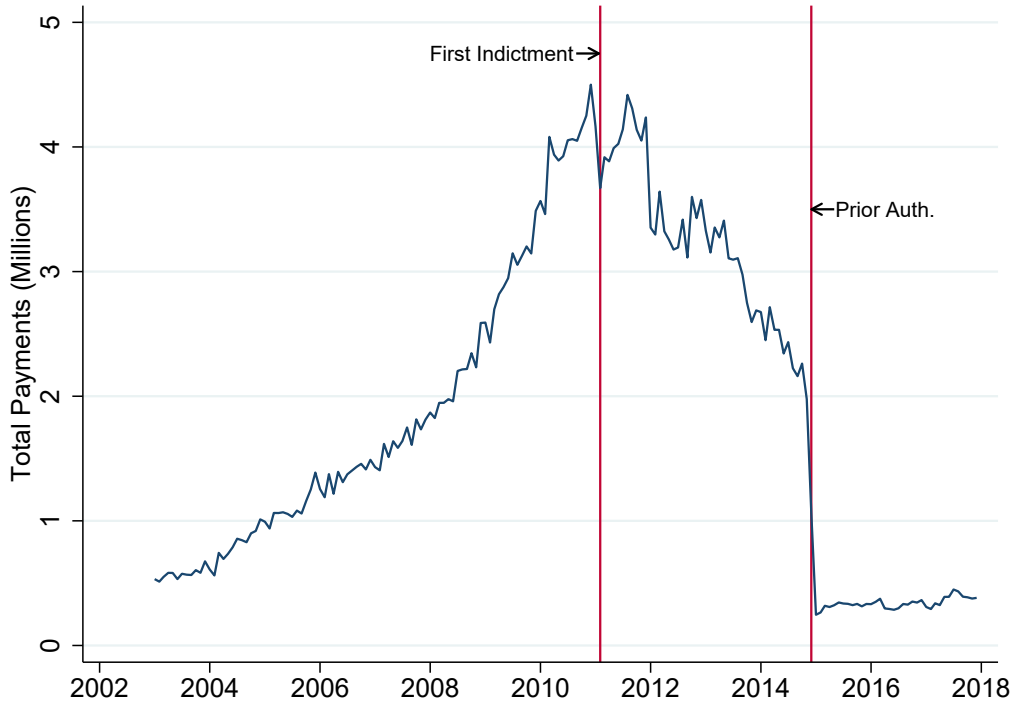
We use the staggered roll out of prior authorization and the differential timing of criminal and civil enforcement across US federal judicial districts to identify the causal effects of these respective approaches for reducing fraudulent rides.[5] For our estimations, we present results using both traditional two-way fixed effects (TWFE) methods as well as those introduced by Callaway and Sant'Anna (2020), which we call "CS" methods. For the traditional TWFE results, we estimate

$$Y_{dt} = \beta Enf_{dt} + \alpha_d + \alpha_t + \varepsilon_{dt} \tag{1}$$

for district d in month t, where Enf_{dt} is an indicator variable that turns on after the district is first subject to enforcement and α_d and α_t are a series of fixed effects for districts and year-month,

[5]There are 94 US federal judicial districts, each of which are wholly contained within states; these are the regions at which the Department of Justice and the US federal court operate, each with its own US attorney and Department of Justice office. We provide a map of these districts in Appendix A.

Figure 3: Total Payment for Ambulance Rides in Pennsylvania East District



Notes: The sample includes non-emergent basic life support ambulance rides from a dialysis facility to a place of residence for dialysis patients from 2003–2017 for patients whose county of residence is within the Pennsylvania East judicial district. The first vertical line marks the first criminal or civil indictment of an ambulance firm in this district and the second vertical line marks the implementation of prior authorization in Pennsylvania.

respectively.[6] For results estimated at the patient level, our estimating equation is

$$Y_{idt} = \beta Enf_{dt} + \alpha_d + \alpha_t + \Gamma X_{idt} + \varepsilon_{idt} \tag{2}$$

for individual i with observable patient and dialysis facility characteristics X_{idt}, which also includes dialysis facility fixed effects.

To assess the dynamics of the treatment effect, we estimate

$$Y_{dt} = \sum_{e=-K}^{-2} \beta_e Enf_{dt} + \sum_{e=0}^{L} \beta_e Enf_{dt} + \alpha_d + \alpha_t + \varepsilon_{dt}, \tag{3}$$

where e is the number of months since the first month of treatment and K and L are positive

[6]We define Enf_{dt} to be equal to one following the first lawsuit in the district, regardless of the number of subsequent lawsuits.

constants. Similarly, for results estimated at the patient level, our estimating equation is

$$Y_{idt} = \sum_{e=-K}^{-2} \beta_e Enf_{dt} + \sum_{e=0}^{L} \beta_e Enf_{dt} + \alpha_d + \alpha_t + \Gamma X_{idt} + \varepsilon_{idt}. \tag{4}$$

In settings that have heterogeneous treatment effects along different dimensions, traditional TWFE models may not recover the average effect of treatment on the treated (ATT).[7] To overcome this issue, we use the methods introduced by Callaway and Sant'Anna (2020) to estimate the effect of treatment separately for each group of districts treated at the same time, using only districts that are never treated as the control group. That is, we estimate (3) for each group of districts treated at the same time and those districts that never receive treatment separately for each group. Under weak assumptions, the CS method recovers the average treatment effect at time t for the group of districts treated at time g, which we refer to as $ATT(g,t)$. To simplify the interpretation of our results, we aggregate the $ATT(g,t)$ of each treatment group across time to obtain a treatment-group specific parameter analogous to the β recovered using traditional TWFE methods. The parameter

$$\theta_{sel}(\tilde{g}) = \frac{1}{\mathcal{T} - \tilde{g} + 1} \sum_{t=\tilde{g}}^{\mathcal{T}} ATT(\tilde{g}, t) \tag{5}$$

gives the average treatment effect on districts treated at time $\tilde{g}$ from the first month in which they are treated until the last month in our data, $\mathcal{T}$.

Because we want to analyze the dynamic effects of treatment parsimoniously even though few districts are treated at any given time, we also aggregate our results across groups to recover the effect of treatment after $e = t - g$ months of exposure to treatment. And because districts are treated at different times, some treatment groups are treated later in our sample period than others, which means we must aggregate the results across groups to account for any compositional changes in treated units at different lengths of exposure. To do this, we only aggregate $ATT(g,t)$ for groups that are treated for at least e' months and recover the average treatment effect for treatment of length e on districts that are treated for at least e' periods,

$$\theta_{es}^{bal}(e; e') = \sum_{g \in \mathcal{G}} \mathbf{1}\{g + e' \leq \mathcal{T}\} ATT(g, g+e) P(G = g | G + e' \leq \mathcal{T}), \tag{6}$$

where $\mathcal{G}$ gives the set of treatment times and $\mathcal{T}$ is the last month in our data.

Finally, to compare the estimates using traditional TWFE and CS methods, we further aggregate $ATT(g,t)$ into a single parameter that gives the average treatment effect for the first

[7] See, for example, Borusyak and Jaravel (2017), de Chaisemartin and D'Haultfuille (2020), Goodman-Bacon (2019), Sun and Abraham (2020), and Athey and Imbens (2018).

e' months of treatment in districts treated for at least e' months. This parameter is given by

$$\theta_{es}^{O,bal}(e') = \frac{1}{e'+1}\sum_{e=0}^{e'}\theta_{es}^{bal}(e,e'), \tag{7}$$

which is simply the unweighted average of the parameters given by (6) across the first e' months of treatment.

4.2 The Effect of Prior Authorization on Rides

We first consider the effect of prior authorization. Table 3 gives the results of estimating (1) with various sets of fixed effects in columns (1)–(3). In column (4), we present the estimate of $\theta_{es}^{O,bal}(23)$, the effect of prior authorization on all treated districts in the two years following treatment, where the outcome is the total payments for non-emergent ambulance rides between a dialysis facility and a patient's home, measured in logs.[8] We find that prior authorization reduces payments for non-emergent ambulance rides by 0.753 log points, or 52.9%.

Table 3: Effect of Prior Auth. on Ambulance Spending

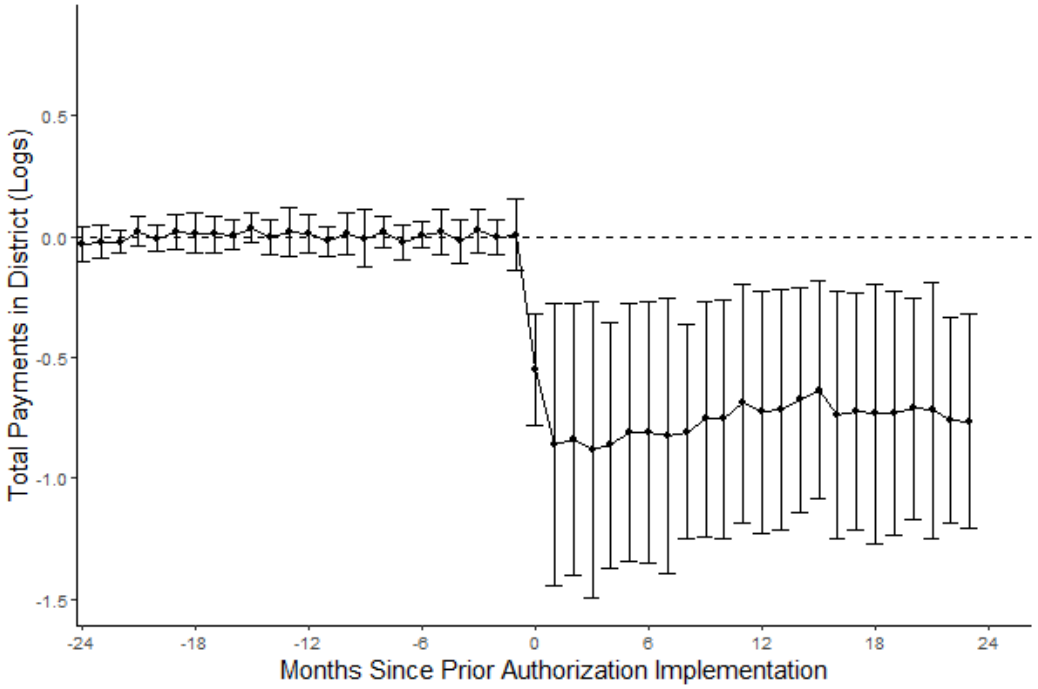
	(1) Total Payments (Log)	(2) Total Payments (Log)	(3) Total Payments (Log)	(4) Total Payments (Log)
Prior Auth.	0.691***	0.920***	-0.724***	-0.753***
	(0.0862)	(0.0969)	(0.0397)	(0.142)
Year-Month FE	0	1	1	1
District FE	0	0	1	1
CS Methods	0	0	0	1
Dep. Var. Mean	11.47	11.47	11.47	11.47
Observations	14507	14507	14507	14507

Notes: All rides are non-emergent basic life support rides between a dialysis facility and a patient's home observed in the USRDS data. These data include rides from 2011–2017. An observation is a district-month. Columns (1)–(3) give estimates of β from equation (1) using traditional TWFE methods. Column (4) gives the estimate of $\theta_{es}^{O,bal}(23)$ using CS methods. For Columns (1)–(3), standard errors are clustered at the district level. For column (4), standard errors are obtained using CS's bootstrap-based procedure. $^{+}$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

Figure 4 shows the dynamic difference-in-differences results, which compare the difference in the level of spending in districts exposed to prior authorization with those that have never been exposed for the two years before and after the regulation was put in place. In particular, the figure presents estimates of $\theta_{es}^{bal}(e;23)$ for $e \in [-24, 23]$. We find that the effect of prior authorization was large, immediate, and persistent.

[8]The second wave of prior authorization occurs two years before the end of our data, meaning that both waves of treatment are included in this parameter.

Figure 4: Effect of Prior Auth. on Ambulance Spending



Because the parameter estimates presented in Figure 4 and column (4) of Table 3 are aggregations across the two waves of prior authorization, we also present the overall treatment effect for the districts in each wave individually. That is, we present $\theta_{sel}(\tilde{g})$ for the two values of $\tilde{g}$ corresponding to the first and second waves of prior authorization. The point estimates for $\theta_{sel}(\tilde{g})$ are -1.196 ($s.e.$ 0.305) for the first wave of prior authorization and -0.498 ($s.e.$ 0.108) for the second, indicating statistically significant drops of 69.8% and 39.2%, respectively. As such, the effect of prior authorization on payments in New Jersey, Pennsylvania, and South Carolina — the states first exposed to prior authorization in December 2015 — was much larger than the effect in the second wave of states exposed 13 months later.[9]

4.3 The Effect of Litigation on Rides

To study whether litigation reduces fraudulent ambulance rides, we use the same CS methodology as in Section 4.2 to estimate the impact of civil and criminal enforcement actions separately. Table 4 provides estimates of $\theta_{es}^{O,bal}(23)$, where the treatment date g is determined by the first enforcement action of each type in the district.[10]

[9] In Appendix B, we perform a similar analysis at the firm-month and patient-month level, finding that the large effect of prior authorization is robust. We also consider a falsification test that shows prior authorization had no impact on the number of emergent rides.

[10] Because Illinois North, Massachusetts, and California Central had civil actions before our sample period and the first civil action in Virginia East was too late in our data, we exclude these districts from our analysis of the effect of civil enforcement. Kentucky East is excluded from our analysis of criminal enforcement for being subject to enforcement too late in our data.

We find that civil enforcement does not have a statistically significant effect on rides or total payments, whereas criminal enforcement reduces monthly payments by 9% and rides by 10% in the two years following enforcement. Figure 6 shows the dynamic effects of the first criminal indictment, followed by the heterogeneous effects of each wave of indictments. The dynamic effects in panel (a) suggest that criminal enforcement reduces payments gradually, inducing a downward trend without an immediate drop. This could mean that the effect of enforcement grows over time as information about the penalties for fraudulent behavior disseminates, or it could indicate a more cautious strategy by fraudsters that results in a gradual slowdown of spending. Panel (b) suggests a degree of heterogeneity in the effect of these enforcement actions, with some having large effects and others having none at all.

Table 4: Effect of Litigation on Ambulance Spending

	(1) Total Ride Payments (Log)	(2) Total Rides (Log)	(3) Total Ride Payments (Log)	(4) Total Rides (Log)
Enforcement	0.070	0.069	-0.095*	-0.106**
	(0.115)	(0.117)	(0.0475)	(0.0488)
Type of Enforcement	Civil	Civil	Criminal	Criminal
Year-Month FE	1	1	1	1
District FE	1	1	1	1
CS Methods	1	1	1	1

Notes: All rides are non-emergent basic life support rides between a dialysis facility and a patient's home observed in the USRDS data. These data include rides from 2011–2017. An observation is a district-month. Columns (1) and (2) give the estimate of $\theta_{es}^{O,bal}(23)$ using CS methods for civil enforcement, while columns (3) and (4) do the same for criminal enforcement. The treatment date is the earliest enforcement action of the relevant type in the district. Standard errors are obtained using CS's bootstrap-based procedure. $^{+}$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

Taken together, our results show that prior authorization was much more effective than litigation at deterring fraud. Prior authorization caused a large and immediate drop in non-emergent ambulance rides that persisted over time, whereas criminal enforcement had only about one-sixth the effect and civil action had no impact whatsoever.

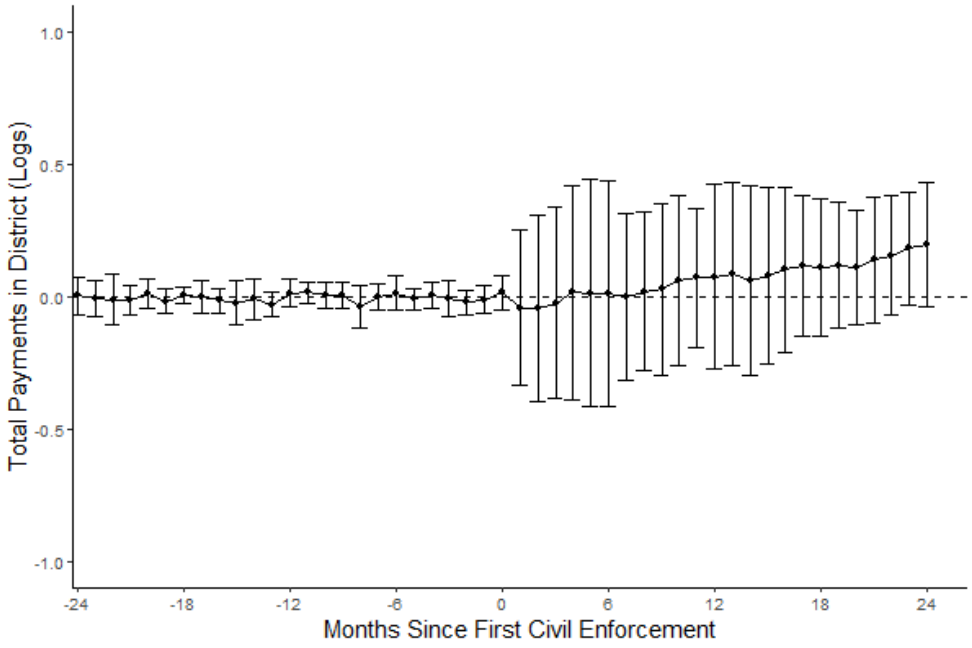
5 The Effect of Enforcement on Market Structure

Not only did prior authorization cause a large drop in the number of non-emergent ambulance rides to dialysis facilities, it also led to a large reduction in the number of firms that provide them. As shown in column (4) of Table 5, prior authorization reduced the number of ambulance companies providing non-emergent dialysis rides by 22.1%.

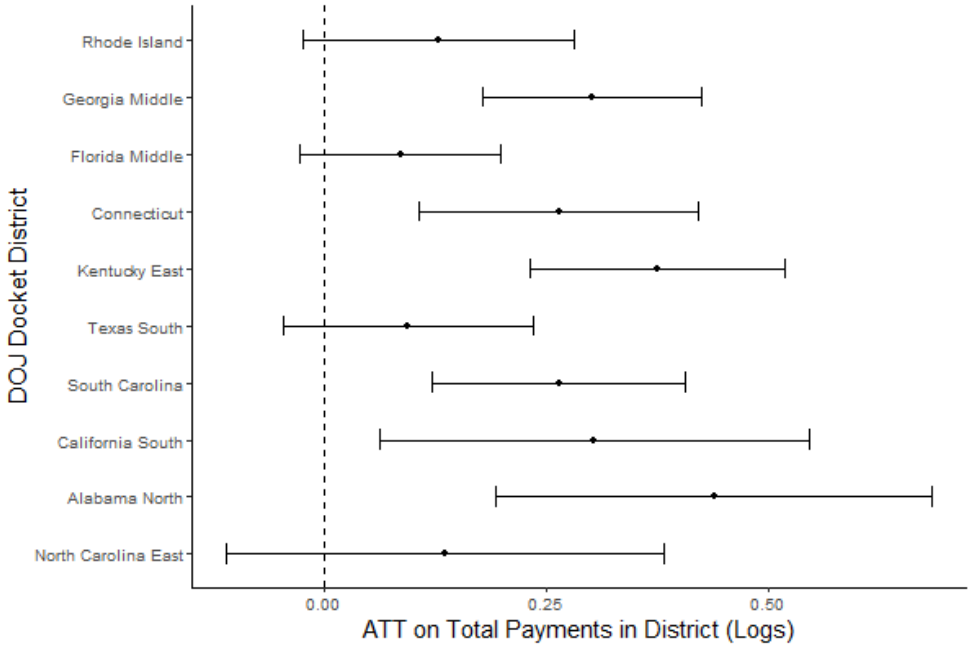
Beyond simply reducing the number of ambulance companies, we find that prior authorization also leads to greater firm specialization: firms with a high share of non-emergent rides are more

Figure 5: The Impact of Civil Cases on Ambulance Payments

(a) Dynamics



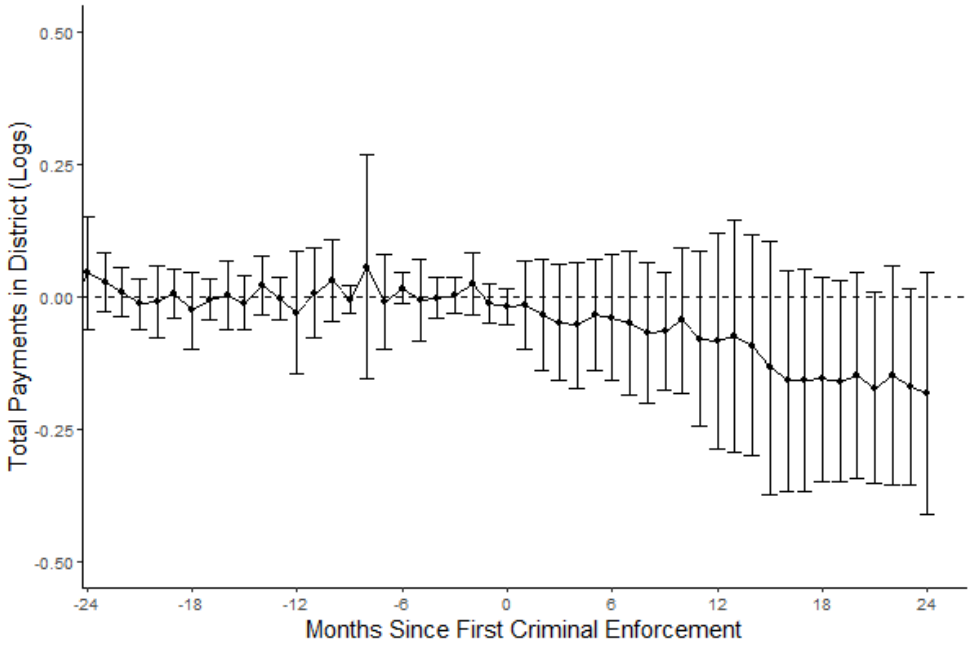
(b) Heterogeneous Effects



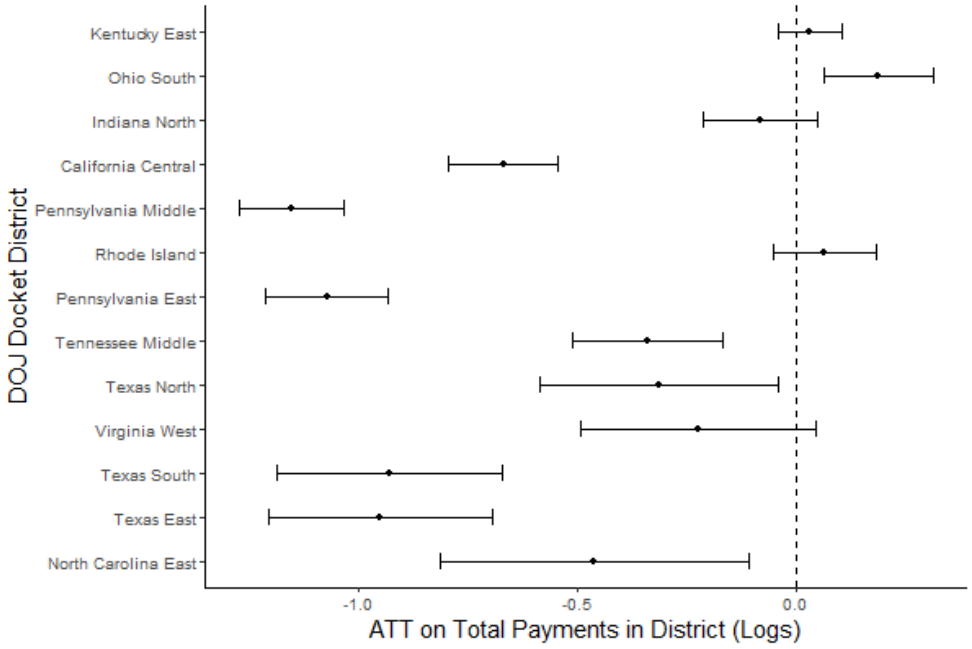
likely to exit at the first wave of prior authorization, while the number of firms providing only non-emergent dialysis rides increases. We can see this in Figure 8, which gives the distribution of firms by the share of non-emergent rides they provide to dialysis patients. Many of the firms that provide non-emergent ambulance rides to dialysis patients provide very few emergent rides to the same population, especially before prior authorization. After prior authorization, fewer

Figure 6: The Impact of Criminal Cases on Ambulance Payments

(a) Dynamics



(b) Heterogeneous Effects



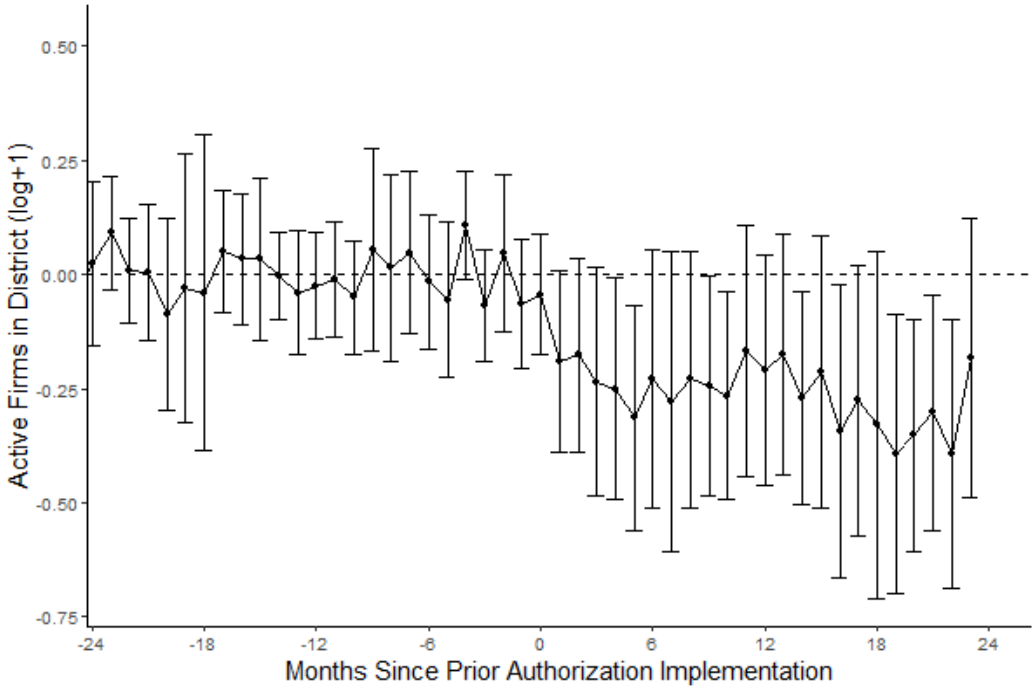
firms provide rides to dialysis patients, particularly firms that provide very few non-emergent rides, but the number of firms that provide *only* non-emergent rides to dialysis patients nearly doubled, increasing from 59 to 102. In other words, prior authorization seems to have split the market in two: some firms provide mainly emergent rides that do not require prior authorization while others successfully navigate the bureaucracy of prior authorization to provide mainly non-

Table 5: Effect of Prior Authorization on Number of Active Firms

	(1) Active Firms (Log+1)	(2) Active Firms (Log+1)	(3) Active Firms (Log+1)	(4) Active Firms (Log+1)
Prior Auth.	0.759***	0.917***	-0.277***	-0.2492**
	(0.0544)	(0.0562)	(0.0193)	(0.0859)
Year-Month FE	0	1	1	1
District FE	0	0	1	1
CS Methods	0	0	0	1
Dep. Var. Mean	1.935	1.935	1.935	1.935
Observations	6768	6768	6768	6768

Notes: Dependent variable is the natural logarithm of one plus the number of firms providing non-emergent basic life support rides between a dialysis facility and a patient's home in a district-month. These data include rides from 2012–2017. An observation is a district-month. Columns (1)–(3) give estimates of β from equation (1) using traditional TWFE methods. Column (4) gives the estimate of $\theta_{es}^{O,bal}(23)$ using CS methods. For Columns (1)–(3), standard errors are clustered at the district level. For column (4), standard errors are obtained using CS's bootstrap-based procedure. $^{+}$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

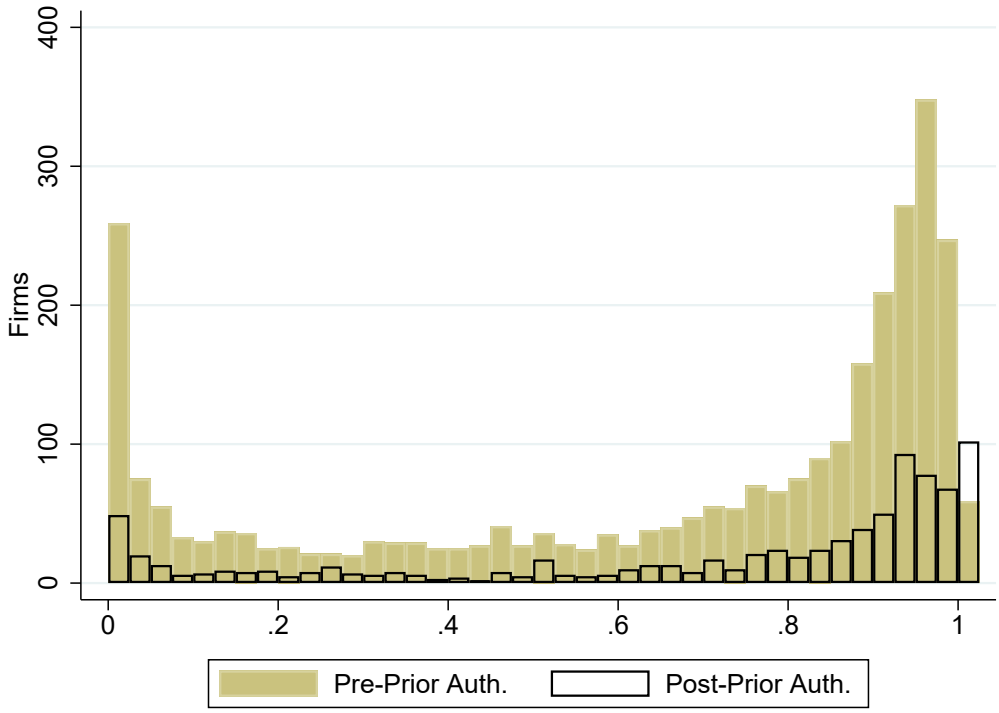
Figure 7: Effect of Prior Authorization on Number of Active Firms



emergent ones.

In contrast to the large impact of prior authorization on the market for non-emergent ambulance rides, we find little evidence that criminal and civil enforcement had any noticeable impact. Table 6 provides one possible explanation as to why: when a firm exits, roughly half of its patients continue to ride in the next month, with this number closer to 60% when the exiting firm was indicted. Among those who stop riding, almost all of them continue to receive dialysis, with only 2% of patients who ride with an exiting firm failing to make it to their sessions in the month after the exit, notwithstanding patients who were hospitalized or died. The fact that

Figure 8: Distribution of Firms by Share of Rides that are Non-Emergent



patients who received rides from indicted firms are more likely than not to continue receiving ambulance rides indicates that litigation did not do much to change their behavior.

Table 6: Summary Statistics for Riders of Exiting Firms

	Status of Exiting Firm	
	Not Indicted	Indicted
Continues Riding	0.511	0.569
Is Treated without Riding	0.369	0.391
Dies This Month	0.060	0.013
Is Hospitalized This Month	0.038	0.007
Is Not Treated Next Month	0.022	0.020
Observations	4481	151

6 Prior Authorization's Effect on Patient Health

Prior authorization reduced the number of non-emergent ambulance rides taken by dialysis patients. Although this regulation caused a sharp decline in potentially fraudulent payments, the additional administrative burden may have resulted in some patients forgoing treatment if they could not find alternative transportation. If these missed sessions resulted in adverse events like hospitalization or death, Medicare's savings from fewer ambulance reimbursements could have been offset by higher costs in other parts of the ESRD program as well as a lower quality of life for the affected patients.

To assess the impact of prior authorization on health outcomes, we estimate equation (2) at the patient-month level, with measures of patients' health as our outcome variables. We control for a rich set of patient and facility characteristics, including facility fixed effects, while clustering our standard errors at the district level. Because the CS methodology is not well suited for estimates at the patient-month level, we use traditional TWFE methods in this section but show in the appendix that CS produces similar results.[11]

Table 7 presents the effects of prior authorization on patient adherence to dialysis, as well as downstream health outcomes like hospitalizations and mortality. We find no evidence that prior authorization led to either decreases in dialysis sessions or increases in adverse events.

Although we find that prior authorization did not harm patients' health on average, it could be that some patients were harmed in ways not captured by our point estimates. To consider this possibility, we restrict our sample to the group of patients most likely to be affected by the policy change: those who relied most heavily on ambulance rides prior to the reform. Specifically, we restrict our sample to patients who took at least 100 non-emergent ambulance rides to dialysis facilities before prior authorization and compare the outcomes of these frequent riders across districts throughout the staggered rollout of prior authorization. Table 8 shows that even for the most-frequent riders, we find no evidence that prior authorization resulted in worse health outcomes.

Another potentially unintended consequence of prior authorization is that some patients who satisfy Medicare's criteria for a reimbursable ride might not receive one if their ambulance company goes out of business. To assess this possibility, Figure 9 shows what happens to riders in the month after their ambulance company exits the market. Compared to patients whose ambulance company exited prior to the first wave of prior authorization, an exit not induced by anti-fraud regulation, those who rode with companies that exited during the first month of

[11]The CS methodology is ill suited for patient-month-level estimation for a few reasons. First, it does not allow for time-varying controls, which may be important for assessing patients' health. Second, using the CS methodology for unbalanced panels (which a panel of patients over such a long period would necessarily be) is computationally intensive, especially given the high number of patients in our data. To avoid these issues, we estimate equations (1) and (4) using the CS methodology on a district-level panel of district-wide average patient health outcomes as a robustness check.

Table 7: Effect of Prior Auth. on Adherence and Adverse Events

	(1) Dialysis Sessions	(2) Mortality	(3) All-Cause Hosp.	(4) Insufficient Dialysis Hosp.
Prior Authorization	0.0180*	0.0000903	-0.00379***	-0.00103***
	(0.00790)	(0.000141)	(0.000774)	(0.000215)
Year-Month FE	1	1	1	1
District FE	1	1	1	1
Pat/Fac Controls	1	1	1	1
Facility FE	1	1	1	1
R-squared	0.0120	0.00494	0.0158	0.00613
Dep. Var. Mean	12.12	0.00994	0.123	0.0117
Observations	17548288	17548288	17548288	17548288

Notes: Table gives estimates of β from equation (2) using traditional TWFE methods at the patient-month level. Controls include incident patient characteristics, age, and tenure on dialysis as well as facility characteristics including chain ownership status, demographic characteristics of the ZIP code, and whether the facility is freestanding or hospital-based. Insufficient dialysis hospitalizations are those for which the primary diagnosis indicates excess fluids. Standard errors clustered at the district level are given in parentheses. $^{+}$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

prior authorization were not less likely to receive treatment even though they were much less likely to continue riding. That is, patients riding with ambulance companies that exited following prior authorization did not miss more sessions than a typical patient whose ambulance company exited before the start of prior authorization. Taken together, these results suggest that prior authorization for non-emergent ambulance rides did not adversely affect patients' health: patients continue to receive treatment at the same rate as before and have no uptick in hospitalizations or mortality.

7 Non-emergent Rides and Prior Authorization

Although prior authorization greatly reduced the number of non-emergent dialysis rides, many patients continue to receive them despite the more-stringent regulations. Several stylized facts about these riders suggest that the regulation had its intended effect of ensuring that the patients who ride in ambulances are the ones who truly need to do so. First, column (1) of Table 9 shows that the probability that a current rider continues riding the following month falls after prior authorization, indicating that ridership is less persistent. Next, comparing the two years before prior authorization to the two years after, Figure 10 shows that the total number of rides taken by each rider decreased substantially. Finally, we find that the median number of months in which a rider takes a non-emergent ride falls from six months to three.

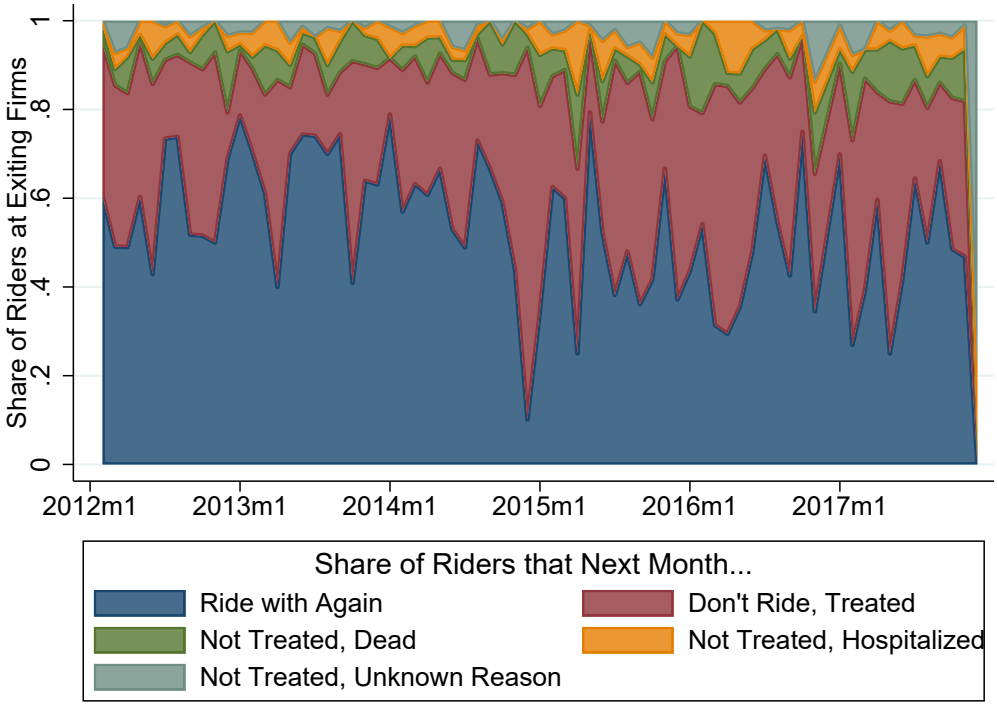
As further evidence that prior authorization resulted in a more-appropriate mix of patients taking ambulance rides, we note that patients who took many rides before prior authorization

Table 8: Effect of Prior Auth. on Frequent Riders

	(1) Dialysis Sessions	(2) Mortality	(3) All-Cause Hosp.	(4) Insufficient Dialysis Hosp.
Prior Authorization	0.121***	-0.00119*	-0.0271***	-0.00327***
	(0.0225)	(0.000593)	(0.00286)	(0.000778)
Year-Month FE	1	1	1	1
District FE	1	1	1	1
Pat/Fac Controls	1	1	1	1
Facility FE	1	1	1	1
R-squared	0.0698	0.00938	0.0255	0.0134
Dep. Var. Mean	11.91	0.0114	0.178	0.0158
Observations	1159603	1159603	1159603	1159603

Notes: Table gives estimates of β from equation (2) using traditional TWFE methods at the patient-month level. Controls include incident patient characteristics, age, and tenure on dialysis as well as facility characteristics including chain ownership status, demographic characteristics of the ZIP code, and whether the facility is freestanding or hospital-based. Insufficient dialysis hospitalizations are those for which the primary diagnosis indicates excess fluids. The sample is limited to patients that took at least 100 non-emergent ambulance rides to dialysis under the non-prior authorization regime. Standard errors clustered at the district level are given in parentheses. $^{+}$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

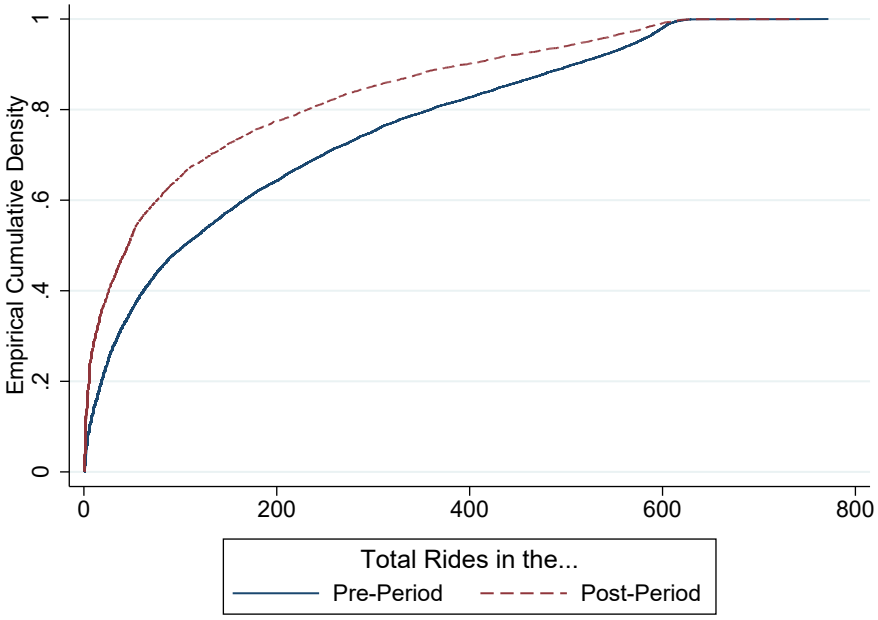
Figure 9: Results of Firm Exit for Riders at Exiting Firms



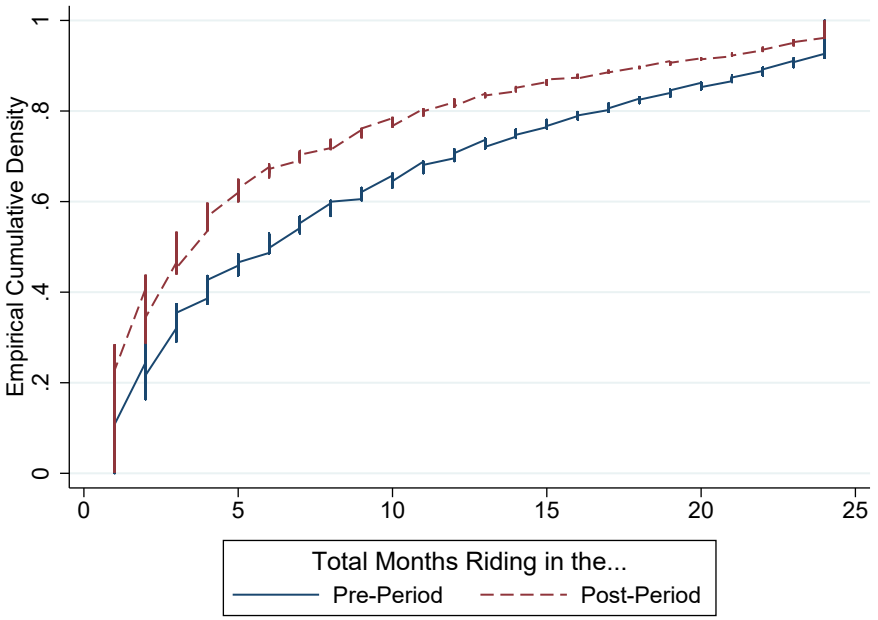
were more likely to continue riding after the policy change. Specifically, we show in Table 9 that, conditional on riding, the total number of rides taken over the life of the patient increased

Figure 10: Empirical CDF of Ridership Among Riders

(a) Total Rides Taken



(b) Months Taking At Least One Ride



following prior authorization. This change occurred suddenly, as shown in Figure 11. Similarly, we find that the probability of suffering an adverse event during the same month a ride is taken — likely reflecting a legitimate need for an ambulance — increased after prior authorization. Taken as a whole, these results indicate that the patients who receive non-emergent ambulance rides after the start of prior authorization are less healthy, which is consistent with Medicare's

aim for the program — to provide rides only when medically necessary.

Table 9: Effect of Prior Auth. on Patient Selection

	(1) Rides Next Month	(2) Lifetime Rides	(3) Hospitalizations	(4) Mortality
Prior Authorization	-0.0718***	22.11**	0.0162***	0.00531***
	(0.00473)	(8.394)	(0.00386)	(0.00114)
Year-Month FE	1	1	1	1
District FE	1	1	1	1
Pat/Fac Controls	1	1	1	1
Facility FE	1	1	1	1
R-squared	0.111	0.310	0.0387	0.0216
Dep. Var. Mean	0.837	558.1	0.251	0.0341
Observations	783420	783420	783420	783420

Notes: Table gives estimates of β from equation (2) using traditional TWFE methods at the patient-month level. Controls include incident patient characteristics, age, and tenure on dialysis as well as facility characteristics including chain ownership status, demographic characteristics of the ZIP code, and whether the facility is freestanding or hospital-based. Sample is limited to patient-months in which the patient receives at least one non-emergent dialysis ambulance ride. Standard errors clustered at the district level are given in parentheses. $^{+}$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

The denial rate for submitted claims provides further evidence that prior authorization resulted in a more-appropriate use of ambulance rides. Although we do not observe requests submitted for prior authorization, we do observe whether a claim was paid if it was submitted after the service was rendered. Figure 12 shows that immediately following prior authorization, the share of denied claims jumped sharply and then declined gradually.[12] That the denial rate decreased following the initial spike indicates that some firms stopped submitting claims that would be denied given the heightened scrutiny of prior authorization. That is, we interpret this result as evidence that prior authorization acts as a screening mechanism that effectively deters fraud.

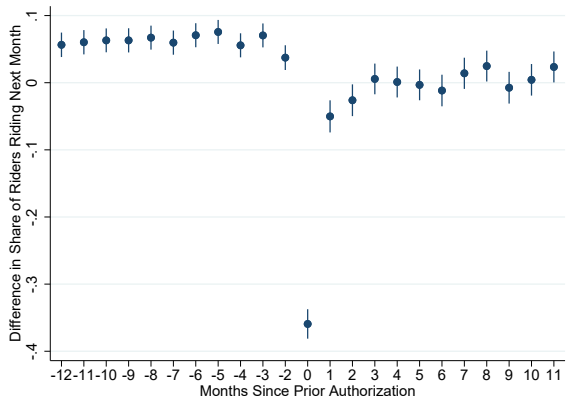
8 Regulation versus Litigation

An extensive literature has considered the benefits of using regulation to combat illegal behavior rather than litigation, mostly in the context of torts and property rights violations where individuals or private parties are harmed. Our work provides a natural extension of these studies to circumstances where the injured party is the government and the type of crime is financial

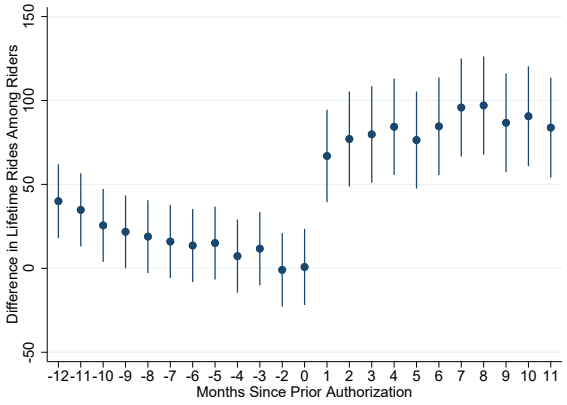
[12]Because these denial rates capture only claims that were submitted after providers could obtain prior authorization for the service, rather than including those that were denied prior authorization, the increase in denial rates after prior authorization is likely a lower bound for the true increase.

Figure 11: Effect of Prior Auth. on Patient Selection

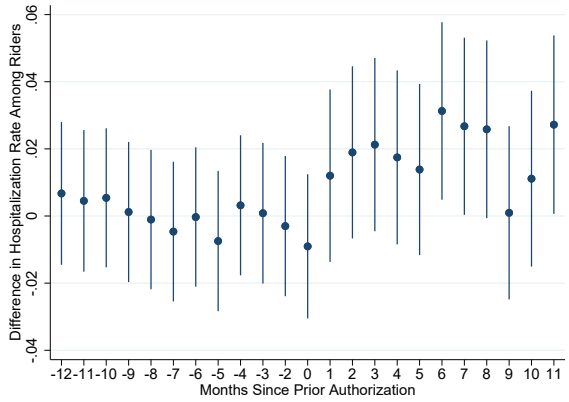
(a) Share of Riders Riding Next Month



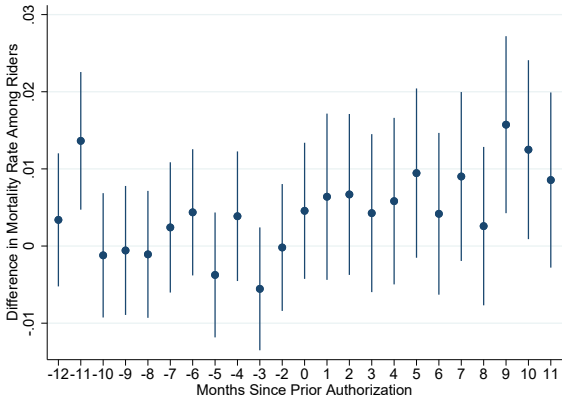
(b) Total Rides Taken



(c) Hospitalization Rate Among Riders



(d) Mortality Rate Among Riders

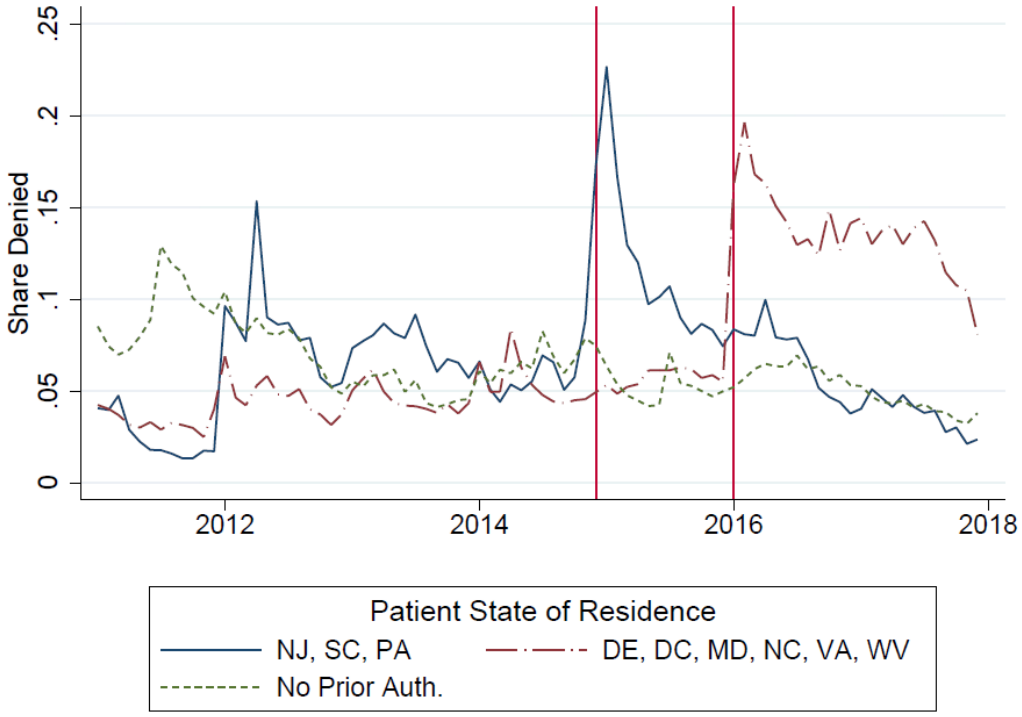


fraud. Moreover, our study makes a novel contribution given that many of the canonical results on deterrence do not apply to this form of illicit activity. In particular, the large literature examining torts and the assignment of property rights, such as Coase (1960), provides little guidance on how to efficiently deter financial fraud against the government. We therefore revisit the question of when and how litigation may effectively deter fraud on its own, or when regulation must be used in conjunction with it.

8.1 Why Litigation Failed

In the case of ambulance fraud, the government faces a number of constraints that make litigation unlikely to have a widespread effect on illegal behavior. First among these is the government's limited ability to levy large penalties on the fraudulent firms and operators we study, a necessary component of effective deterrence (Becker, 1968). Litigation may fail to curtail illicit behavior if severe penalties cannot be enforced (Shavell, 1984), like in the case of fly-by-night ambulance companies that may shut down or spend their ill-gotten gains before authorities

Figure 12: Claim Denial Rates by Prior Authorization Status



Notes: The sample includes non-emergent basic life support ambulance rides from a dialysis facility to a place of residence for ESRD patients from 2011–2017. State is determined by the transported patient's state of residence. Vertical lines mark the implementation of prior authorization in NJ, SC, and PA, and in DE, DC, MD, NC, VA, and WV. The share of claims denied is the share of rides for which the submitted claim was not paid any positive amount.

can recover the financial penalties imposed by the courts. Even among successfully prosecuted firms, the likelihood of receiving full restitution is low. Despite regularly reaching millions of dollars, the Department of Justice itself warns that restitution for criminal penalties is often difficult to enforce, writing, "Realistically, however, the chance of full recovery is very low...it is rare that defendants are able to fully pay the entire restitution amount owed" (Department of Justice, 2021). The difficulty of enforcing financial penalties may explain why we find that civil lawsuits are less effective than criminal enforcement. Civil lawsuits only impose monetary penalties or exclusion from the Medicare program, penalties that will not have much effect on firms that can simply shut down rather than change their behavior. Conversely, criminal lawsuits can impose jail time on the owners or operators of fraudulent firms, a non-monetary penalty that can be enforced even in the absence of recoverable funds.

Litigation may also have been ineffective because federal attorneys were unwilling to prosecute beneficiaries for being complicit in ambulance fraud. Patients were often a key part of the schemes, with criminal lawsuits alleging they received kickbacks for riding and for referring others

— in one particularly striking case, a patient threatened to report ambulance company employees to the police if they did not increase her kickback. Moreover, about 2,000 patients immediately stopped riding in the first three states subject to prior authorization, perhaps reflecting a large faction of complicit beneficiaries. Despite such compelling evidence of their widespread involvement, the government has only criminally prosecuted a handful of dialysis patients for ambulance fraud, like owing to the generally sympathetic nature of dialysis patients as well as the exorbitant costs of imprisoning them in one of the six overcrowded Bureau of Prisons Medical Centers, the highest-severity institutions (Office of the Inspector General, 2015; of Prisons Clinical Guidance, 2019).

In addition to the challenge of levying high penalties against proven lawbreakers, litigation may be hampered by the difficulty of detecting and successfully prosecuting illicit behavior at a sufficiently large scale: over 3,000 firms participated in non-emergent ambulance transportation of dialysis patients over our sample period, yet only 98 companies or individuals were ever prosecuted. That prior authorization was so effective at deterring medically unnecessary rides even after litigators and prosecutors had already made concerted efforts to stop them further reflects a low detection rate. For example, the ten lawsuits filed by the Pennsylvania East judicial district were more than any other district, but despite such active litigation, the number of active firms still fell from 83 to 47 in the three months immediately following prior authorization and the number of rides fell even further: an astounding 87.5% drop from 10,653 to 1,327. The large number of firms exiting in the face of regulation, even after extensive prosecution, suggests that criminal enforcement did not do much to deter fraudulent behavior.

A lack of specialization may partly explain these low detection rates (Landis, 1938). Twenty-three different judicial districts were involved in the lawsuits we study, which means 23 different sets of investigators, attorneys, and judges were responsible for understanding the complex nature of this fraud in order to successfully prosecute it. Moreover, the Department of Justice attorneys who work on health care fraud are responsible for enforcing many other parts of the federal criminal and civil code, as are the judges who try the cases.

Health care fraud may also be difficult to prove after the fact. Criminal lawsuits require a "beyond a reasonable doubt" evidentiary standard, and establishing a lack of medical necessity to this standard is challenging: the Department of Justice must amass incontrovertible evidence, such as video recordings of purportedly bedridden patients walking. With over 3,000 firms participating in nonemergency transportation, such cases cannot be widely prosecuted given the limited resources of the Department of Justice and the Federal Bureau of Investigation.

That the injured party is the government in our setting, rather than a private party, is another reason why litigation may not deter much illicit behavior on its own. Behrer et al. (2021) and Mookherjee and Png (1992) argue that litigation alone will be ineffective when the harm in question affects a large number of individuals and the private reporting of harm is insufficient.

The injured party in the case of health care fraud is every American taxpayer, and individuals are not empowered to protect the public interest. The government also faces agency problems, because the stolen money does not directly impact the federal employee. That is, the failure to detect health care fraud has limited consequences for those responsible for combating it.

8.2 Why Regulation Succeeded

In contrast to criminal and civil enforcement, regulation effectively deterred the type of health care fraud we study. To better understand why, we place our results within the framework of Glaeser and Shleifer (2003) that compares pure litigation-based enforcement to a regime that also uses administrative rules. Most relevant for our setting, they find that adding administrative rules is optimal in cases where litigation can be subverted. As noted above, litigation can be most effective when the enforcer is able to assess large penalties (Becker, 1968), yet larger penalties provide a stronger incentive for subversion (Glaeser and Shleifer, 2003). Although not addressed in prior work, the unwillingness of prosecutors to pursue complicit beneficiaries and the challenge of recovering stolen funds from fly-by-night firms are both forms of subversion that also make litigation ineffective at assigning liability. Conversely, prior authorization prevented fraudulent funds from ever being paid in the first place, making it unnecessary to assign ex post liability.

Regulation may also be superior to litigation because it is easier to enforce at a large scale. Compared to the low rates of detection and punishment through the courts, claim denial rates rose above 20% after the start of prior authorization. In this case, MAC administrators successfully detected noncompliant rides en masse, underscoring the benefits of using regulations when litigation faces capacity constraints (Glaeser and Shleifer, 2003). Moreover, regulations may improve detection rates by making noncompliance more obvious. Although courts may find it difficult to assess medical necessity, regulations can include "bright-line rules" that are easy to monitor (Kaplow, 1992; Glaeser and Shleifer, 2002). For prior authorization, it is much simpler to show that a firm failed to submit paperwork than it is to prove a patient did not have a legitimate medical reason for taking an ambulance ride.

Relatedly, administrative enforcers can be more specialized than judges or prosecutors, which facilitates monitoring (Landis, 1938). In the case of ambulance fraud, assessing medical necessity requires specialized knowledge by the enforcer. Compared to Department of Justice attorneys, the MAC administrators responsible for checking prior authorization requirements focus entirely on Medicare regulations. Thus, while the Department of Justice must convince unspecialized judges and juries that care was not medically necessary, MAC administrators can more competently and efficiently examine supporting documentation and decide whether a reimbursement is justified.

8.3 Broader Effects

In addition to deterrence, regulation and litigation can produce other effects that are difficult to measure empirically. In response to increased scrutiny, some may choose to forgo fraud in the first place, a general deterrence effect of unknown magnitude (Shavell, 1991; Leder-Luis, 2019). Conversely, individuals intent on committing health care fraud may substitute away from one particular scheme and pursue others that are more difficult for authorities to detect.

Regulation and litigation both have costs that affect their relative efficiency. Because monitoring paperwork for prior authorization is much simpler than ex post enforcement against fraudulent claims, regulation can likely accomplish the same level of deterrence at a much lower cost. The chief actuary for CMS estimated the cost of implementing prior authorization nationwide at only "$38.1 million in the first expansion year and $28.6 million per year in subsequent years," substantially less than the potential cost of prosecuting all fraudulent ambulance firms (Spitalnic, 2018). Furthermore, litigation expends scarce Department of Justice and Federal Bureau of Investigation resources, which may come with the opportunity cost of less effort in other areas.

On the other hand, regulation may be costly if it results in care being rationed inefficiently (American Medical Association, 2021) or imposes large hassle costs on patients (Herd and Moynihan, 2018) and providers (Dunn et al., 2021). In the context of non-emergent ambulance rides for dialysis patients, we find no evidence supporting the former concern, as discussed in Section 6, although this result may not hold for regulations in other settings. In terms of hassle costs, it is unlikely that patients or physicians bear large costs from prior authorization, as ambulance companies are largely responsible for supplying a proof of medical necessity to MAC administrators. On the other hand, our finding that some ambulance companies became more specialized in non-emergent rides after the reform, as shown in Figure 8, could reflect barriers to entry for the supply of non-emergent rides. This finding is also consistent with theoretical evidence that regulations can be efficient even when some firms profit from regulatory capture (Glaeser and Shleifer, 2003).

One potential benefit of imposing hassle costs through regulation, however, is the potential for regulation to act as a screening mechanism. If the regulation is well targeted, only medically necessary services will be rendered, as providers and patients anticipate that only proper claims will be approved (Zeckhauser, 2021). The pattern of denial rates shown in Figure 12 provides evidence of this phenomenon: following an initial spike in denied claims, the denial rate starts to fall once firms begin submitting claims for only medically necessary rides.

9 Conclusion

In this paper, we show that prior authorization is much more effective than criminal or civil enforcement at reducing fraudulent ambulance rides. Prior authorization caused an immediate and persistent drop in non-emergency ambulance rides of nearly 53%, whereas criminal and civil lawsuits had a much smaller effect. Had the federal government required prior authorization when it first became aware of the fraudulent activity in 2003, it would have saved $4.8 billion.[13] Given the relative costs of litigation and regulation, we find that implementing a prior authorization regime is much more efficient.

Importantly, we show that the decrease in non-emergent rides did not come at the expense of patient health even though it drove many ambulance companies out of the market. Following prior authorization, the dialysis patients who continued taking non-emergent ambulance rides were in poorer health, suggesting that the Medicare benefit was being used more efficiently.

We connect our results to the economic theory that explains why regulation is necessary — and litigation alone insufficient — for successfully combatting ambulance fraud, which also applies more broadly throughout the health care system. Criminal and civil penalties are often too low given prosecutors' unwillingness or inability to levy high penalties against patients or fly-by-night firms, and prosecution rates are held back by the challenges of detecting fraud, the diffuse nature of the harm, and the limited resources of unspecialized enforcers. In addition, litigation is unlikely to deter fraud more generally because health care fraud requires more specialization than other criminal activities. This points to health care fraud as being an area in need of regulatory innovations to complement any legal enforcement that comes through prosecution.

Medicare has recently moved in this direction by expanding prior authorization to other areas of health care that may be susceptible to fraud, including power mobility devices, home health services, and hyperbaric oxygen.[14] Our results indicate that these efforts are likely to be successful, motivating our future work on this topic.

[13]This is the sum of the amount paid from 2003–2017 for non-emergent dialysis ambulance rides above the mean levels within prior authorization waves from 2003 and 2004 ($4.1 billion dollars) and the difference between realized spending and a linear projection of the spending within prior authorization waves from the time of the first wave of prior authorization to the end of 2017, fitting the trend to the five previous years ($703 million).

[14]For more information on CMS's prior authorization programs, see `https://www.cms.gov/research-statistics-data-systems/medicare-fee-service-compliance-programs/prior-authorization-and-pre-claim-review-initiatives`.

References

(2018, Mar). Find your united states attorney.

American Medical Association (2021). 2020 ama prior authorization (pa) physician survey. Technical report, American Medical Association.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy 76*(2), 169–217.

Behrer, A. P., E. L. Glaeser, G. A. M. Ponzetto, and A. Shleifer (2021). Securing property rights. *Journal of Political Economy*.

Callaway, B. and P. H. Sant'Anna (2020). Difference-in-differences with multiple time periods. *Journal of Econometrics*.

Centers for Medicare and Medicaid Services (2020, Dec). Repetitive, scheduled non-emergent ambulance transport (RSNAT) prior authorization model frequently asked questions. Woodlawn, MD. `https://www.cms.gov/Research-Statistics-Data-and-Systems/Monitoring-Programs/Medicare-FFS-Compliance-Programs/Prior-Authorization-Initiatives/Downloads/AmbulancePriorAuthorization_ExternalFAQ_121517.pdf`.

CMS (2020). Fact sheet: 2020 estimated improper payment rates for centers for medicare and medicaid services (cms) programs.

Coase, R. H. (1960). The problem of social cost. *Journal of Law and Economics 3*(2), 1–44.

Dafny, L. S. (2005). How do hospitals respond to price changes? *American Economic Review 95*(5), 1525–1547.

Department of Justice (2021). Restitution process.

Dunn, A., J. D. Gottlieb, A. Shapiro, D. J. Sonnenstuhl, and P. Tebaldi (2021, July). A denial a day keeps the doctor away. Working Paper 29010, National Bureau of Economic Research.

Eliason, P. J., B. Heebsh, R. C. McDevitt, and J. W. Roberts (2020). How acquisitions affect firm behavior and performance: Evidence from the dialysis industry. *Quarterly Journal of Economics 1*(135), 221–267.

Fang, H. and Q. Gong (2017). Detecting potential overbilling in medicare reimbursement via hours worked. *American Economic Review 107*(2), 562–591.

Glaeser, E. L. and A. Shleifer (2002, 11). Legal Origins. *The Quarterly Journal of Economics 117*(4), 1193–1229.

Glaeser, E. L. and A. Shleifer (2003, June). The rise of the regulatory state. *Journal of Economic Literature 41*(2), 401–425.

Herd, P. and D. P. Moynihan (2018). *Administrative Burden: Policymaking by Other Means*. Russell Sage Foundation.

Howard, D. H. and I. McCarthy (2021). Deterrence effects of antifraud and abuse enforcement in health care. *Journal of Health Economics 75*, 102405.

Kaplow, L. (1992). Rules versus standards: An economic analysis. *Duke Law Journal 42*(3), 557–629.

Landis, J. M. (1938). *The administrative process*. Yale University Press.

Leder-Luis, J. (2019). Whistleblowers, the false claims act, and the behavior of healthcare providers. *Working Paper RePEc*. `https://ideas.repec.org/jmp/2019/ple1069.pdf`.

Levitt, S. D. (1997). Using electoral cycles in police hiring to estimate the effect of police on crime. *The American Economic Review 87*(3), 270–290.

Mookherjee, D. and I. P. L. Png (1992). Monitoring vis-á-vis investigation in enforcement of law. *The American Economic Review 82*(3), 556–565.

of Prisons Clinical Guidance, F. B. (2019). Care level classification for medical and mental health conditions or disabilities.

Office of the Inspector General, U. D. o. J. (2015). The impact of an aging inmate population on the federal bureau of prisons.

O'Malley, A. J., T. A. Bubolz, and J. S. Skinner (2021). The diffusion of health care fraud: A network analysis. Technical report, National Bureau of Economic Research.

Sanghavi, P., A. B. Jena, J. P. Newhouse, and A. M. Zaslavsky (2021). Identifying outlier patterns of inconsistent ambulance billing in medicare. *Health Services Research 56*(2), 188–192.

Shavell, S. (1984). Liability for harm versus regulation of safety. *The Journal of Legal Studies 13*(2), 357–374.

Shavell, S. (1991, 10). Specific versus General Enforcement of Law. *Journal of Political Economy*.

Silverman, E. and J. Skinner (2004). Medicare upcoding and hospital ownership. *Journal of Health Economics 23*(2), 369–389.

Spitalnic, P. (2018). Certification of medicare prior authorization model for repetitive scheduled non-emergent ambulance transport (rsnat). Technical report, Centers for Medicare and Medicaid Services.

Wilson, W. (1913). *The new freedom: A call for the emancipation of the generous energies of a people*. Doubleday.

Zeckhauser, R. (2021). Strategic sorting: the role of ordeals in health care. *Economics & Philosophy 37*(1), 64–81.

APPENDIX

A US Court Districts

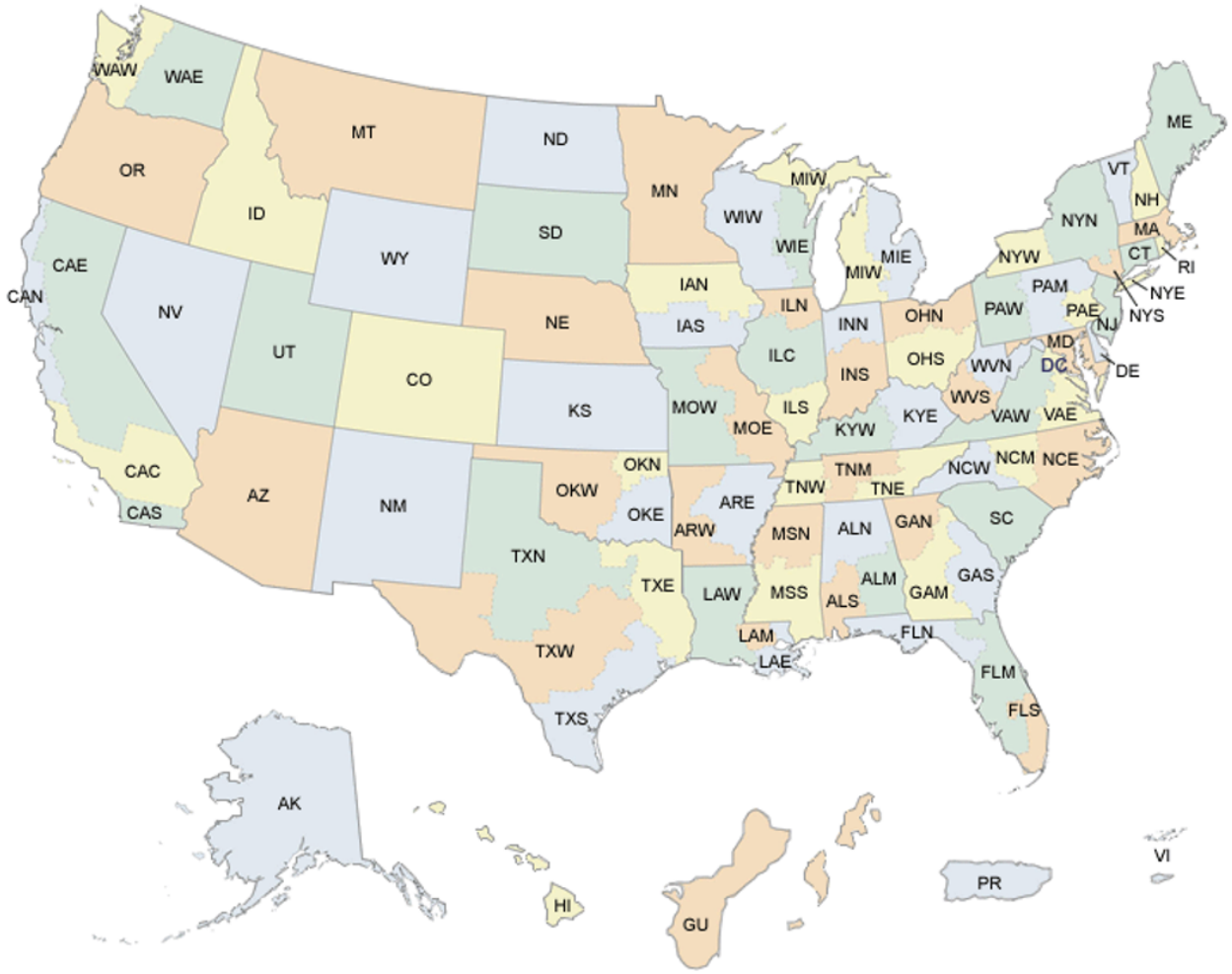


Figure A1: Map of US Court Districts from DOJ (2018)

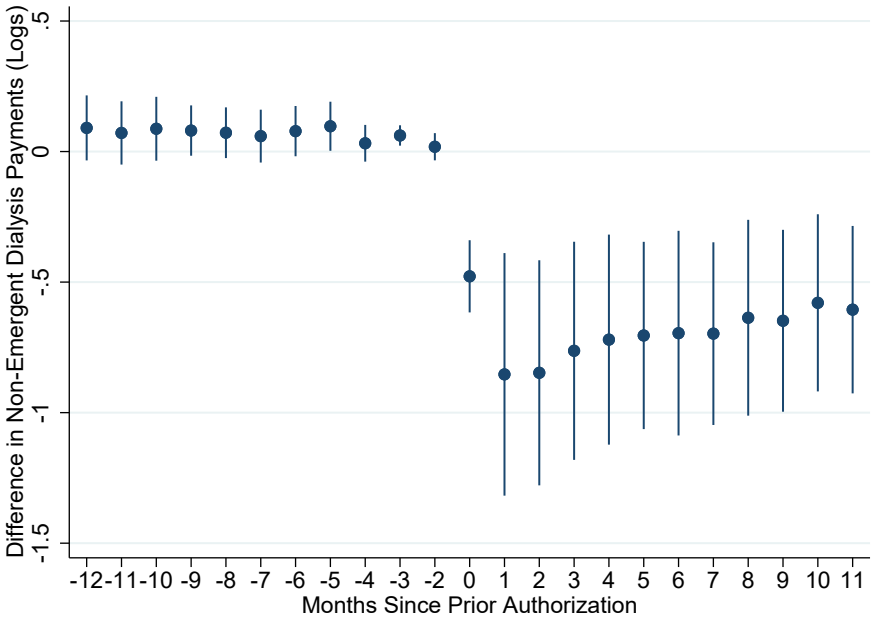
B More Results for Effects Prior Authorization

In this appendix we present additional results on the effects of prior authorization that we refer to throughout the paper.

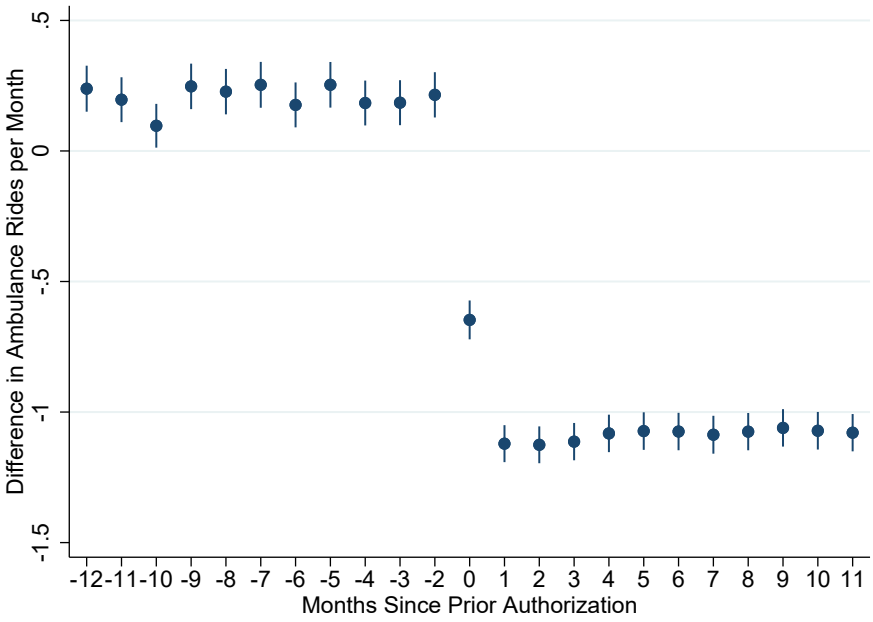
First we show in Figure B2 that our estimate of the large effect of prior authorization on rides is robust to analysis at the firm-month and patient-month level using traditional TWFE methods.

Figure B2: Effect of Prior Auth. on Ridership

(a) Firm-Level Effect on Non-Emergent Dialysis Rides (Log)



(b) Patient-Level Effect on Non-Emergent Dialysis Rides



As a placebo, we also show in Table B1 and Figure B3 that prior authorization had no impact on the number of emergent rides.

Finally, we present results from Sections 6 and 7 using the CS methodology.

Table B1: Effect of Prior Auth. on Emergency Ambulance Spending

	(1) Payments for Emergent Rides (Log)	(2) Payments for Emergent Rides (Log)	(3) Payments for Emergent Rides (Log)	(4) Payments for Emergent Rides (Log)
Prior Auth.	0.575***	0.581***	-0.0175*	0.0112
	(0.0346)	(0.0388)	(0.00794)	(0.0266)
Year-Month FE	0	1	1	1
District FE	0	0	1	1
CS Methods	0	0	0	1
Dep. Var. Mean	11.32	11.32	11.32	11.32
Observations	7810	7810	7810	7810

Notes: Dependent variable is the natural logarithm of the total payments for emergency ambulance transports in the district-month. These data include rides from 2011–2017. An observation is a district-month. Columns (1) and (2) give the estimate of $\theta_{es}^{O,bal}(23)$ using CS methods for civil enforcement, while columns (3) and (4) do the same for criminal enforcement. The treatment date is the earliest enforcement action of the relevant type in the district. Standard errors are obtained using CS's bootstrap-based procedure. $^{+}$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

Figure B3: Effect of Prior Auth. on Payments for Emergency Rides

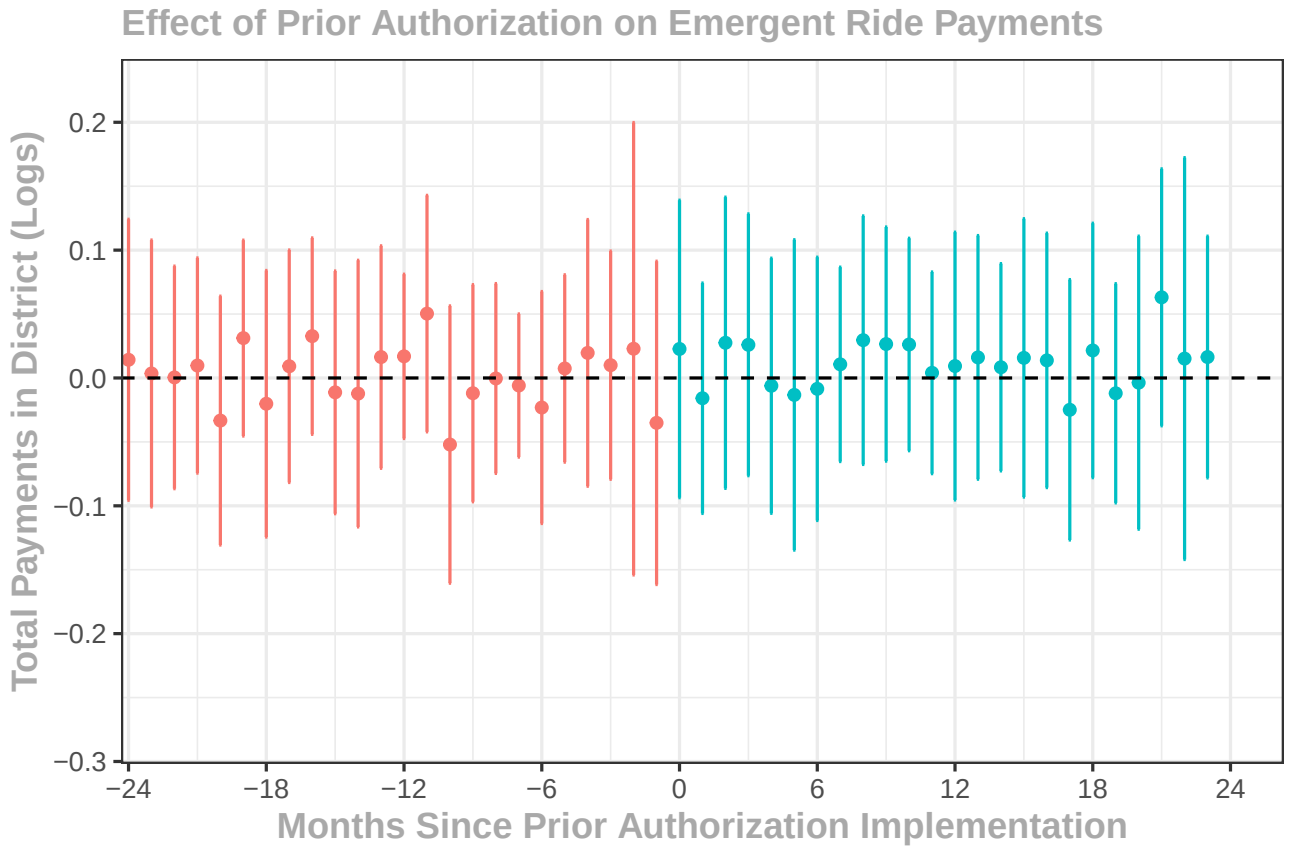


Table B2: Effect of Prior Auth. on Adherence and Adverse Events

	(1) Dialysis Sessions per Patient	(2) Mortality Rate	(3) All-Cause Hosp.	(4) Insufficient Dialysis Hosp.
Prior Authorization	-0.0491	0.0001	-0.0033	-0.0024*
	(0.0314)	(0.0006)	(0.0023)	(0.0012)
Year-Month FE	1	1	1	1
Pat/Fac Controls	0	0	0	0
District FE	1	1	1	1
CS Methods	1	1	1	1

Notes: Table gives estimates of $\theta_{es}^{O,bal}(23)$ using the CS methods. An observation is a district-month. All outcomes are per patient-month. Insufficient dialysis hospitalizations are those for which the primary diagnosis indicates excess fluids. Standard errors are obtained using CS's bootstrap-based procedure. $^{+}$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

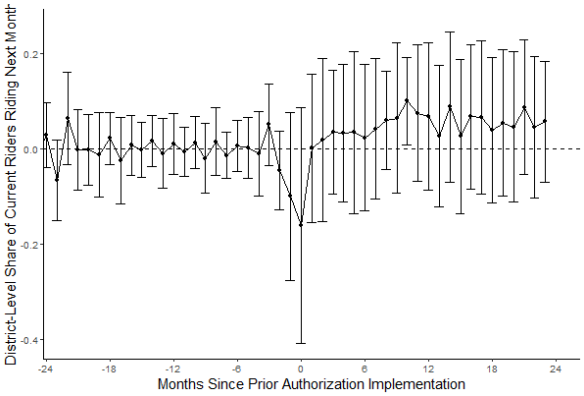
Table B3: Effect of Prior Auth. on Patient Selection

	(1) Rides Next Month	(2) Total Lifetime Rides per Patient	(3) Hospitalization Rate Among Riders	(4) Mortality Rate Among Riders
Prior Authorization	0.0419	90.3404***	0.0226	0.0046
	(0.0487)	(26.8898)	(0.0202)	(0.0104)
Year-Month FE	1	1	1	1
Pat/Fac Controls	0	0	0	0
District FE	1	1	1	1
CS Methods	1	1	1	1

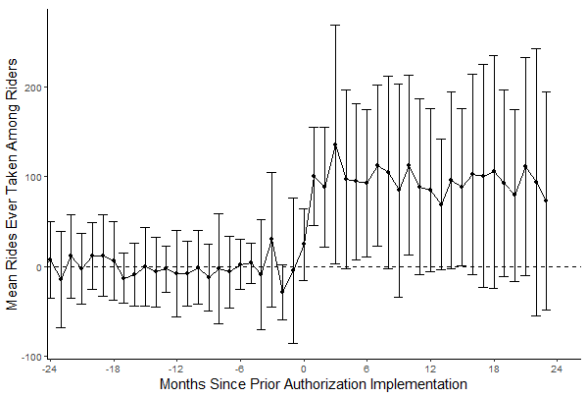
Notes: Table gives estimates of $\theta_{es}^{O,bal}(23)$ using the CS methods. An observation is a district-month. All outcomes are per patient-month in which the patient takes at least one non-emergent dialysis ride. Standard errors are obtained using CS's bootstrap-based procedure. $^{+}$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

Figure B4: Effect of Prior Auth. on Patient Selection

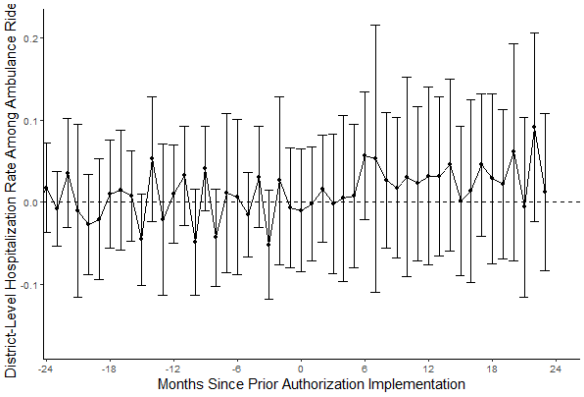
(a) Share of Riders Riding Next Month



(b) Total Rides Taken



(c) Hospitalization Rate Among Riders



(d) Mortality Rate Among Riders

