Regulation and Diffusion of Innovation Under Information Spillovers: The Case of New Medical Procedures*

Riley League†

September 2024

The value of innovative technologies and practices is often initially uncertain, forcing policy-makers to weigh the potential benefit of promoting a valuable innovation against the cost of encouraging adoption of an unsuccessful one. Complicating this tradeoff is the potential for wider adoption of the innovation to dispel uncertainty about its efficacy. In this paper, I examine these issues in the context of new medical procedures, where the value of each innovation is highly uncertain and Medicare contractors must decide whether to reimburse health care providers for the procedure. Using geographic variation in the coverage rules issued by these contrators, I show that they exert significant influence on providers' adoption of new procedures. Next, I leverage the resulting variation in the incentives of providers to adopt new procedures to identify information spillovers from individual providers' experiences with the new procedures. I present evidence that social learning is an important determinant of the spread of innovation in this context. Finally, in light of this evidence, I estimate a structural model of innovation adoption and provider learning to determine the optimal Medicare coverage policy for new procedures. In counterfactual simulations, I find that universally covering all new procedures would result in large welfare gains.

*This research was supported by the National Institute on Aging, grant number T32-AG000186.
†University of Illinois Urbana-Champaign, league@illinois.edu

1 Introduction

Regulators often must make decisions about how high to set the evidentiary threshold for allowing the proliferation of innovations of uncertain value. These regulators must trade off the potential gains from the rapid adoption of a valuable innovation against the risk of allowing a wasteful or even harmful innovation to become more widely employed. The question of how much evidence policymakers should require before facilitating the adoption of an innovation is complicated by the potential for social learning—increased adoption of the innovation can itself reduce uncertainty about its value as it becomes more familiar to a wider audience. In this paper, I address these issues in the context of Medicare coverage for new medical procedures.

New medical procedures are an excellent context in which to study these issues. First, new medical procedures represent a large class of innovations for which the value is extremely uncertain. Over 700 new medical services have been introduced since 2002, but only around 60% of these procedures have been successfully adopted into medical practice. The successful innovations can be a source of such valuable advances as improved health, cost savings, and increased productivity, while the unsuccessful ones can lead to increased costs, uncertain health gains, and waste (Cutler and McClellan, 2001; Cutler and Huckman, 2003; Chernew and Newhouse, 2011; Baicker et al., 2012; Skinner and Staiger, 2015; Hwang et al., 2016). The combination of high potential payoffs from successful innovations and considerable uncertainty around which procedures will prove to be valuable makes the tradeoff between allowing early adoption and preventing the spread of low-value innovations highly salient to policymakers.

The regulatory system for new medical services highlights this tradeoff while also presenting a unique opportunity for identification of learning spillovers across health care providers. The determinations of whether Medicare will reimburse providers for each of these services are made by the privately-owned companies that contract with the government to administer Medicare. Importantly, each of these administrators, called Medicare Administrative Contractors or MACs, has a distinct geographic jurisdiction in which they make coverage determinations. These administrators frequently issue conflicting coverage rules, resulting in significant variation in where and when new procedures will be reimbursed by Medicare as coverage is rolled out in a staggered fashion for each new procedure. This variation results in providers in some jurisdictions being incentivized to adopt the innovation early when evidence surrounding its efficacy is scarce while others must wait and learn from the experience of providers elsewhere until the procedure is covered in their jurisdiction.

I use this exogenous variation in the incentives to adopt new procedures in different information environments to identify the extent to which the adoption of new procedures is driven by social learning and determine the optimal evidentiary threshold on the part of Medicare administrators in light of this phenomenon. To do this, I first assess the ability of Medicare administrators to influence the adoption of new procedures by health care providers. Leveraging the variation in timing of coverage across jurisdictions, I find that granting local coverage to a new procedure leads to a 6-fold increase in utilization. This novel evidence of the ability of local administrators to impact the adoption of new innovations indicates that the question of how these administrators should set their evidentiary thresholds is of great importance to welfare. Furthermore since administrators have such an ability to influence utilization, if there are large spillovers of knowledge between providers, then these coverage rules must grapple with the information externality their decisions entail: providers in other jurisdictions benefit by learning from the experience of providers elsewhere.

I provide evidence that this learning externality is large. First, I show that utilization responds to changes in coverage in other jurisdictions. Importantly, these spillovers are positive for procedures that turn out to be high value, while they are negative for innovations that eventually are found to be of low value. This result is consistent with providers receiving information about the efficacy of new procedures from the increased utilization in areas with different coverage rules and updating their beliefs accordingly. Similarly, I show that idiosyncratic shocks to the initial level of utilization of a procedure are persistent for successful innovations but not for unsuccessful ones. This is true conditioning on a proxy for the initial beliefs about the efficacy of the new procedure, as well as instrumenting for past utilization using the relative size of the jurisdictions in which the procedure was previously covered. Once again, this indicates that the more quickly the medical community gains experience with a procedure, the more quickly it adopts high value innovations and de-adopts low value ones. Finally, I present suggestive evidence that when a procedure becomes covered in larger jurisdictions, the resulting increase in the rate of adoption for successful procedures is larger than for smaller jurisdictions as larger jurisdictions generate more signals of the innovation's true quality. Conversely, utilization of unsuccessful procedures falls more quickly in larger jurisdictions for the same reason. I also present evidence that other potential drivers of diffusion patterns including learning from clinical trials and technological change are unable to explain the patterns of adoption I observe in this context.

In light of the evidence that social learning is an important factor in the diffusion of new medical procedures, I address the question of how Medicare should set its evidentiary threshold for coverage by estimating a structural model of physician learning in response to Medicare coverage decisions. This model allows me to quantify the tradeoffs faced by policymakers and inform welfare-enhancing policy changes to Medicare coverage policies. I model physicians as Bayesian learners generating a noisy public signal of the procedure's quality each time they perform it. Identification here is very difficult in most contexts: whether a potential adopter incorporates an innovation early or late is endogenous to the agent's beliefs about the value of the innovation, meaning that late adopters may behave differently than early adopters for reasons other than the development of evidence by the early adopters. In my context though, I can leverage exogenous variation in the incentives faced by providers stemming from differences in local Medicare coverage decisions to identify the model. In particular, I am able to use this variation to identify physicians that adopt the new procedure early (when uncertainty about its value is rampant) or late (when the uncertainty is lower) for reasons beyond the physician's perceived value of the innovation. This novel use of an understudied institutional detail allows me to quantify the value of learning spillovers from early- to late-adopting physicians. Using this structural model, I am able to simulate counterfactual adoption patterns under more or less stringent coverage rules, finding that transitioning to a regime of universal coverage of these new procedures would result in large welfare gains relative to the current regime, achieving 93% of the welfare gains that covering all and only procedures that are better than the respective incumbent procedure would achieve, indicating that allowing for additional experimentation and learning would be extremely valuable to patients.

The tradeoff faced by Medicare administrators between allowing for early experimentation and learning or waiting until the evidence is more certain is something regulators must grapple with across many different arenas. There is substantial policy debate over whether the standards chosen by policymakers are appropriate in areas as disparate as Food and Drug Administration approval of pharmaceuticals (Jewett, 2022; Makary, 2021) to investments in green technology (Storrow, 2021). In line with the importance of this tradeoff, there have been a number of academic studies highlighting that this tradeoff exists and asking whether the regulatory regime is optimal. The most prominent examples of these studies are in the medical context, including studies pointing out that shortened drug review times lead to more adverse events (Grabowski and Wang, 2008; Olson, 2008) and studies of the medical device industry arguing that the regulatory incentives manufacturers face lead to underinvestment in clinical trials and restrict new device entry (Budish et al., 2016; Stern, 2017). Perhaps most closely related to my study, Grennan and Town (2020) compare the review processes for new medical devices in the United States and European Union and find that the higher US standard is indistinguishable from one that maximizes total surplus. In the context of new medical procedures, by contrast, I find that welfare could be increased by dramatically lowering the regulatory standard.

The potential for social learning complicates the tradeoff between allowing rapid diffusion of high-potential innovations at the risk of allowing more low-value ones to proliferate. Social learning is the process of agents updating beliefs about the efficacy of an innovation through the receipt of public signals generated by the innovation's use. It is important to note that this type of learning refers to acquiring more accurate beliefs, rather than learning how to deploy resources more efficiently and achieve better outcomes. While both types of learning are often present in medical contexts, the social learning that I study is distinct from this more commonly studied phenomenon of learning-by-doing.[1] Nonetheless, social learning has long been thought to be important for physicians (American Medical Association, 2010). National conferences, medical society meetings, and communications allow for the rapid, wide diffusion of new information about physician experiences through word-of-mouth, and continuing education requirements lead health care providers to be exposed to new developments in their profession (McKinlay, 1981). Indeed, empirical research has corroborated that health care providers learn from the experiences of those they come in contact with (Allen et al., 2019; Soumerai et al., 1998). This literature focuses on documenting evidence of social learning at the local level: knowledge spreading between physicians that are geographically clustered (Agha and Molitor, 2018) or socially connected (Zheng et al., 2010). In contrast, my paper focuses on the global knowledge spillovers from each physician to the entire medical community. To my knowledge, no research has empirically documented these global spillovers despite their likely importance.

The global social learning I study is thought to be an important driver of the diffusion of innovations even beyond medicine, with phenomena as diverse and important as the Industrial Revolution (Mokyr, 2016), the use of US states as laboratories of democracy (Brandeis, 1932; Callander and Harstad, 2015), and the proliferation of pro-market economic policies (Buera et al., 2011) having been attributed to social learning. However, given the universal nature of the learning spillovers in many contexts, identification of social learning has proven difficult. Due to the endogeneity of innovation adoption decisions, to credibly identify global learning spillovers from one agent to all others, the research design needs exogenous variation in the timing of adoption. In light of this difficulty, the vast majority of empirical research on social learning has focused on documenting particular channels through which knowledge can spread, including social networks (Allen et al., 2019; Foster and Rosenzweig, 1995; Ryan and Gross, 1943), geography (Chandra and Staiger, 2007; Conley and Udry, 2010; Agha and Molitor, 2018), and professional connections (Kellogg, 2011). Existing empirical studies of social learning that consider its global nature generally lack exogenous variation in the information environment and incentives of agents to adopt the innovation (e.g Fafchamps et al., 2016; Moretti, 2011; Covert, 2015). One notable exception is Gilchrist and Sands (2016), who find that variation in weather on the opening weekend of films leads to persistent differences in viewership over the entire theater run, arguing that this phenomenon is driven by a desire for shared experiences rather than learning about the quality of the movie (as had been posited by Moretti (2011)).

[1] There are a number of excellent studies about physicians updating their beliefs about the efficacy of new treatments, including Coscelli and Shum (2004), Crawford and Shum (2003), and Ferreyra and Kosenok (2011), which study the adoption of a new anti-ulcer drug by Italian physicians and consider how providers learn about the drug's efficacy for different types of patients. In contrast, there is a distinct literature studying learning-by-doing. For example, Gowrisankaran et al. (2006) and Hockenberry and Helmchen (2014) find evidence of learning-by-doing for surgical procedures, while Gong (2017) considers both types of learning and finds them present in brain aneurysm treatment. Outside of medical contexts, there is a very large literature studying learning-by-doing, which in similar spirit to this paper has been shown to involve spillovers across agents (Thornton and Thompson, 2001; Stoyanov and Zubanov, 2012; Yang, 2022).

Similarly, I will be able to exploit exogenous variation in the initial information available about each innovation, albeit in a context in which policymakers play a much more active role.

Finally, my finding that the actions of Medicare administrators are powerful drivers of the adoption of new medical procedures relates to a fast-growing literature on the importance of administrative actions by health insurers to the provision of medical services. While recent research has highlighted the potential for administratively determined prices (Clemens and Gottlieb, 2014), denials rates (Dunn et al., 2021; League, 2023), prior authorization policies (Brot-Goldberg et al., 2022; Eliason et al., 2021), and audits (Shi, 2022) to influence medical practice, none have studied the particularly stark administrative decision about coverage or non-coverage. Furthermore, few studies of Traditional Medicare have recognized the decentralized administrative structure of Medicare as contributing to variation in these administrative rules (League, 2023). A small number of studies have noted the high level of variation in posted rules about coverage across contractors (Foote and Town, 2007; Levinson, 2014) while others have highlighted discrete cases where differences in these rules may lead to differences in medical practice (Wilk et al., 2018; Carlson et al., 2009; Foote et al., 2008). Nonetheless, none of these studies focus on new procedures—a context in which coverage rules are both particularly stark and particularly variable—nor provide systemic evidence of the impact of these differences across a broad class of medical services.

The remainder of this paper is laid out as follows. In Section 2, I discuss the relevant institutional details for the context of my study. In Section 3, I present the data used for this project as well as summary statistics on the coverage decisions of Medicare contractors and the success of new medical innovations. In Section 4, I exploit the differential timing of coverage across jurisdictions to show the sizable impact of coverage decisions on the adoption of new procedures. In Section 5, I present evidence of social learning and discuss alternative explanations for the patterns observed in my data. In Section 6, I estimate a structural model of physician adoption of new procedures, discussing policy-relevant and economically interesting counterfactuals. Finally, in Section 7, I conclude and discuss directions for future research.

2 Institutional Context

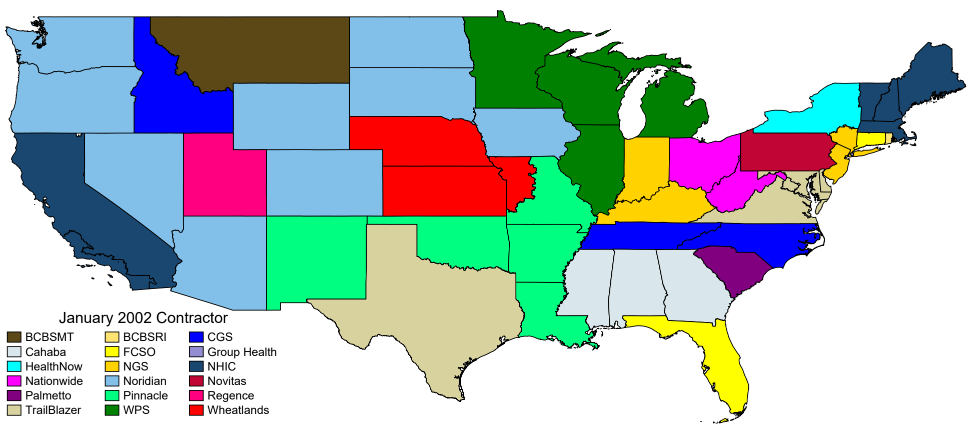
2.1 Medicare Administrative Contractors and Coverage Rules

Traditional Medicare is often thought of as a monolithic, federally run insurance program (NBPAS, 2021). But while the government bears all actuarial risk and determines the vast majority of Medicare policy, the day-to-day administrative operations are performed by private contractors called Medicare Administrative Contractors, or MACs. The administrative tasks performed by these contractors include processing medical claims and prior authorization requests, enrolling health care providers in the Medicare program, and determining the conditions under which Medicare will reimburse providers for various health care services.

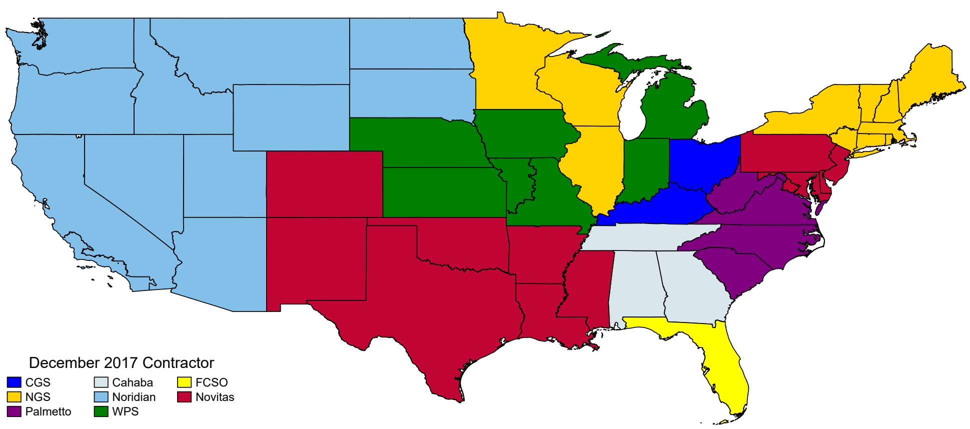
These administrators are contracted to provide administrative services for distinct regional jurisdictions. Figure 1 shows the areas administered by each administrative company in January 2002 and December 2017. At the beginning of my sample, there were 19 active administrative companies operating jurisdictions that sometimes spanned state borders (e.g., the Washington, DC area) or were strict subsets of states (e.g. New York). Over time, Medicare has combined administrative jurisdictions, leading many companies to exit the market and reducing the geographic variation in coverage rules. Over my entire sample, 19 administrative companies were active across 57 jurisdictions. A more detailed description of the jurisdiction combination process and its effects on the health care system is available in League (2023).

Figure 1: Map of MAC Jurisdictions

(a) 2002



(b) 2017



Notes: Each panel reports the administrative company responsible for processing Medicare Part B claims in each jurisdiction of the continental United States in the relevant month. Panel (a) reports this data for January 2002 while panel (b) reports data for December 2017.

While there are statutory guidelines as to the type of medical services Medicare is intended to pay for, these administrators have wide discretion over how to implement these broad standards. The coverage standard the administrative contractors must implement is to avoid payment for services that "are not reasonable and necessary for the diagnosis or treatment of illness or injury or to improve the functioning of a malformed body member" (Social Security Act, 1965a). While there are a few examples of the federal government providing more specific guidance on whether certain services meet this standard, in general, these determinations are left to the local contractors.[2] In particular, MACs tend to disagree quite frequently on coverage of new procedures. There is anecdotal evidence that this variation can be attributed to the fact that the employees who develop the coverage rules vary widely in their propensity to allow coverage of new procedures.[3] In fact, one reason for the recent consolidation of administrative jurisdictions was a desire on the part of policymakers to mitigate the impact of the apparently arbitrary differences in coverage across jurisdictions (Levinson, 2014).

Because MACs decide Medicare coverage rules for the jurisdictions in which they administer Medicare, this results in geographic variation in coverage at any given point in time. For example, an inspector general report found that in 2011 almost two-thirds of procedures were subject to local coverage restrictions in at least one jurisdiction, but among these, only 59% were subject to restrictions in all jurisdictions (Levinson, 2014). Not only are there differences in coverage at a single point in time, but these differences also change over time. Administrators continually update their coverage rules in light of new evidence on the efficacy of treatments, leading to significant variation in coverage rules over time.

2.2 New Medical Procedures

New medical procedures go through a much lighter touch of regulatory process than other medical innovations, such as pharmaceuticals or devices. After the procedure is created, if the procedure is truly different from established practice, the American Medical Association assigns the procedure a category III Current Procedural Terminology, or CPT, code. CPT codes are used by health care providers to inform health insurers what services they've rendered to patients in order to be reimbursed. Category III codes in particular are temporary and are meant to track

[2]The federal government can specify coverage rules legislatively or administratively. Legislative rules must go through the normal legislative process and so are uncommon. A rare example of this is regulation on the allowed frequency of various screenings, including mammography and colonoscopy (Social Security Act, 1965b). More common are administratively created rules. Analogous to the Local Coverage Determinations (LCDs) issued by the local contractors are the National Coverage Determinations (NCDs) issued by the Centers for Medicare and Medicaid Services. NCDs supersede LCDs and are made when "the service is the subject of substantial controversy" surrounding the item or service (Centers for Medicare and Medicaid Services, 2003). One prominent recent example of this is the NCD limiting coverage of the controversial Alzheimer's drug Aduhelm (Centers for Medicare and Medicaid Services, 2022).

[3]One MAC administrator attributed much of the variation to differences in the scrutiny that administrators apply to the evidence for or against coverage. Another explained the variation as coming from the fact that "every administrator is different."

the adoption of new procedures. Since the introduction of category III codes in 2002, there have been codes generated over 700 new procedures.

After a new procedure is assigned a category III code, Medicare Administrative Contractors determine whether Medicare will reimburse providers in their jurisdiction for performing the procedure. Because category III codes represent new procedures, MACs have significant leeway over coverage rules for these procedures. In fact, all MACs currently have a presumptive non-coverage rule for all category III codes, with coverage only extended on a procedure-by-procedure basis.

After a period that generally lasts five years, the American Medical Association reassigns the procedure to either a permanent category I code or deletes the code.[4] Procedures reported using category I codes are the vast majority of codes, they're much more likely to be paid by insurers, and in this context can be thought of as marking the innovation's success (Dranove et al., 2021).

This regulatory process has received little attention from academics. One notable exception is Dranove et al. (2021), who find that utilization of these procedures rises when the procedures are promoted from category III to category I codes and argue that administrative barriers and a lack of property rights in this context depress innovation and slow the diffusion of new procedures. The authors abstract from the main regulatory player studied in this paper: Medicare Administrative Contractors.

This regulatory path from innovation to acceptance highlights the high level of uncertainty about the quality of the new procedures. Even after being assigned a category III code, new procedures are subject to disagreements among Medicare administrators about whether they meet Medicare's coverage standards after which they are often deleted from the CPT system by the American Medical Association. Indeed, Dranove et al. (2021) find that for the period they analyze, only 29% of procedures are promoted to category I status on time. Similarly, many of these procedures fail to ever become widely accepted while others fall out of favor after initial excitement. For example, He et al. (2019), Steinbuch et al. (2017), and Gazzeri et al. (2015) each highlight innovations represented by category III codes that have been found to be of very limited utility.

3 Data

The primary source of data for this paper is a 20% random sample of Medicare claims for physician services (called the carrier file) from 2002-2017. This dataset includes encounter-level information on patient diagnoses, procedures performed, payments made by the patient and insurer, and many attributes of the provider and patient for millions of patients enrolled in

[4]Category III codes are meant to be archived after five years, but many codes are archived much later while others are archived earlier.

Table 1: Summary Statistics

	Mean	Std. Dev.
Uses Per Million Beneficiaries	366.93	1,489.02
Unique Patients Treated	1,447.19	6,833.64
Unique Providers Using	245.23	883.08
Payment per Procedure	\$496.04	\$1,752.26
Percentage of Claims Denied	76.79%	30.02%
Months Active in Data	62.61	39.90
Adopted?	0.606	0.489
AMA Promotion Status		
Promoted	0.429	0.496
Outstanding	0.294	0.456
Deleted	0.277	0.448
Distinct Procedures in Data	343	

Notes: Sample consists of all professional claims reporting Category III CPT codes for a 20% sample of Traditional Medicare beneficiaries from 2002-2017. An observation is a procedure.

Traditional Medicare.

In Table 1, I present procedure-level summary statistics on the utilization of these procedures. We see that there is wide variation in the overall level of utilization, as well as the in the average reimbursement for the procedure and the share of claims denied.

In addition to utilization information, this data includes whether the administrator paid or denied the claim for reimbursement. I use the administrator's propensity to deny claims for each new procedure in combination with incomplete posted coverage rules to infer the coverage status of each procedure for each administrator in each month. Appendix A gives more details on this classification process.

As it turns out, the administrators disagree about whether these procedures meet this standard quite regularly. This results in a great deal of variation in when procedures are covered by Medicare across different jurisdictions. Table 2 reports the share of procedure-months in which each administrative contractor covers the procedure fully or on a case-by-case basis. There is very wide variation across administrators in the propensity to cover new procedures, with the most generous administrator covering almost a quarter of the time while the least generous never

Table 2: Coverage by Administrative Contractor

	Percentage of Procedure-Months		
Administrator	Non-Covered	Case-by-Case	Covered
Novitas	76.5%	13.3%	10.2%
Cahaba	80.7%	14.4%	4.9%
NHIC	82.0%	13.5%	4.5%
Noridian	82.2%	6.4%	11.5%
Palmetto	82.3%	7.2%	10.5%
TrailBlazer	83.7%	11.4%	4.9%
HealthNow	86.3%	5.6%	8.1%
NGS	86.8%	7.7%	5.5%
FCSO	87.0%	8.6%	4.4%
WPS	87.1%	8.0%	4.9%
Pinnacle	87.3%	8.4%	4.3%
CGS	88.0%	6.9%	5.1%
Wheatlands	93.8%	3.2%	3.0%
Regence	98.0%	1.5%	0.5%
Group Health	98.4%	0.7%	0.9%
Triple-S	98.5%	0.8%	0.7%
BCBSMT	98.9%	0.7%	0.4%
Nationwide	99.0%	0.0%	1.0%
BCBSRI	100.0%	0.0%	0.0%
Overall	79.0%	14.8%	6.2%

Notes: Sample consists of all professional claims reporting Category III CPT codes for a 20% sample of Traditional Medicare beneficiaries from 2002-2017. An observation is a MAC-procedure-month tuple. The table reports the share of procedure-months at each coverage level separately for each MAC as well as collectively. Administrators are sorted in ascending order by non-coverage rate.

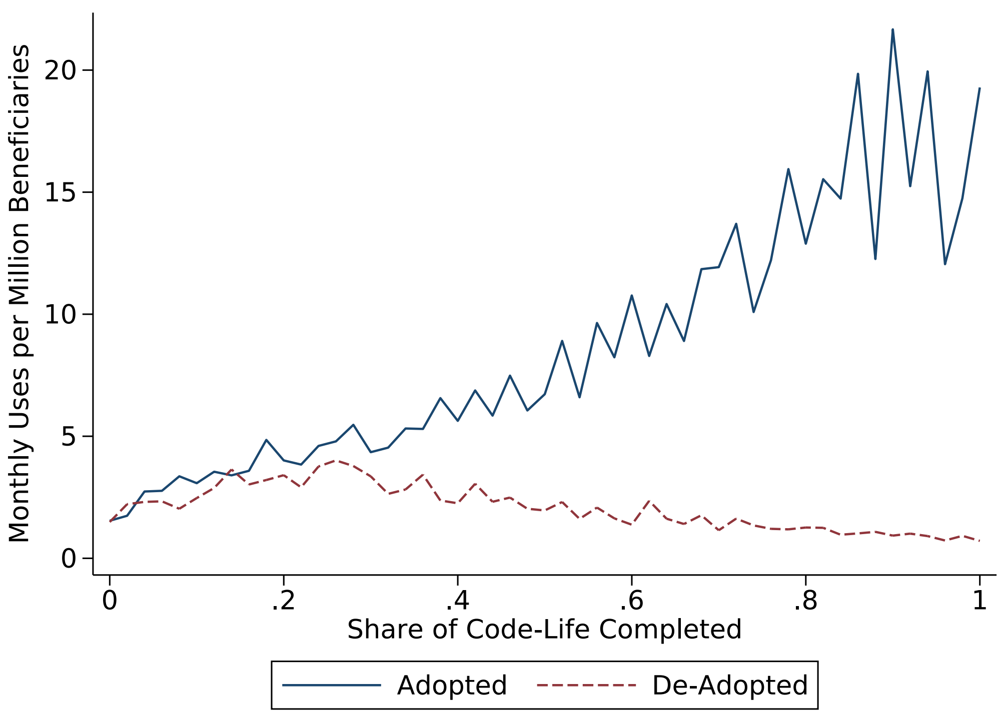
covers any procedure.[5] This variation results in 57% of procedure-months featuring some difference in coverage level across jurisdictions, while among procedures covered in any jurisdiction, it is covered in all jurisdictions only 7% of the time.

This prevalence of disagreement is important for a few reasons. First, it indicates meaningful uncertainty on the part of the administrators as to which procedures meet Medicare's standards for coverage. Second, it gives me the opportunity to identify learning spillovers from physicians in jurisdictions where the procedure is covered earlier to those that have to wait and learn until their local administrator grants coverage to the procedure.

[5]Many of the administrators listed in the table were not active for the entire sample period and so made coverage decisions on different new procedures. The model I present in Section 5 takes this issue into account, and in estimating this model I continue to find very large differences in evidentiary standards across MACs. See Figure 9 for this result.

In line with this uncertainty, we see that roughly half of these procedures fail to be adopted by the medical community. I classify procedures as adopted or de-adopted based on whether their use grows or falls over time.[6] Of the 343 procedures in my data, 208 see their use grow over time while the remaining 135 see their use fall, as is reported in Table 1. Figure 2 shows the average utilization of each of these classes of procedures over their time in my data. That 40% of the procedures fail is evidence that there is meaningful uncertainty about the value of these procedures. Furthermore, the fact that on average the procedures that are adopted and those that are de-adopted start out at roughly the same level of utilization is evidence that the *ex-ante* beliefs about the efficacy of these procedures are similar. In light of this, I will use this measure of adoption to indicate the *ex-post* value of each innovation.

Figure 2: Utilization by Adoption Status



Notes: The figure reports the average utilization per million beneficiaries of procedures whose use rises (adopted procedures) or falls (de-adopted procedures) over time. The horizontal axis scales the length of time the code covering each procedure is in the data to be equal to one.

Additionally, I supplement this data with hand-collected information on whether the American Medical Association has promoted or deleted the code covering each procedure. This com-

[6] Specifically, I estimate the model $Y_t = \beta_0 + \beta_1 T_t + \varepsilon_t$ separately for each procedure, where Y_t is the total nationwide uses of the procedure and T_t is the number of months since the introduction of the code covering the procedure. I classify each procedure as adopted if the estimate for β_1 is greater than zero and as de-adopted otherwise.

plementary measure of a procedure's *ex-post* value yields a similar classification of successful and unsuccessful innovations, with these shares reported in the final rows of Table 1. For codes about which the AMA has made a decision by January 2022, 95 have been deleted while 147 have been promoted. This success rate of 60.7% is nearly identical to the 60.6% success rate implied by the more comprehensive measure based on utilization trends.[7] Despite over 60% of procedures succeeding (regardless of the measure used), administrators are very hesitant to cover these procedures. As shown in Table 2, almost 80% of the time these procedures are non-covered, while even the most generous administrator only fully covers less than 12% of procedure-months.

4 Impact of Coverage

In this section I present evidence that local coverage decisions made by Medicare Administrative Contractors impact the adoption of new medical procedures. This result is important both in its own right and instrumentally for addressing the question of how these regulators should set their evidentiary thresholds. On its own, the impact of local Medicare coverage decisions has implications for geographic variation in medical practice and health outcomes (Fisher et al., 2003a,b; Finkelstein et al., 2016). Furthermore, there is very limited evidence on the influence of Medicare contractors on health care practice (League, 2023; Wilk et al., 2018; Carlson et al., 2009; Foote et al., 2008). Instrumentally, understanding the magnitude of the effect of coverage on utilization is key to understanding how Medicare administrators should set their evidentiary standards.

To shed light on this question, I estimate event studies of coverage changes. The primary specification I use is

$$Y_{pjt} = \beta_0 + \beta_1 Covered_{pjt} + \gamma_{pj} + \gamma_{pt} + \varepsilon_{pjt}, \tag{1}$$

where $Covered_{pjt}$ is an indicator for Medicare coverage of procedure p in jurisdiction j at time t, γ_{pj} and γ_{pt} are series of jurisdiction-by-procedure and procedure-by-month fixed effects, and ε_{pjt} is the econometric error term. Dependent variables Y_{pjt} are measures of utilization, including an indicator for any utilization within the jurisdiction-month, as well as the number of procedures performed per million beneficiaries. Because there are likely spillovers in utilization from jurisdictions in which coverage changes to those in which it does not, I also estimate a model without time fixed effects. In so doing, I compare the utilization of each procedure after coverage to its own earlier use in the same jurisdiction rather than comparing the relative changes in utilization across jurisdictions in which coverage does and does not change. Finally, because

[7] Throughout the paper, I use whether a procedure's use rises or falls over time as my primary measure of its success. In Appendix B, I discuss this decision in greater depth and demonstrate the robustness of my results to using AMA classifications instead.

Table 3: Effect of Coverage on Utilization

	(1) Use Rate	(2) Use Rate	(3) Any Uses	(4) Any Uses
Case-by-Case	4.894***	3.884*	0.0573***	0.0351***
	(1.160)	(1.547)	(0.00337)	(0.00315)
Covered	31.61**	19.98	0.129***	0.0618***
	(12.11)	(14.22)	(0.00792)	(0.00538)
Jurisdiction FEs	1	1	1	1
Time FEs	0	1	0	1
Dep. Var. Mean	4.867	4.867	0.0623	0.0623
Observations	1,240,776	1,240,776	1,240,776	1,240,776

Notes: An observation is a procedure-jurisdiction-month. Sample consists of monthly observations from 2002-2017. Use rate is the number of uses of the procedure per million beneficiaries. Regressions include jurisdiction-by-procedure and procedure-by-month fixed effects where indicated. Standard errors are clustered at the jurisdiction-by-procedure level. +, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

of concerns about the conventional two-way fixed effects estimator (Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021), in Appendix C I present additional results limiting the treatment window and employing the stacked regression estimator of Cengiz et al. (2019).

The results of estimation of Equation (1) are presented in Table 3. We see that following a change in Medicare coverage, utilization rises dramatically. When a procedure is fully covered it is used over 30 more times per million beneficiaries each month relative to when it is non-covered, representing an increase over 6 times greater than the mean utilization rate. Relative to jurisdictions in which coverage does not change, this increase is somewhat attenuated—potentially indicating spillovers to other jurisdictions—but is still very large.

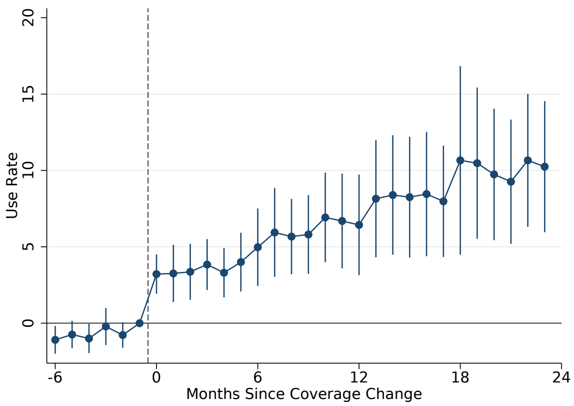
To understand the dynamic effect of coverage changes, I also estimate

$$Y_{pjt} = \sum_{e=-6}^{-2} \beta_e T_{pjt}(e) + \sum_{e=0}^{24} \beta_e T_{pjt}(e) + \gamma_{pj} + \gamma_{pt} + \varepsilon_{pjt}, \tag{2}$$

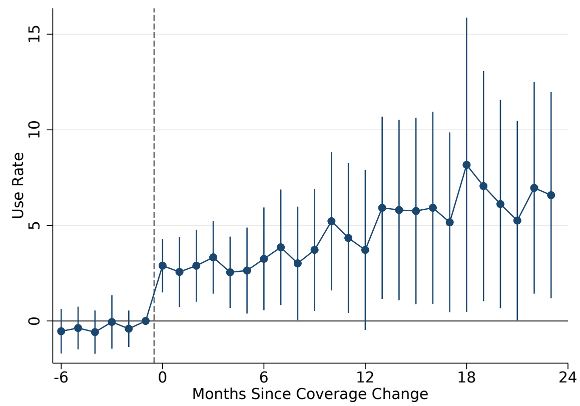
where $T_{pjt}(e)$ is an indicator for a jurisdiction being e months from a change in coverage. Figure 3 reports estimates of β_e for specifications with and without time fixed effects in the model as well as estimated using stacked regression. Across all three specifications, we see consistent results indicating a lack of differential trends in utilization prior to coverage being extended followed by utilization increasing dramatically immediately upon coverage and then gradually continuing to grow thereafter. That coverage has an immediate impact on utilization indicates that the initial lack of coverage prevented providers from performing procedures they otherwise would, while the following gradual increase in utilization is consistent with learning over time about the efficacy of the newly covered procedure. In the next section of the paper, I will more closely examine this and other potential causes of the growth in utilization following coverage.

Figure 3: Change in Utilization at Coverage Change

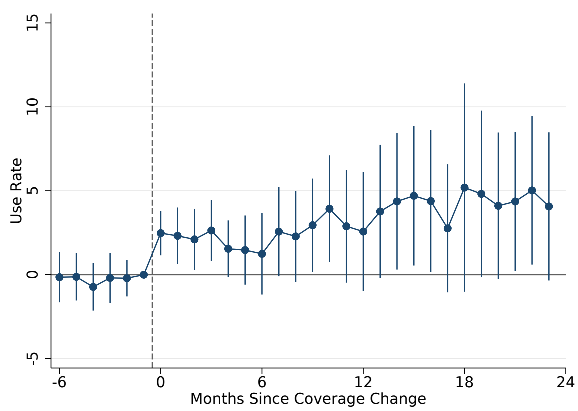
(a) No Time Fixed Effects



(b) Two-Way Fixed Effects



(c) Stacked Regression



Notes: The figures report estimates of β_e from Equation (2) for $e \in \{-6, \dots, 24\}$. An observation is a jurisdiction-procedure-month tuple. Panel (a) includes estimates without time fixed effects. Panel (b) uses the two-way fixed effects estimator. Panel (c) uses the stacked regression estimator. For all three specifications, only treated jurisdiction-procedure pairs in the data for the entire treatment window are included along with untreated observations. Error bars give the pointwise 95% confidence intervals. Standard errors are clustered at the jurisdiction-procedure level.

5 Evidence of Social Learning

Social learning is a phenomenon widely believed to be important in medicine. In opposing intellectual property rights in medical procedures, the AMA Code of Ethics notes that going back to the time of Hippocrates, the role of the physician has included being "a teacher who imparts knowledge of skills and techniques to colleagues, and a student who constantly seeks to keep abreast of new medical knowledge" (American Medical Association, 2010). Furthermore, this learning process generally occurs through channels accessible to the entire medical community, including word-of-mouth endorsements at national conferences, mass media, and notes in medical journals (McKinlay, 1981). In line with this sentiment, empirical research has documented that health care providers learn from the experiences of one another (Coleman et al., 1957; Allen et al., 2019; Soumerai et al., 1998; Zheng et al., 2010; Agha and Molitor, 2018). However, to my knowledge, no empirical research has investigated the global knowledge spillovers from each physician to the entire medical community that I document here.

To document social learning, I will exploit the unique opportunity granted by the variability in Medicare coverage rules along with the high degree of uncertainty in the value of the innovations I study. Separating social learning from other phenomena that may be driving the adoption or de-adoption of new innovations is generally difficult because the global spillovers across providers means there is rarely a control group that does not have access to the common pool of information. Similarly, adoption decisions are generally made in response to the beliefs of the potential adopters about the value of the innovation. To overcome these issues, I exploit variation in the incentives of providers to adopt new procedures, as well as heterogeneity in the underlying value of each innovation. Heterogeneity in the patterns of diffusion between successful and unsuccessful innovations is key to separating learning from other drivers of diffusion because while many potential drivers of adoption can predict greater adoption over time, very few of them can predict de-adoption. In contrast, social learning makes opposite predictions for the trends of adoption for innovations that are good or bad.[8] To that end, I will present heterogeneity in the effect of increased information by whether the procedure turned out to be of high or low value.

In this section, I present three pieces of evidence that social learning is important in this setting. The first is that there are spillovers in utilization from jurisdictions that see a change in coverage to jurisdictions that do not, consistent with providers responding to the experiences of providers in other jurisdictions. The second is that idiosyncratic increases in utilization cause persistent changes in utilization. For high value innovations, positive shocks to early use lead to persistently higher utilization as providers more quickly learn the value of the procedure while for low value innovations, positive shocks to early use have no such effect. The third piece of evidence is that the effect of coverage may differ by the size of the jurisdiction, with coverage having a qualitatively larger effect for larger jurisdictions in which the opportunity to generate quality signals is higher, with this effect flipping sign for de-adopted procedures.

[8] The idea that learning can be identified by a negative reaction to adverse information has also been used by Anand and Shachar (2011), who demonstrate that being exposed to advertisements for television shows lowers the likelihood of watching for viewers whose characteristics indicate the show is a bad match for them while having the opposite effect on viewers who are good matches.

5.1 Spillovers from Coverage

As shown in Section 4, Medicare coverage of a new procedure increases its utilization in jurisdictions in which it is covered. Social learning predicts that by increasing the utilization of the procedure, coverage will also decrease uncertainty about the value of the procedure for all providers. Importantly, this decreased uncertainty will occur not only for providers in jurisdictions in which the coverage rules are loosened but for all providers. Furthermore, while this reduction in uncertainty will increase utilization for procedures that are better than many providers believe, for low value procedures the additional utilization will reveal that the innovation is worse than many of the providers already using the procedure believe, leading utilization to fall. Thus, for successful procedures, social learning predicts that extending coverage will lead to increased utilization even in jurisdictions that are not subject to a change in local coverage while for unsuccessful procedures, the opposite will be the case.

To test this prediction, I use stacked regression to estimate

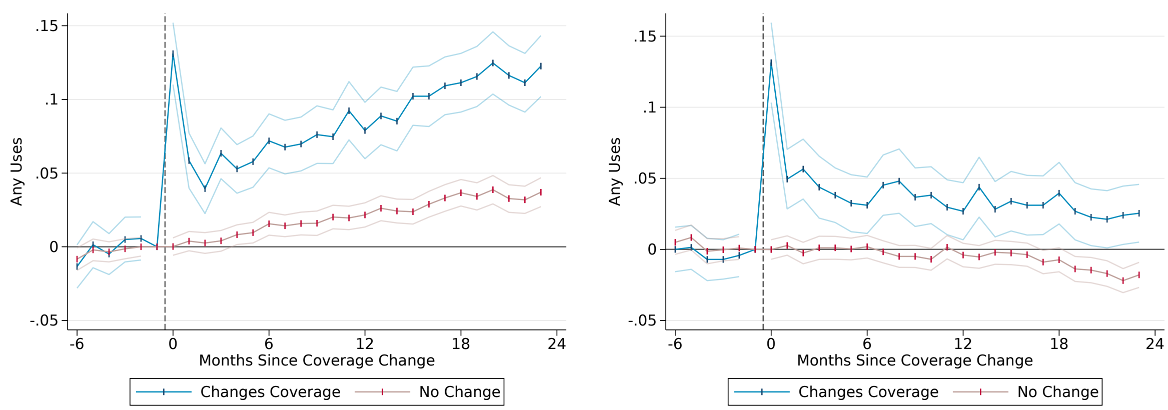
$$(3)\qquad Y_{pjtg} = \sum_{e=-6}^{-2} \beta_e T_{pjtg}(e) + \sum_{e=0}^{24} \beta_e T_{pjtg}(e) + \sum_{e=-6}^{-2} \tau_e E_{pjtg}(e) + \sum_{e=0}^{24} \tau_e E_{pjtg}(e) + \gamma_{pjg} + \varepsilon_{pjtg},$$

where $E_{pjtg}(e)$ is an indicator for being e months from treatment date $g = e + t$ for both jurisdictions j that see coverage change in month g for procedure p along with the control jurisdictions in which jurisdiction does not change during the event window. Estimation of this equation relies on the stacked regression estimator's explicit matching of jurisdictions in which coverage changes with suitable comparison jurisdictions in which coverage does not change to create many 2x2 difference-in-differences events. The series of τ_e coefficients thus give the time series change in utilization in jurisdictions in which coverage does not change as coverage in another jurisdiction changes. Under the assumption that the only thing determining time series variation in utilization is changes in beliefs about the efficacy of the procedure coming from social learning, τ_e identifies the knowledge spillovers from jurisdictions in which coverage changes to those in which it does not. As in Equation (2), β_e identifies the differential change in utilization in jurisdictions in which coverage changes and (under the assumption of parallel trends) the effect of coverage on utilization independent of the resulting generation of any universally accessible knowledge.

Figure 4a presents estimates of τ_e and $\beta_e + \tau_e$ for procedures that are *ex post* successful while Figure 4b presents estimates using a sample limited to those procedures whose use falls over time. Notice that for both groups of procedures, coverage increases utilization in jurisdictions in which coverage changes (the blue line) as providers' financial and regulatory incentives to use the new procedure become more favorable. However, the impact of coverage on jurisdictions in which coverage does not change differ by the underlying value of the procedure. For high-value procedures, use of the new procedure in these jurisdictions rises, while utilization falls for low-value procedures. This is consistent with providers everywhere updating their beliefs in response to the information generated by the increased utilization in jurisdictions in which coverage changes. Further supporting this interpretation are the facts that there are not pre-trends in the utilization of these procedures before the change in coverage and that the spillovers in utilization only occur gradually as the medical community's experience with the procedure grows.

Figure 4: Change in Utilization at Coverage Change for Treatment and Control Jurisdictions

(a) Adopted Procedures (b) De-Adopted Procedures



Notes: The figures report estimates of τ_e (in red) and $\beta_e + \tau_e$ (in blue) from Equation (3) for $e \in \{-6, \ldots, 24\}$. An observation is a jurisdiction-procedure-month-group tuple, where groups are defined by the stacked regression procedure defined in Appendix C. Panel (a) presents estimates for the sample limited to procedures whose use rises over time, while panel (b) presents estimates for procedures whose use falls over time. 95% confidence intervals are given by the translucent tickless lines of the relevant color. Standard errors are clustered at the group level.

5.2 Persistence in Shocks to Use

In the previous subsection, I used administrator coverage decisions as a source of variation in utilization that I could use to examine the impact of shocks to utilization in one jurisdiction on utilization in another. In this subsection, I use administrator coverage decisions to learn the beliefs of the medical community about the value of each procedure. Because administrators are tasked with covering procedures for which the available evidence supports its use, a greater share of administrators offering coverage of a procedure indicates that the evidence supporting its use is stronger. In Appendix D, I both prove this formally using the model of administrator behavior outlined in Section 6 as well as present empirical evidence that provider beliefs about a procedure's efficacy correspond to those of the Medicare administrators.

Table 4: Persistence of Idiosyncratic Shocks to Use

	(1) Use Rate	(2) Use Rate	(3) Use Rate
Lagged Use	1.349***	1.155***	0.945***
	(0.259)	(0.212)	(0.197)
Lagged Use × De-Adopted	-0.685**	-0.671**	-0.794***
	(0.256)	(0.205)	(0.212)
Months of Lag	6	12	24
Dep. Var. Mean	6.127	6.473	6.176
Observations	19,710	17,658	13,842

Notes: Estimates of β_1 and β_2 from Equation (4). An observation is a procedure-month pair. Standard errors are clustered at the procedure level. $^{+}$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

Using the share of administrators covering a procedure as a proxy for beliefs about its value, I can isolate idiosyncratic shocks to use and test their persistence over time. In particular, I can estimate heterogeneity in the persistence of these shocks to use by whether the procedure turns out to be of high or low value. In so doing, I can compare procedures that initially have the same support among administrators (i.e., about which the medical community has the same *ex-ante* beliefs). Social learning predicts that for procedures that are better than the medical community believes, idiosyncratic increases in utilization should lead to greater utilization later on as providers learn about the procedure's value more quickly while for procedures that are worse than expected, these positive shocks should not be persistent. To make this comparison, I estimate

$$Y_{pt} = \beta_1 Y_{pt}^{l} + \beta_2 Y_{pt}^{l} \times LowValue_p + \beta_3 ShareCov_{pt}^{l} + \beta_4 ShareCase_{pt}^{l} + \gamma_t + \gamma_p + \varepsilon_{pjt}, \tag{4}$$

where Y_{pt} is the utilization of procedure p at time t and Y_{pt}^{l} is utilization at time $t-l$, $LowValue_p$ is an indicator for whether procedure p is ultimately de-adopted, and $ShareCov_{pt}^{l}$ and $ShareCase_{pt}^{l}$ are the shares of administrators in period $t-l$ that cover the procedure fully or on a case-by-case basis, respectively.

Table 4 presents estimates of β_1 and β_2 from Equation (4). We see that for successful procedures, positive shocks to past use positively predict future use, while this is not true for de-adopted procedures. This is consistent with social learning: idiosyncratic shocks to use create unexpected signals of a procedure's value that affect later utilization.

In addition to exploiting general idiosyncratic shocks, I can also use idiosyncratic shocks of a particular type: those induced by the size of the jurisdictions in which the procedure is covered. As shown in Figure 5, there is significant variation in the share of beneficiaries covered by each administrator. Furthermore, while the Centers for Medicare and Medicaid Services has recently attempted to balance the workload across administrators (CMS, 2005), there remains significant variation over time as the Medicare Administrative Contractor jurisdictions were consolidated. These differences in the number of providers subject to the coverage determinations of each administrator can be used as an instrument for early utilization of a procedure. The intuition is that while the total amount of utilization of a procedure is affected by the *size* of the jurisdiction in which the procedure is covered, the underlying quality of the innovation is only related to the *share* of administrators that deem the procedure worthy of Medicare coverage. To implement this strategy, I estimate the equation

$$Y_{pt} = \beta_1 Y^l_{pt} + \beta_3 ShareCov^l_{pt} + \beta_4 ShareCase^l_{pt} + \gamma_t + \gamma_p + \varepsilon_{pjt}, \tag{5}$$

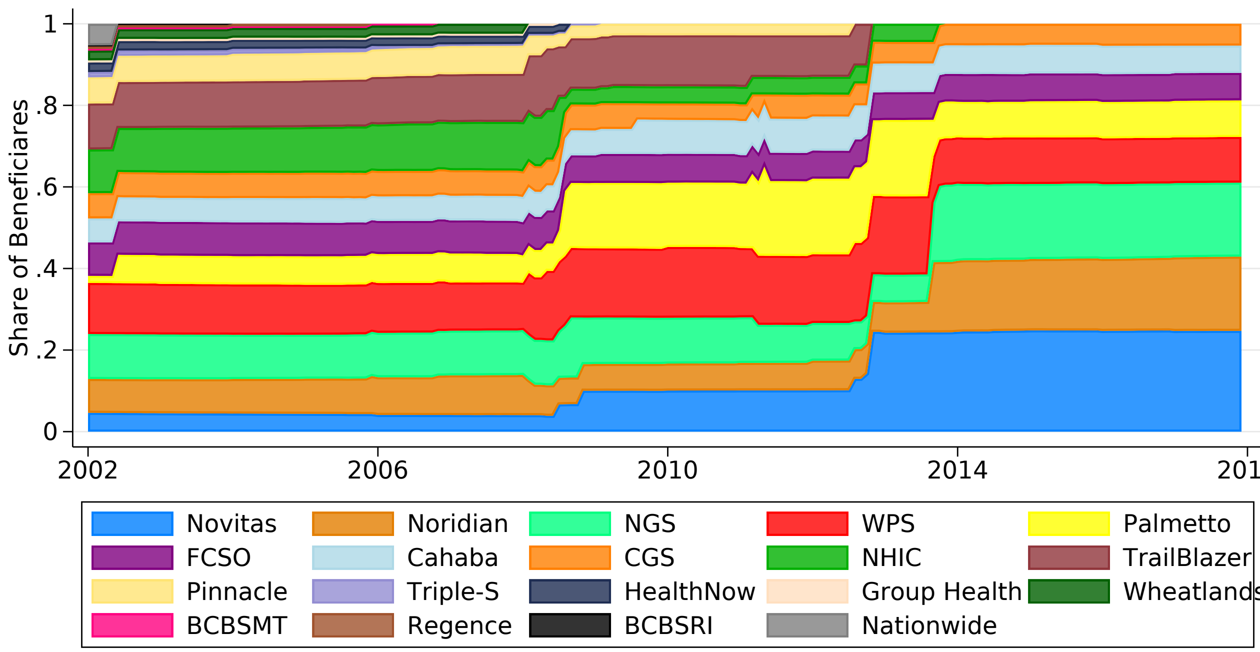
using two-stage least squares, where the first stage equation is given by

$$Y^l_{pt} = \alpha_1 PopCov^l_{pt} + \alpha_2 PopCase^l_{pt} + \alpha_3 ShareCov^l_{pt} + \alpha_4 ShareCase^l_{pt} + \delta_t + \delta_p + \tilde{\varepsilon}_{pjt}, \tag{6}$$

where $PopCov^l_{pt}$ and $PopCase^l_{pt}$, the excluded instruments for lagged utilization, are the number of beneficiaries living in jurisdictions in which the procedure is covered fully or on a case-by-case basis, respectively. For this instrumental variables strategy to be valid, it must be the case that $PopCov^l_{pt}$ and $PopCase^l_{pt}$ are uncorrelated with ε_{pjt}. In this context, this entails assuming that the size of the jurisdictions in which the procedures are covered earlier is only related to later use of the procedure through its effect on earlier utilization. In addition to this exclusion restriction, the other key assumption is relevance: that the size of the jurisdictions in which the procedure is covered actually affects the total use of the procedure. As shown in Figure 6, the number of beneficiaries living in jurisdictions in which the procedure is covered is positively related to the total uses of that procedure even conditional on the share of administrators deeming the procedure fit for coverage.

Table 5 gives the two-stage least squares estimates of β_1 separately for adopted and de-adopted procedures. As before, we see that for successful innovations, more early utilization positively affects later utilization, while the opposite is true for unsuccessful procedures. Here the shocks to early use are isolated to those caused by the size of the jurisdictions in which it is covered. Again, the heterogeneity in the persistence of early shocks by whether the procedure is eventually found to be effective is key evidence that the mechanism is social learning, but it is also important for validating the identification strategy. That for de-adopted procedures, the size of the early-covering administrative jurisdictions is negatively related to its later use refutes the potential worry that the determinations of contractors administering larger areas hold greater sway. In these cases, utilization fell more quickly despite (or through its impact on early utilization and information availability, because of) the endorsement by the administrators

Figure 5: Share of Population of Each Jurisdiction



Notes: Figure presents the share of beneficiaries living in jurisdictions administered by each administrative company in each month.

of large jurisdictions.

5.3 Heterogeneity by Jurisdiction Size

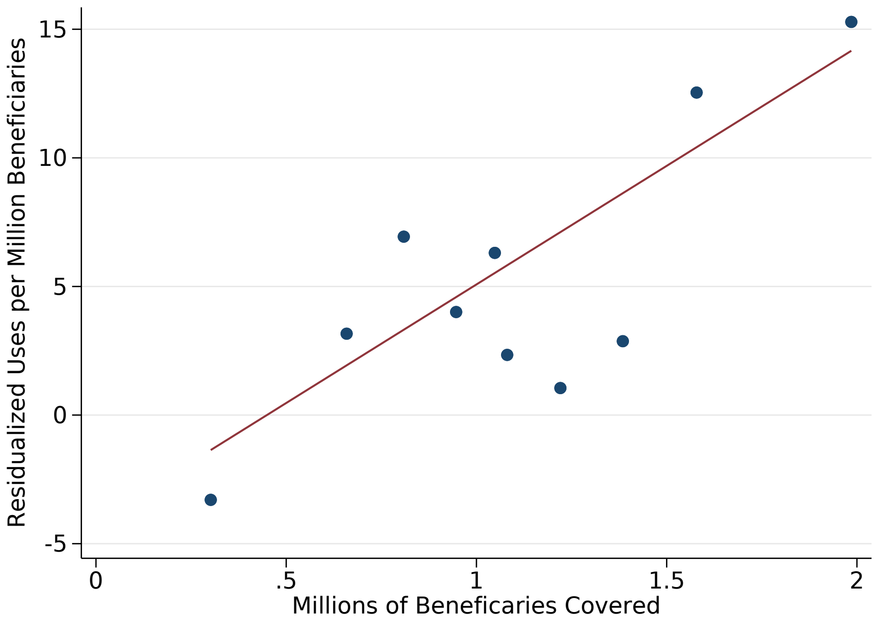
The third piece of evidence of social learning I present is suggestive evidence of heterogeneity in the size of the effect of coverage by the size of the jurisdiction. In particular, I show that when coverage is extended to a new procedure, the response in larger jurisdictions is qualitatively more sensitive to the underlying value of the innovation. That is, for successful innovations, larger jurisdictions adopt the procedure more rapidly than smaller jurisdictions, while for unsuccessful ones, they de-adopt them more rapidly. This phenomenon can be explained by social learning as larger jurisdictions create more signals of a procedure's quality.

I estimate the equation

$$(7)\quad Y_{pjt} = \sum_{e=-6}^{-2} \beta_e T_{pjt}(e) + \sum_{e=0}^{24} \beta_e T_{pjt}(e) + \sum_{e=-6}^{-2} \phi_e T_{pjt}(e) \times Size_{jt} + \sum_{e=0}^{24} \phi_e T_{pjt}(e) \times Size_{jt} + \gamma_{pj} + \varepsilon_{pjt},$$

where $Size_{jt}$ is the total population covered by the administrator. The set of ϕ_e coefficients are the coefficients of interest as they give the differential change in utilization depending on the covering jurisdiction's size. Notice that there are no time fixed effects, so this specification

Figure 6: Relationship Between Covered Population and Use



Notes: Figure presents the average monthly utilization per million beneficiaries for each of nine quantiles of the total number of beneficiaries living in jurisdictions in which the procedure is covered fully or on a case-by-case basis, adjusted for differences in the share of administrators covering the procedure. An observation is a procedure-month pair. The red line gives the predicted values from estimates of Equation 6.

compares how utilization changes at coverage between larger and smaller jurisdictions but does not compare jurisdictions in which coverage changes to those in which it does not. This is important because with global knowledge spillovers, the extra information generated when a procedure becomes covered in a larger jurisdiction would affect utilization in all other jurisdictions as well, meaning there would likely be no differential change relative to the jurisdictions without coverage changes depending on the size of the covering jurisdiction.

Figure 7 presents estimates of ϕ_e in Equation (7) separately for procedures that are eventually adopted and those that are not. We see that for adopted procedures, utilization grows more following a loosening of the coverage rules for larger jurisdictions compared to smaller ones, although this difference is not statistically significant. After there being no discernible difference immediately following the change in coverage, the difference grows as time passes. This is what would be expected if social learning is driving the difference: the additional information generated by the additional utilization accumulates slowly and builds on itself to increasingly lead providers to adopt the procedure. For de-adopted procedures, we see the opposite pattern. For these procedures, larger jurisdictions see less utilization after coverage. This too is consistent with the higher level of use in the larger jurisdiction generating more negative signals of the procedure's quality, leading providers to become less likely to perform the procedure.

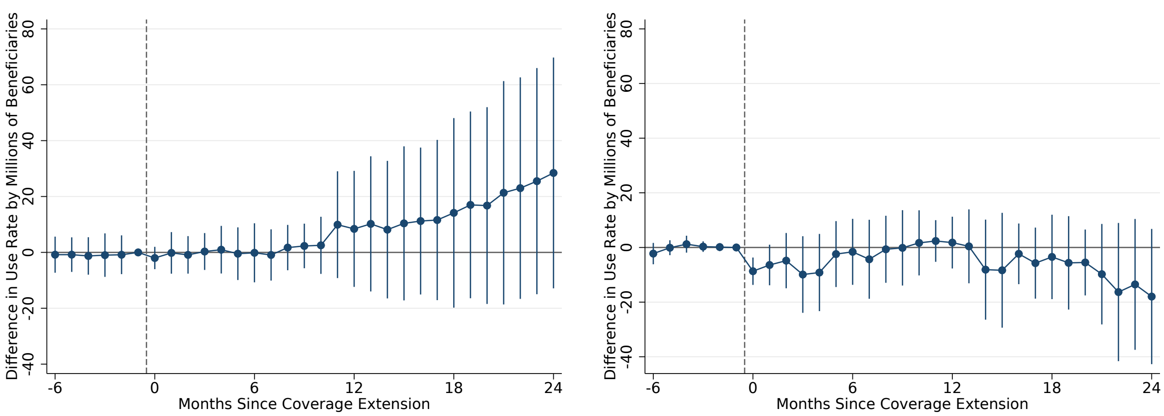
Table 5: Persistence of Idiosyncratic Shocks to Use, Jurisdiction Size IV

	Adopted		De-Adopted	
	(1) Use Rate	(2) Use Rate	(3) Use Rate	(4) Use Rate
Lagged Use	3.156**	3.088**	-2.582	-3.352
	(0.962)	(0.939)	(3.868)	(4.664)
Months of Lag	24	24	24	24
Current Cov. Cont.	0	1	0	1
Dep. Var. Mean	9.600	9.600	2.141	2.141
Observations	7,488	7,488	6,354	6,354

Notes: Two-stage least squares estimates of β_1 from Equation (5). An observation is a procedure-month pair. Dependent variable is the number of uses of the procedure per million Traditional Medicare beneficiaries. Models in columns (2) and (4) include controls for the contemporary coverage status of the procedures. Standard errors are clustered at the procedure level. $^{+}$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

Figure 7: Change in Utilization at Coverage Change for Treatment and Control Jurisdictions

(a) Adopted Procedures (b) De-Adopted Procedures



Notes: The figures report estimates of ϕ_e from Equation (7) for $e \in \{-6, \ldots, 24\}$. An observation is a MAC-procedure-month tuple. Dependent variable is the number of uses of the procedure per million Traditional Medicare beneficiaries. Panel (a) presents estimates for the sample limited to procedures whose use rises over time, while panel (b) presents estimates for procedures whose use falls over time. Error bars give the pointwise 95% confidence intervals. Standard errors are clustered at the MAC-procedure level.

5.4 Other Potential Explanations

While I have presented multiple pieces of evidence consistent with social learning being important in this setting, this does not rule out the presence of other important drivers of innovation adoption and diffusion. In this subsection, I discuss a few of the most plausible alternative explanations and explain why I do not believe them to be first-order issues for the diffusion of these innovations.

The first alternative driver of diffusion in this context is technological change. Perhaps the reason I observe increasing use over time for many of these procedures is because the actual value of the procedure is increasing, either because of improvements exogenous to the amount of utilization of the procedure or as a result of increased use as providers tinker with how to best perform the procedure. A few institutional details and patterns in the data indicate that changes to the value of the procedure are unlikely to occur. First, improvement in the value of a procedure could not explain the widespread de-adoption of many of the procedures that I observe. Second, many of these procedures are related to medical devices, which to be altered require approval from the Food and Drug Administration, making rapid improvement difficult.[9] Finally, the codes created by the AMA to classify these procedures are quite specific and are revised to reflect evolving ways of doing procedures, but this happens only very infrequently, indicating that the nature of the new procedures is quite stable once they are introduced.[10] While these points cast doubt on change in the actual value of the innovations being first-order, the possibility of improvement cannot be categorically ruled out. In light of this, I should point out that if over time providers are learning how best to perform a procedure and this collective improvement process is a result of the wider employment of the procedure, then the implications for administrators' coverage decisions would be largely the same as under social learning: there is a positive information externality from the wider use of the procedure, so evidentiary thresholds should be lower than otherwise.

Another alternative explanation is learning from clinical trials. The purpose of clinical trials is to generate information that influences medical practice, and there is extensive literature showing that clinical trials do just that (Depalo et al., 2019; Avdic et al., 2018). Furthermore, unfavorable clinical trial results can generate the de-adoption patterns I observe (Grennan and Town, 2020). However, clinical trials are likely only a minor determinant of the spread of the new procedures I study. This is because clinical trials of these procedures are surprisingly rare. As noted by Dranove et al. (2021), from 2008 to 2017 only 20% of the procedures approved by the American Medical Association were supported by randomized controlled trials. Because so few of these procedures are subject to randomized trials, there is limited scope for these trials to impact their diffusion.

So in light of the lack of evidence supporting the factors considered here along with the extensive evidence consistent with social learning, I conclude that social learning is a primary driver of the diffusion of new medical procedures. To that end, in the next section, I write down a model of provider procedure adoption that focuses on social learning in order to both estimate

[9] Dranove et al. (2021) find that just over half of approved category III CPT applications from 2008 and 2017 were associated with new devices.

[10] Out of the 337 codes operative from 2002 to 2014, there were only 7 total revisions to category III CPT codes.

the size of the social learning externality as well as to understand how it impacts the optimal evidentiary standard for Medicare coverage of new medical procedures.

6 Model

Having demonstrated the importance of social learning in this context, I now build a model of innovation adoption and diffusion around this phenomenon in order to assess the current evidentiary standards for coverage used by Medicare administrators. The model I employ for this purpose is similar to that used by a number of studies of learning by health care providers. For example, Coscelli and Shum (2004) model the adoption of a new anti-ulcer drug by Italian physicians and consider how providers learn about the drug's efficacy for different types of patients. Other studies highlight the often dynamic nature of technology adoption decisions (Crawford and Shum, 2003; Ferreyra and Kosenok, 2011) and the role of human capital accumulation and depreciation (Gowrisankaran et al., 2006; Hockenberry and Helmchen, 2014; Gong, 2017) in patient outcomes. These studies form an important foundation for the model of physicians as Bayesian learners updating their beliefs about the value of a new innovation that I employ in this paper.

I depart from these existing models in a number of important ways. First, I incorporate spillovers in knowledge across providers to account for social learning. I am able to do this because, unlike previous researchers, I am able to exploit an exogenous shifter in the national body of knowledge available to a physician at the time he or she first utilizes the innovation: variation in Medicare coverage. In line with the evidence I present of social learning in this context, I estimate that informational spillovers from a doctor to the wider medical community meaningfully impact utilization. Next, rather than examining a single innovation, I consider an entire class of innovations, which is crucial to asking the policy question of how generous coverage should be in general rather than for an innovation that *ex-post* was successful or not. This is particularly important in light of the large degree of uncertainty and heterogeneity in innovation value that I find. Finally, my incorporation of Medicare coverage decisions lends my results to natural policy counterfactuals. I find that Medicare could improve welfare either by reducing or increasing coverage.

I model the physician's decision to adopt a new procedure as a static one depending on the physician's beliefs about the value of the new procedure, the local coverage rules, the physician's unobservable type. Specifically, the utility of physician i from adopting the new procedure p at time t is given by

$$U_{ipt} = \mathbb{E}_{ipt}[\delta_p^*] - \beta_{1p} Case_{ipt} - \beta_{2p} Noncov_{ipt} - X_{ip},$$

where X_{ip} is the physician's ability to perform procedure p, $Case_{ipt}$ and $Noncov_{ipt}$ are indicators for whether procedure p is covered in the physician's jurisdiction at time t on a case-by-case basis or fully non-covered, and $E_{ipt}[\delta_p^*]$ is the belief of physician i about the value of procedure p

relative to the outside option with value normalized to zero.

The key difficulty for physicians in this model is that rather than being able to observe δ_p^* directly, they are uncertain about the quality of the new procedure and so must base their adoption decision on their expectation about its value. The value of the outside option is normalized to zero, and physicians face no uncertainty about their type, so a physician will adopt the new procedure for all eligible patients the physician treats if and only if $\mathbb{E}_{ipt}[\delta_p^*] > X_{ip} + \beta_{1p} Case_{ipt} + \beta_{2p} Noncov_{ipt}$. This means that the higher physicians believe the quality of the new procedure to be, the more will adopt it conditional on the coverage rules they face. Furthermore, physicians maximize per-period utility in a myopic way: they adopt the procedure when they believe adoption will give them positive utility in the current period and do not consider how their adoption decision might affect their future beliefs.

The initial beliefs of physicians are distributed normally (independent of physician type) with a mean of δ_{0p} and a standard deviation of $\sigma_{\delta p}$. Note that δ_{0p} may or may not equal δ_p^* (physicians may be optimistic or pessimistic) and that $\sigma_{\delta p}$ captures the level of initial uncertainty about the efficacy of the new procedure.

Each time a physician utilizes the new procedure, it creates a noisy signal of the true value distributed normally with a mean of the true value δ_p^* and a standard deviation $\sigma_{\nu p}$. Importantly, this signal is observable not only to the physician performing the procedure but all physicians. The entire medical community is thus able to learn from the experience of each physician and socially learn about the true value of the new procedure. The size of this social learning externality depends crucially on the precision of the signals generated through the procedure's use such that the less noisy the signal (the smaller is $\sigma_{\nu p}$) the more able the medical community is to aggregate the information generated by each discrete use of the procedure. To model this signal, I assume that each provider receives an independent signal each time the procedure is performed. That is, the medical community as a whole receives a distribution of signals centered around the true value of the procedure each time it is utilized.

Physicians update their beliefs about the efficacy of the procedure in light of the signals generated through its use according to Bayes rule. This means that after receiving n signals $\{\nu_{ipk}\}_{k\in\{1,\ldots,n\}}$, the beliefs of physician i are distributed normally with mean μ_{ipn} and standard deviation σ_{ipn}, where

$$\sigma_{ipn} = \frac{\sigma_{\nu p}\sigma_{\delta p}}{\sqrt{\sigma_{\nu p}^2 + n\sigma_{\delta p}^2}} \qquad \text{and} \qquad \mu_{ipn} = \sigma_{ipn}^2 \left(\frac{\delta_{0ip}}{\sigma_{\delta p}^2} + \sum_{k=1}^{n} \frac{\nu_{ipk}}{\sigma_{\nu p}^2} \right).$$

Notice that all physicians—regardless of their initial beliefs—are equally uncertain about the value of the procedure at any point in time, but their beliefs about the value of the procedure differ because of their different priors and the different values of the signals received. That is,

$\sigma_{ipn} = \sigma_{all,pn}$ for all i, while μ_{ipn} is distributed normally with mean μ^a_{pn} and standard deviation σ^a_{pn}, where

$$\sigma^m_{pn} = \frac{\sigma_{\nu p}\sigma_{\delta p}}{\sqrt{\sigma^2_{\nu p} + n\sigma^2_{\delta p}}} = \sigma_{all,pn} \qquad \text{and} \qquad \mu^m_{pn} = \sigma^2_{all,pn}\left(\frac{\delta_{0p}}{\sigma^2_{\delta p}} + \frac{n\delta^*_p}{\sigma^2_{\nu p}}\right).$$

Proofs that the values of μ_{ipn} are so distributed are given in Appendix E.

There are two types of patients: those for whom the new treatment is appropriate and those for whom it is not. Physicians are able to observe patients' types, so if the physician has adopted the procedure, the new treatment will be used for all and only patients for whom it is appropriate. The issue is that patients should be treated only by physicians able to perform the procedure. So the welfare loss from uncertainty comes from physicians unable to effectively employ the procedure performing it and physicians who are able foregoing the opportunity to use it. More explicitly, the utility of patient j treated using the new procedure p by provider i at time t is given by

$$V_{iptj} = \eta_{pj}(\delta^*_p - X_{ip}),$$

where η_{pj} is an indicator equal to 1 if the treatment is appropriate for patient j and 0 otherwise and X_{ip} is physician i's ability to perform procedure p. η_{pj} is a Bernoulli random variable with mean η_p.

Similar to physicians, administrators are also initially uncertain about the value of the new procedure and make coverage decisions based on their beliefs about its value. Each administrator a sets its evidentiary standard for coverage s_{ap} for procedure p according to

$$s_{ap} = \theta_a + \mu_p + \varepsilon_{ap},$$

where θ_a and μ_p are administrator and procedure fixed effects and ε_{ap} is a mean-zero error term. The administrator will cover the new procedure if the evidence supporting its use e_{tp} is greater than s_{ap}.

6.1 Identification and Estimation

The unknown parameters in my model of provider innovation adoption that I must estimate are β_{1p}, β_{2p}, δ^*_p, $\sigma_{\delta p}$, $\sigma_{\nu p}$, η_p, and δ_{0p}. Briefly, β_{1p} and β_{2p} are identified by differences in utilization across jurisdictions with differences in coverage within the same time period. δ^*_p is identified by projecting the trend in utilization to its steady state. $\sigma_{\delta p}$ and $\sigma_{\nu p}$ are identified by how tightly the convergence of utilization to its steady state level tracks utilization, including utilization in other jurisdictions coming from changes in the coverage rules in those jurisdictions. δ_0 is identified by the difference in utilization of the new procedure when it is first introduced and the end of the sample period. Finally, η_p is identified by the share of patients treated using the new procedure by physicians who have adopted the procedure. I assume physician types X_{ip} are normally distributed and normalize the mean and standard deviation of the distribution corresponding to each procedure to be zero and one, respectively.

In order to ensure identification of all model parameters for each procedure, I limit the estimation sample to procedures for which I observe both coverage and non-coverage. As discussed above, without changes in coverage of the procedure, the model is not identified. Furthermore, this restriction should be thought of as limiting the sample to those for which the contractors could have plausibly decided to have a different coverage decision. This restriction results in a sample of 195 procedures.

I estimate the parameters of the physician adoption decision by maximum likelihood estimation. The log-likelihood function is given by

$$\mathcal{L}(\theta_p|X_{cov,pt}) =$$

$$\sum_{t\in\{1,\dots,T\}} \sum_{cov\in\{-1,0,1\}} \log\left(\binom{g_{cov,pt}}{y_{cov,pt}}\right) + y_{cov,pt}\log(\alpha_{cov,pt}\eta_p) + (g_{cov,pt} - y_{cov,pt})\log(1-\alpha_{cov,pt}\eta_p),$$

where

$$\alpha_{cov,pt} = 1 - \Phi\left(-\frac{\mu^m_{pn} - \beta_{1p}Case_{ipt} - \beta_{2p}Noncov_{ipt}}{\sqrt{\left(\sigma^m_{pn}\right)^2 + 1}}\right).$$

Note that n_{pt} is the total previous uses of procedure p prior to time t, $g_{cov,pt}$ is the total number of patients in the consideration sets of physicians in jurisdictions with coverage level $cov \in \{0,1\}$, and $y_{cov,pt}$ is the total number of patients treated using procedure p at time t in jurisdictions with coverage level cov.[11] I define the consideration set of patients potentially suitable for the treatment to be all patients with a primary diagnosis shared by any patient ever treated with the procedure who is treated by a physician whose specialty is that of any physician who ever uses the procedure.

The model of administrator behavior can be estimated separately from the provider learning model. In the model of administrator behavior, evidence e_{tp} supporting procedure p at time t can be parameterized as a procedure-month fixed effect. Under the assumption that ε_{ap} is type-one extreme value, θ_a can be estimated using logistic regression where the coverage probability is given by

$$P(Cover_{atp}) = \frac{\exp(\theta_a + \gamma_{tp})}{1 + \exp(\theta_a + \gamma_{tp})} \tag{8}$$

where $Cover_{atp}$ is an indicator for whether administrator a fully covers procedure p at time t and

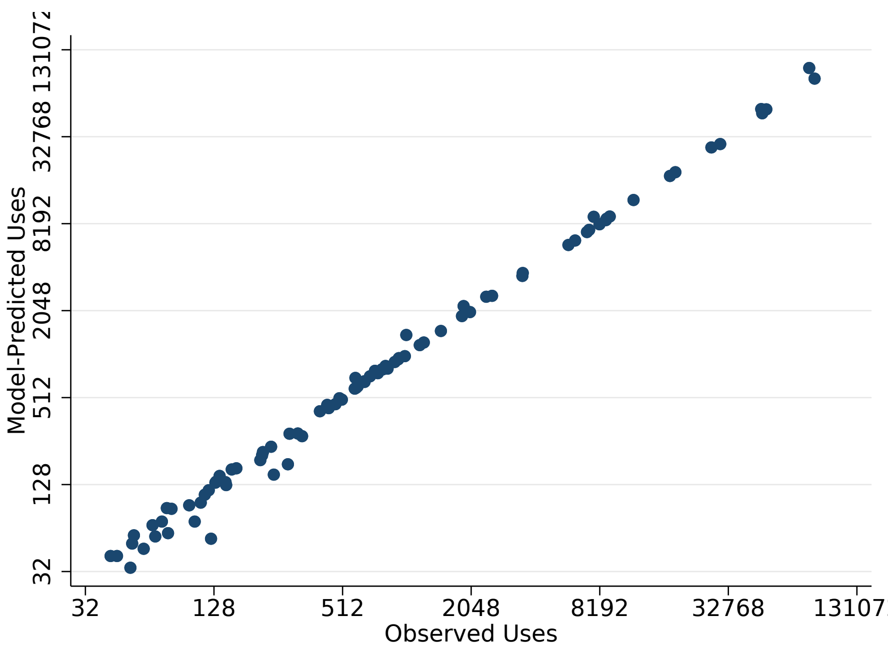
[11] The derivation of this likelihood function is presented in Appendix E.

γ_{tp} is a series of procedure-by-month fixed effects. Thus the administrator-specific propensities to cover new procedures are identified by within-procedure-month differences in coverage.

6.2 Estimation Results

First, I assess the fit of my model to the data. Figure 8 below plots the observed and model-predicted level of utilization for each procedure in the sample. Note that with few exceptions, the predicted uses of the procedure are nearly identical to the observed level of utilization.

Figure 8: Observed and Predicted Levels of Utilization

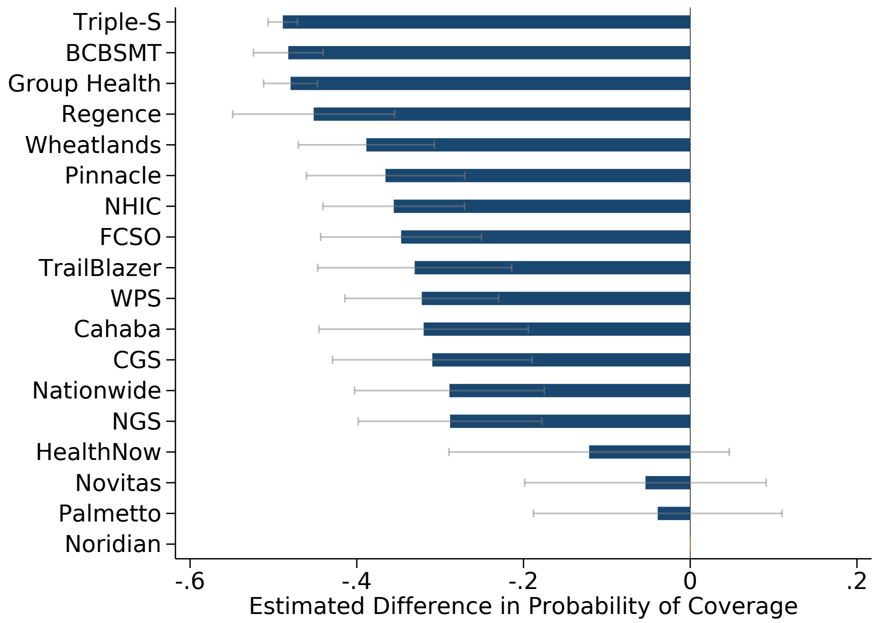


Notes: The figure reports the observed uses of each procedure along with the model-predicted number of uses for each procedure in the structural estimation sample. An observation is a procedure. Sample is limited to the 195 procedures for which I observe both coverage and non-coverage in the data.

Next, I report the estimated differences in the probability of coverage for each administrator for a marginal new procedure. Figure 9 reports these differences relative to Noridian, the administrator most generous with coverage of new procedures. I find wide heterogeneity in the propensity of each administrator to cover new procedures. These differences highlight the wide range of plausible evidentiary standards for the coverage of new procedures and make clear the uncertainty of policymakers about the appropriate standard.

Table 6 reports summary statistics of the estimated parameters for each procedure. These results suggest that relative to having the procedure fully covered, having the procedure covered on a case-by-case basis makes the procedure somewhat less appealing to providers on average, while having it completely non-covered makes it much less appealing. In particular, I estimate

Figure 9: Estimated Differences in Coverage Probabilities Across Administrators



Notes: The figure reports estimates of θ_a from Equation (8) transformed to marginal effects. An observation is a MAC-procedure-month tuple. Dependent variable is an indicator for full coverage of the procedure. Error bars give the 95% confidence intervals. Standard errors are clustered at the MAC-procedure and procedure-month levels.

that β_{2p} is positive for 89% of procedures is never statistically significantly negative at the 95% confidence level. These results are consistent with the reduced form evidence presented in Section 4.

These results indicate there is significant uncertainty about the value of each procedure. When a procedure is introduced, the distribution of beliefs across providers is 25 times wider than the distribution of ability to perform the procedure, on average. Furthermore, the median provider is able to correctly distinguish a new procedure's value from the incumbent procedure at the 95% confidence level only 13% of the time. And while additional utilization does generate meaningful information about the procedure's value, the standard deviation of this signal is almost 20 times greater than that of physicians' priors. This means that it takes 1,243 subsequent signals to match the precision of the median physician's prior.

My estimates of η_p indicate that among patients in the consideration set for being treated with the new procedure, the new procedure is appropriate 8.9% of them on average. In Appendix F, I present the full distribution of the estimates of η_p, showing that for some procedures, the value of η_p is much higher. In the same appendix, I also validate this distribution by estimating η_p separately from all other model parameters under a behavioral assumption.

Interpreting the δ parameters, we see that on average these new procedures are better than

Table 6: Summary Statistics of Parameter Estimates

Parameter	Mean	Std. Dev.
β_{1p}	3.293	32.18
β_{2p}	19.40	49.61
$\sigma_{\delta p}$	25.46	33.85
$\sigma_{\nu p}$	501.6	96.96
δ_p^*	10.45	99.34
δ_{0p}	-29.40	51.14
η_p	0.089	0.254

Notes: Estimates of structural model parameters. An observation is a procedure. Sample is limited to the 195 procedures for which I observe both coverage and non-coverage in the data.

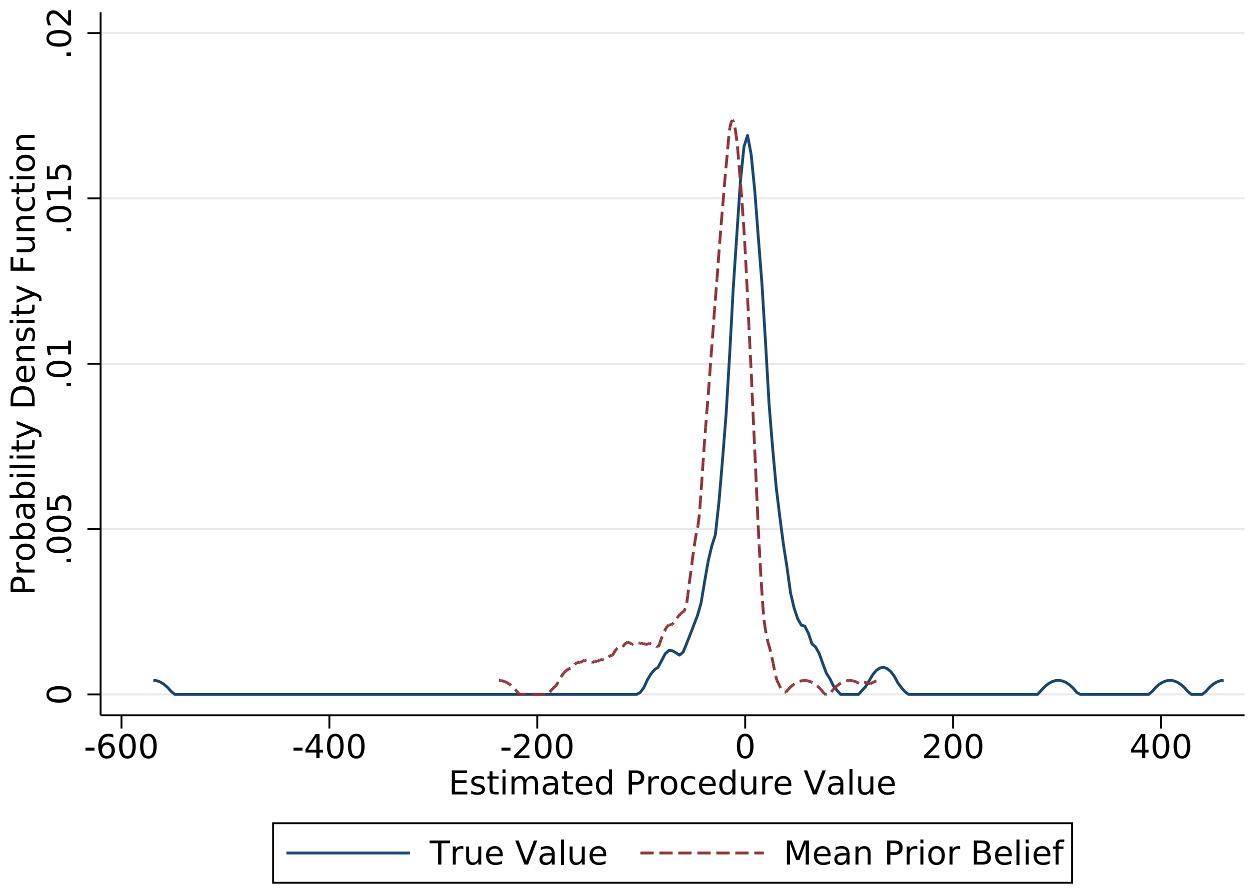
the outside option although providers initially believe they are worse. The estimate of δ_p^* is positive for 58.2% of procedures, meaning that while the new procedures are better on average and the median innovation is an improvement over the incumbent procedure, a significant portion of procedures are worse than the outside option.

The full distributions of providers' initial beliefs and the true value of each procedure are presented in Figure 10. The distributions are quite similar, although the modal prior is somewhat lower than the modal true value. The distribution of priors being below that of the true values is consistent with either pessimism or risk aversion on the part of providers. Insofar as the distributions have the same shape, it is consistent with physician priors being drawn from the same distribution as the true values, albeit with a level shift from pessimism or risk aversion. Interpreting the difference as being attributable to risk aversion, these distributions are consistent with rational expectations on the part of providers.

Regardless of the similarity of the distributions of the priors and true procedure values, I estimate that physicians are overconfident in their beliefs. For example, for only 59.3%, 51.6%, and 39.8% of procedures does the median physician have the true procedure value in their 99%, 95%, and 90% confidence intervals of beliefs, respectively. Furthermore, there is little correlation between physicians' prior beliefs and the true value of the procedure, with a correlation coefficient between δ_p^* and δ_{0p} of 0.05. The median physician is pessimistic about 76.9% of procedures and only significantly optimistic for 4.4%.[12]

[12]By "significantly", I mean the median physician's prior 95% confidence interval does not contain the true procedure value.

Figure 10: Estimated Distribution of Physician Priors and True Procedure Values



Notes: The solid blue line reports the estimated probability density function of δ^*_p, while the dashed red line reports the estimated probability density function of δ_{0p}. An observation is a procedure. Sample is limited to the 195 procedures for which I observe both coverage and non-coverage in the data.

6.3 Welfare

With the model parameters in hand, I can assess the welfare impact of each new procedure as well as assess the welfare consequences of counterfactual policies. Because the welfare to patient j from being treated with the outside option is 0, total welfare $W_{cov,pt}$ from procedure p under coverage regime cov at time t is given by

$$W_{cov,pt} = y_{cov,pt}\left(\delta^*_p - \mathbb{E}[X_{ip}|U_{ipt} > 0]\right) = y_{cov,pt}\left(\delta^*_p - \mathbb{E}[X_{ip}|X_{ip} < \Gamma_{cov,pt}]\right)$$

where $\Gamma_{cov,pt} \equiv \mu^a_{pn} - \beta_{1p} Case_{ipt} - \beta_{2p} Noncov_{ipt}$ for $a \in \{s, m\}$. Because X_{ipt} is standard normally distributed, we have that

$$\mathbb{E}[X_{ip}|X_{ip} < \Gamma_{cov,pt}] = -\frac{\exp\left(-\frac{\Gamma^2_{cov,pt}}{2}\right)}{\sqrt{2\pi}\left(1 - \Phi(\Gamma_{cov,pt})\right)},$$

so the total welfare from procedure p, $\mathcal{W}_p$, is given by

$$\mathcal{W}_p = \sum_{t\in\{1,\dots,T\}} \sum_{cov\in\{-1,0,1\}} W_{cov,pt} = \sum_{t\in\{1,\dots,T\}} \sum_{cov\in\{-1,0,1\}} y_{cov,pt} \left(\delta^*_p + \frac{\exp\left(-\frac{\Gamma^2_{cov,pt}}{2}\right)}{\sqrt{2\pi}\left(1 - \Phi(\Gamma_{cov,pt})\right)} \right).$$

I consider welfare under 4 scenarios. First, I estimate the level of welfare currently. Then I consider welfare under counterfactual coverage rules, including universal coverage and non-coverage of new procedures. I also estimate the possible welfare gains from contractors covering all and only procedures that are better than their outside options ($Case_{ipt} = Noncov_{ipt} = 0 \Leftrightarrow \delta^*_p \geq 0$).[13] In order to speak to the welfare costs of uncertainty and the value of physician learning, I estimate welfare under each of these scenarios under the status quo information environment as well as in the cases that physicians were able to perfectly observe the value of each new procedure and were they unable to learn from the experience of others. As a benchmark, note that were the use of these new procedures forbidden, total welfare would be equal to 0.

Table 7 reports welfare estimates under each of these scenarios. Focusing first on the results maintaining the current physician learning parameters reported in the column titled "Status Quo", I find that welfare under the current coverage rules is positive, indicating that welfare would be reduced by forbidding the use of these procedures completely. Note that this is consistent with the procedures being better than the outside options on average. Similarly, never covering any of the procedures would meaningfully reduce welfare, albeit even in this scenario, physicians would, on average, continue to occasionally use high value procedures, meaning welfare would be higher than a "no Category III procedure use" baseline. Similarly, fully covering all procedures would dramatically raise welfare relative to the current regime. In fact, were Medicare to cover only procedures that are better than the incumbent procedure, the welfare gains would be only modestly larger than if coverage were universal, with universal coverage achieving 93% of the welfare gains from the infeasible "correct" coverage policy over the status quo.

The model also allows me to quantify the welfare costs of uncertainty and the value of

[13]Note that this is not the optimal coverage policy for a few reasons, even within the limited context of the model. First, the contractors do not take physician beliefs or uncertainty into account when deciding coverage rules. For example, it could be welfare enhancing to fail to cover a procedure that is only slightly better than the outside option if physicians' priors are extremely optimistic, avoiding a scenario of significant overuse. Second, this is a partial equilibrium analysis ignoring the fact that were physicians to recognize that the contractors were following this coverage rule, they would update their beliefs about the value of the procedure in light of the contractors' coverage decisions.

Table 7: Welfare Estimates

	Learning Environment		
Coverage Rules	No Learning	Status Quo	No Uncertainty
Actual Coverage	-19,198	23,172	1,886,914
Universal Non-Coverage	-24,588	1,524	1,757,800
Universal Coverage	-18,799	162,730	1,955,349
"Correct" Coverage	-11,447	173,055	1,971,605

Notes: Estimates of total welfare from simulations of the model with current and counterfactual coverage rules and learning parameters. "Correct" coverage means full coverage for procedures with $\delta_p^* \geq 0$ and non-coverage for all other procedures. The no learning environment does not allow providers to update their beliefs using signals of the procedure's quality, while the no uncertainty environment sets all physicians beliefs about the value of each new procedure to its true value with no uncertainty.

physician learning. Comparing coverage rules within rows of Table 7, we see that the welfare costs of imperfect information dwarf those of imperfect coverage rules. Moving from the status quo to the correct coverage rule achieves only 8.0% of the welfare gains on maintain the current coverage rules but eliminating physician uncertainty. Another way of interpreting this difference is that in terms of welfare, the status quo is much closer to a scenario of no learning by physicians than to an environment with no uncertainty. Note finally that without any learning, the new procedures are (in the context of the model) a net welfare loss even under infeasible coverage rules because of the inaccuracy of physician priors. Furthermore, coverage is less welfare-improving without learning, as allowing physicians to experiment and learn from experience has no value if they are unable to update their beliefs and practices.

7 Conclusion

In this paper, I consider the tradeoff between allowing experimentation with early access to innovations and requiring more evidence to support diffusion in the context of new medical procedures. After showing that the coverage decisions of local Medicare administrators greatly impact the diffusion of these procedures, I present ample evidence of social learning: health care providers learn about the value of new procedures from the experiences of other providers. This represents an important externality from the use of new procedures and is something the regulators must consider when determining whether to allow the spread of innovations. In order to quantify this externality and determine the optimal evidentiary standard and coverage policies in this context, I estimate a structural model of provider learning. The results of this model indicate that the current coverage regime results in higher welfare than having no utilization of the new procedures at all, but that more generous coverage rules would lead to large welfare gains. Furthermore, I find that the welfare costs of physician uncertainty are large and that learning and generous coverage rules are complementary.

Beyond offering prescriptions for how to improve Medicare coverage of new procedures, the evidence I present of social learning highlights its potential importance in other contexts. As policymakers weigh the tradeoffs of encouraging the promotion of innovations of uncertain value, the potential for early experimentation to dispel this uncertainty should lead policymakers to consider the potential welfare gains from allowing experimentation and encouraging social learning.

References

Agha, L. and D. Molitor (2018, 03). The Local Influence of Pioneer Investigators on Technology Adoption: Evidence from New Cancer Drugs. *The Review of Economics and Statistics 100*(1), 29–44.

Allen, T., K. Bilir, Z. Chen, and C. Tonetti (2019). Knowledge diffusion through networks. *Available at SSRN 3509835*.

American Medical Association (2010, February). AMA code of medical ethics ' opinions on patenting procedures and devices. *AMA Journal of Ethics 12*(2), 96–96.

Anand, B. N. and R. Shachar (2011). Advertising, the matchmaker. *The RAND Journal of Economics 42*(2), 205–245.

Avdic, D., S. v. H. K. Scholder, B. Lagerqvist, C. Propper, and J. Vikstrom (2018). Information shocks and provider adaptation: Evidence from interventional cardiology. Technical report, Technical report.

Baicker, K., A. Chandra, and J. S. Skinner (2012). Saving money or just saving lives? improving the productivity of us health care spending. *Annual Review of Economics 4*(1), 33–56.

Brandeis, L. (1932). New state ice co. v. liebmann.

Brot-Goldberg, Z., S. Burn, T. Layton, and B. Vabson (2022). Rationing medicine through bureaucracy: authorization restrictions in medicare. *Working Paper*.

Budish, E., B. N. Roin, and H. Williams (2016, May). Patents and research investments: Assessing the empirical evidence. *American Economic Review 106*(5), 183–87.

Buera, F. J., A. Monge-Naranjo, and G. E. Primiceri (2011). Learning the wealth of nations. *Econometrica 79*(1), 1–45.

Callander, S. and B. Harstad (2015). Experimentation in federal systems. *The Quarterly Journal of Economics 130*(2), 951–1002.

Callaway, B. and P. H. C. Sant'Anna (2021). Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics 225*(2), 200–230.

Carlson, M. D. A., J. Herrin, Q. Du, A. J. Epstein, E. Cherlin, R. S. Morrison, and E. H. Bradley (2009). Hospice characteristics and the disenrollment of patients with cancer. *Health Services Research 44*(6), 2004–2021.

Cengiz, D., A. Dube, A. Lindner, and B. Zipperer (2019). The Effect of Minimum Wages on Low-Wage Jobs. *Quarterly Journal of Economics 134*(3), 1405–1454. _eprint: https://academic.oup.com/qje/article-pdf/134/3/1405/29173920/qjz014.pdf.

Centers for Medicare and Medicaid Services (2003). Medicare program; revised process for making medicare national coverage determinations.

Centers for Medicare and Medicaid Services (2022, Apr). CMS finalizes medicare coverage policy for monoclonal antibodies directed against amyloid for the treatment of alzheimer's disease. *Newsroom*.

Chandra, A. and D. O. Staiger (2007). Productivity spillovers in health care: evidence from the treatment of heart attacks. *Journal of political Economy 115*(1), 103–140.

Chernew, M. E. and J. P. Newhouse (2011). Chapter one - health care spending growth. In M. V. Pauly, T. G. Mcguire, and P. P. Barros (Eds.), *Handbook of Health Economics*, Volume 2 of *Handbook of Health Economics*, pp. 1–43. Elsevier.

Clemens, J. and J. D. Gottlieb (2014, April). Do physicians' financial incentives affect medical treatment and patient health? *American Economic Review 104*(4), 1320–49.

CMS (2005, February). Report to congress medicare contracting reform: A blueprint for a better medicare.

Coleman, J., E. Katz, and H. Menzel (1957). The diffusion of an innovation among physicians. *Sociometry 20*(4), 253–270.

Conley, T. G. and C. R. Udry (2010, March). Learning about a new technology: Pineapple in ghana. *American Economic Review 100*(1), 35–69.

Coscelli, A. and M. Shum (2004). An empirical model of learning and patient spillovers in new drug entry. *Journal of econometrics 122*(2), 213–246.

Covert, T. R. (2015). Experiential and social learning in firms: the case of hydraulic fracturing in the bakken shale. Technical report.

Crawford, G. S. and M. Shum (2003). Uncertainty and learning in pharmaceutical demand: Anti-ulcer drugs. *Econometrica*.

Cutler, D. and R. Huckman (2003). Technological development and medical productivity: The diffusion of angioplasty in new york state. *Journal of Health Economics 22*(2), 187–217.

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

Depalo, D., J. Bhattacharya, V. Atella, and F. Belotti (2019). When technological advance meets physician learning in drug prescribing. Technical report, National Bureau of Economic Research.

Dranove, D., C. Garthwaite, C. Heard, and B. Wu (2021, October). The economics of medical procedure innovation. Working Paper 29438, National Bureau of Economic Research.

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. Series: Working Paper Series.

Eliason, P. J., R. J. League, J. Leder-Luis, R. C. McDevitt, and J. W. Roberts (2021). Ambulance taxis: The impact of regulation and litigation on health care fraud. Technical report, National Bureau of Economic Research.

Fafchamps, M., M. Soderbom, and M. vanden Boogaart (2016). Adoption with social learning and network externalities. Technical report, National Bureau of Economic Research.

Ferreyra, M. M. and G. Kosenok (2011). Learning about new products: An empirical study of physicians'behavior. *Economic Inquiry 49*(3), 876–898.

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

Fisher, E. S., D. E. Wennberg, T. A. Stukel, D. J. Gottlieb, F. Lucas, and E. L. Pinder (2003a). 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.

Fisher, E. S., D. E. Wennberg, T. A. Stukel, D. J. Gottlieb, F. Lucas, and E. L. Pinder (2003b, February). The implications of regional variations in medicare spending. part 2: health outcomes and satisfaction with care. *Annals of internal medicine 138*(4), 288—298.

Foote, S. B. and R. J. Town (2007). Implementing evidence-based medicine through medicare coverage decisions. *Health Affairs 26*(6), 1634–1642.

Foote, S. B., B. A. Virnig, R. J. Town, and L. Hartman (2008). The impact of medicare coverage policies on health care utilization. *Health Services Research 43*(4), 1285–1301.

Foster, A. D. and M. R. Rosenzweig (1995). Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of political Economy 103*(6), 1176–1209.

Gazzeri, R., M. Galarza, M. Neroni, C. Fiore, A. Faiola, F. Puzzilli, G. Callovini, and A. Alfieri (2015). Failure rates and complications of interspinous process decompression devices: a european multicenter study. *Neurosurgical focus 39*(4), E14.

Gilchrist, D. S. and E. G. Sands (2016). Something to talk about: Social spillovers in movie consumption. *Journal of Political Economy 124*(5), 1339–1382.

Gong, Q. (2017). Physician learning and treatment choices: Evidence from brain aneurysms. Technical report, working paper, University of Pennsylvania.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics 225*(2), 254–277. Publisher: Elsevier.

Gowrisankaran, G., V. Ho, and R. J. Town (2006). Causality, learning and forgetting in surgery. *Federal Trade Commission, draft*.

Grabowski, H. and Y. R. Wang (2008). Do faster food and drug administration drug reviews adversely affect patient safety? an analysis of the 1992 prescription drug user fee act. *The Journal of Law and Economics 51*(2), 377–406.

Grennan, M. and R. J. Town (2020, January). Regulating innovation with uncertain quality: Information, risk, and access in medical devices. *American Economic Review 110*(1), 120–61.

He, Q., K. Xiao, L. Peng, J. Lai, H. Li, D. Luo, and K. Wang (2019). An effective meta-analysis of magnetic stimulation therapy for urinary incontinence. *Scientific reports 9*(1), 1–10.

Hockenberry, J. M. and L. A. Helmchen (2014). The nature of surgeon human capital depreciation. *Journal of health economics 37*, 70–80.

Hwang, T. J., D. Carpenter, J. C. Lauffenburger, B. Wang, J. M. Franklin, and A. S. Kesselheim (2016, 12). Failure of Investigational Drugs in Late-Stage Clinical Development and Publication of Trial Results. *JAMA Internal Medicine 176*(12), 1826–1833.

Jewett, C. (2022, Mar). F.D.A. rushed a drug for preterm births. did it put speed over science?

Kellogg, R. (2011). Learning by drilling: Interfirm learning and relationship persistence in the texas oilpatch. *The Quarterly Journal of Economics 126*(4), 1961–2004.

League, R. J. (2023). Administrative burden and consolidation in health care: Evidence from medicare contractor transitions. Working paper.

Levinson, D. R. (2014, January). Local coverage determinations create inconsistency in medicare coverage. Technical report, Department of Health and Human Services.

Makary, M. (2021, Oct). The FDA can save thousands of lives today.

McKinlay, J. B. (1981). From "promising report" to "standard procedure": Seven stages in the career of a medical innovation. *The Milbank Memorial Fund Quarterly. Health and Society 59*(3), 374–411.

Mokyr, J. (2016). *A Culture of Growth*. Princeton University Press.

Moretti, E. (2011). Social learning and peer effects in consumption: Evidence from movie sales. *The Review of Economic Studies 78*(1), 356–393.

NBPAS (2021, Sep). What is traditional medicare vs. medicare advantage?: Prior authorization.

Olson, M. K. (2008). The risk we bear: the effects of review speed and industry user fees on new drug safety. *Journal of health economics 27*(2), 175–200.

Ryan, B. and N. C. Gross (1943). The diffusion of hybrid seed corn in two iowa communities. *Rural sociology 8*(1), 15.

Shi, M. (2022). The costs and benefits of monitoring providers: Evidence from medicare audits. *Available at SSRN 4063930*.

Skinner, J. and D. Staiger (2015). Technology diffusion and productivity growth in health care. *Review of Economics and Statistics 97*(5), 951–964.

Social Security Act (1965a). 42 U.S.C. § 1395y.

Social Security Act (1965b). 42 U.S.C. § 1395m.

Soumerai, S. B., T. J. McLaughlin, J. H. Gurwitz, E. Guadagnoli, P. J. Hauptman, C. Borbas, N. Morris, B. McLaughlin, X. Gao, D. J. Willison, et al. (1998). Effect of local medical opinion leaders on quality of care for acute myocardial infarction: a randomized controlled trial. *Jama 279*(17), 1358–1363.

Steinbuch, J., A. C. van Dijk, F. Schreuder, M. Truijman, J. Hendrikse, P. Nederkoorn, A. van der Lugt, E. Hermeling, A. Hoeks, and W. Mess (2017). Definition of common carotid wall thickness affects risk classification in relation to degree of internal carotid artery stenosis: the plaque at risk (parisk) study. *Cardiovascular ultrasound 15*(1), 1–8.

Stern, A. D. (2017). Innovation under regulatory uncertainty: Evidence from medical technology. *Journal of public economics 145*, 181–200.

Storrow, B. (2021, Oct). Congress eyes $235 billion in clean-energy subsidies.

Stoyanov, A. and N. Zubanov (2012, April). Productivity spillovers across firms through worker mobility. *American Economic Journal: Applied Economics 4*(2), 168–98.

Thornton, R. A. and P. Thompson (2001). Learning from experience and learning from others: An exploration of learning and spillovers in wartime shipbuilding. *The American Economic Review 91*(5), 1350–1368.

Wilk, A. S., R. A. Hirth, W. Zhang, J. R. C. Wheeler, M. N. Turenne, T. A. Nahra, K. K. Sleeman, and J. M. Messana (2018). Persistent variation in medicare payment authorization for home hemodialysis treatments. *Health Services Research 53*(2), 649–670.

Yang, K. K. (2022). Experience effects and technology adoption: Evidence from aortic valve replacement. Working Paper.

Zheng, K., R. Padman, D. Krackhardt, M. P. Johnson, and H. S. Diamond (2010). Social networks and physician adoption of electronic health records: insights from an empirical study. *Journal of the American Medical Informatics Association 17*(3), 328–336.

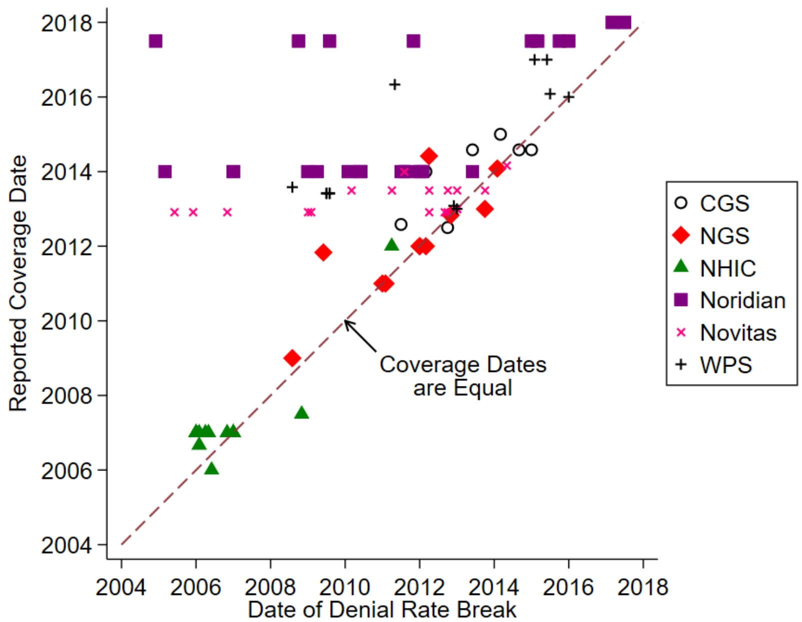
A Determining Coverage Rules

MACs are tasked with determining whether each individual claim for reimbursement meets the standard of medical necessity and accords with Medicare regulations. In practice, this claims processing process is highly systemized, with claims generally being automatically checked against general rules for whether the claim will be paid out (called "claim edits"). MACs often inform providers in their jurisdiction of changes to their coverage rules through publicly available announcements called local coverage decisions. These announcements are available at `https://www.cms.gov/medicare-coverage-database/`. Unfortunately, not all claim edits are publicly announced and available online. For example, while up to 18 administrative companies were active at one time during my sample period, the maximum number of MACs making their coverage rules for category III codes publicly available at one time is only eight. For this reason, I use the claim denial information in the claims data to infer coverage and validate this process using the coverage rules I do observe.

First, I detect structural breaks in level and trend of denial rates by MAC for each procedure. Performing a Chow test comparing the fitted model to the denial rate time series before and after each month, I flag potential structural breaks as the month corresponding to the smallest p-value rejecting the hypothesis that the time series is the same before and after the potential break, limiting to breaks corresponding to p-values below 0.1. Next, I manually screen the potential breaks to limit to those that represent readily apparent coverage changes (i.e. eliminating very gradual or temporary denial rate changes). Third, I limit the remaining potential coverage changes to those that represent a change of at least 15 percentage points in the denial rate. This process results in 106 detected coverage changes. Figure A1 shows that these coverage changes generally closely correspond to the announced coverage change dates for the MACs for which these announced rules are available. This is not the case for two MACs (Noridian and Novitas) for which it is clear that their reported coverage dates are inaccurate batch announcements of coverage changes that have already occurred.

Having detected changes in coverage, I then turn to classifying the coverage level as fully covered, covered on a case-by-case basis, or non-covered. These levels correspond to the language used in the available posted coverage rules and reflect the fact that even for claims reporting procedures that are not categorically denied, the denial rate tends to be much higher than for more established procedures. For example, in my sample, the across-procedure average denial rate is 77% with a standard deviation of over 30 percentage points. In contrast, League (2023) reports that in 2017 Medicare's denial rate for medical procedures was only 10 percent. I classify procedures as fully covered if the denial rate is less than 20%, non-covered if the denial rate is over 80%, and covered on a case-by-case basis otherwise. Finally, I restrict the earliest date of coverage (full or case-by-case) to be the later of the date the code became active (to eliminate

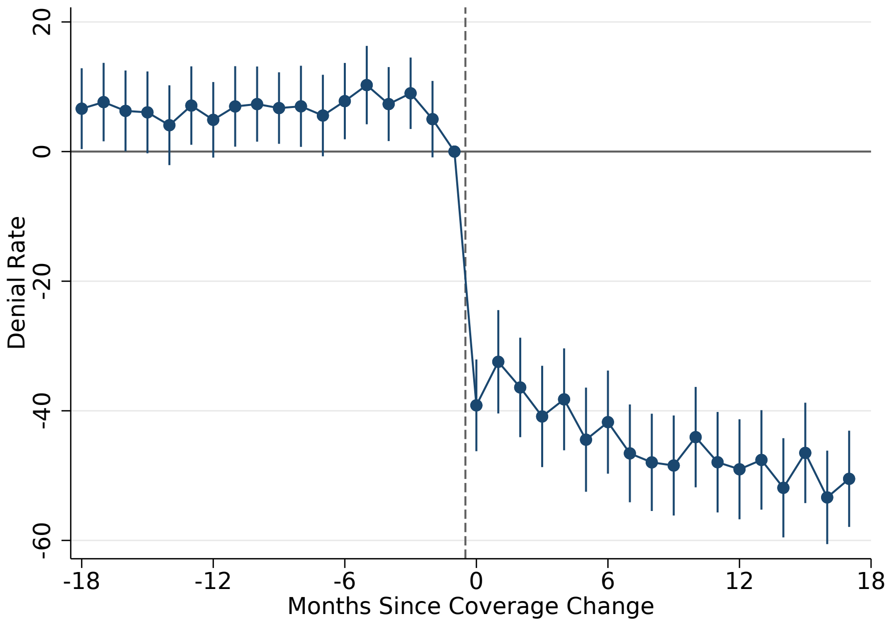
Figure A1: Correspondence of Detected Coverage Changes and Posted Announcements



Notes: The horizontal axis reports the detected coverage change date while the vertical axis reports the earliest coverage date reported in the posted coverage rules for all MAC-procedure pairs with a detected coverage change date.

improper payments) and the date of the first covered use (to account for the fact that non-covered procedures are unlikely to be used and all MACs for which I have posted coverage rules report a presumption of non-coverage for this class of procedures). As shown in Figure A2, the changes in coverage I detect correspond to large and immediate changes in the denial rate. Using this process, I find that 79% of MAC-procedure-month triples correspond to non-coverage, 15% correspond to coverage on a case-by-case basis, and 6% correspond to full coverage.

Figure A2: Change in Denial Rate at Coverage Change



Notes: The figure reports estimates for the change in share of claims denied in the 18 months before and after a change in coverage from non-coverage to full or case-by-case coverage or from case-by-case coverage to full coverage relative to the month before the coverage change. An observation is a MAC-procedure-month tuple. Sample is limited to a balanced panel of MAC-procedure pairs subject to a change in coverage. 95% confidence interval is shaded in blue. Standard errors are clustered at the MAC-procedure level.

B Using AMA Decisions as a Measure of Procedure Success

While the AMA decisions to promote or delete each code may also be used as a measure of a procedure's value, I choose to use whether the procedure's utilization grew or shrank over time for a few reasons. First, the AMA has yet to make a determination for many of the procedures introduced late in my sample period. For these procedures, I am not able to assess their success on this measure while I am for the measure I use in the main text. Second, my measure better captures how the beliefs of the medical community evolve. By measuring whether utilization grows or falls over time, my main measure does a better job of capturing procedures that providers learn are better or worse than previously believed. That said, my results are robust to using either measure, which isn't surprising given their general concordance, as shown in Table A1.

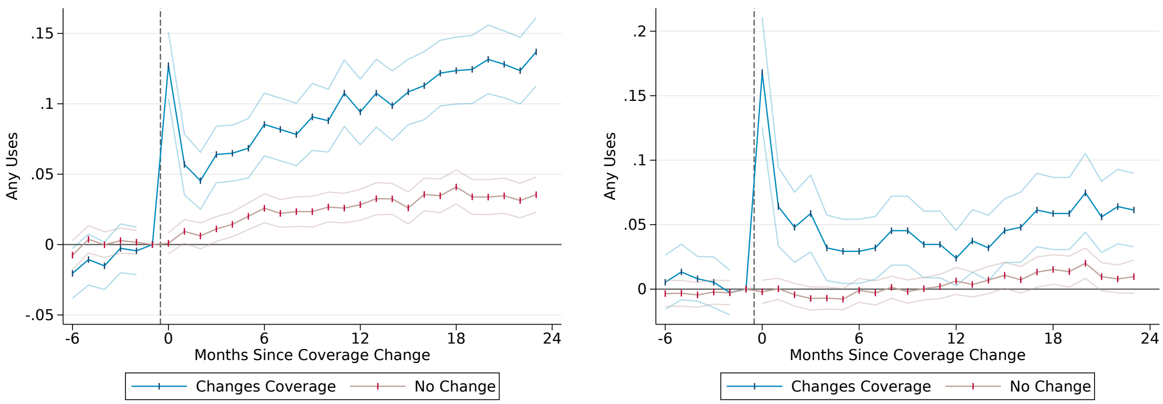
Next, I demonstrate the robustness of all the results in the main text to using this measure instead of my main one. This robustness is not surprising given the similarity of the measures. Figure A3 shows that the spillover from jurisdictions in which coverage changes to other jurisdictions similarly differ by whether the procedure was successful along this measure. Table A2 shows that idiosyncratic shocks to use are persistent only for procedures associated with promoted codes.

Table A1: Concordance of Success Measures

	AMA Code Status			
Adoption Status	Deleted	Outstanding	Promoted	Total
Adopted	42	56	110	208
	20.2%	26.9%	52.9%	60.6%
De-Adopted	53	45	37	135
	39.3%	33.3%	27.4%	39.4%
Total	95	101	147	343
	27.7%	29.4%	42.9%	100.0%

Figure A3: Change in Utilization at Coverage Change for Treatment and Control Jurisdictions

(a) Promoted Procedures (b) Deleted Procedures



Notes: The figures report estimates of τ_e (in red) and $\beta_e + \tau_e$ (in blue) from Equation (3) for $e \in \{-6, \ldots, 24\}$. An observation is a jurisdiction-procedure-month-group tuple, where groups are defined by the stacked regression procedure defined in Appendix C. Panel (a) presents estimates for the sample limited to procedures for which the associated code was eventually promoted to Category I status by the AMA, while panel (b) presents estimates for procedures for which the associated code was deleted. 95% confidence intervals are given by the translucent tickless lines of the relevant color. Standard errors are clustered at the group level.

Table A2: Persistence of Idiosyncratic Shocks to Use

	(1) Use Rate	(2) Use Rate	(3) Use Rate
Lagged Use	1.445***	1.226***	0.849***
	(0.231)	(0.229)	(0.176)
Lagged Use × Deleted	-0.795***	-0.589*	-0.491
	(0.218)	(0.248)	(0.421)
Months of Lag	6	12	24
Dep. Var. Mean	6.127	6.473	6.176
Observations	19,710	17,658	13,842

Notes: Estimates of β_1 and β_2 from Equation (4). An observation is a procedure-month pair. Standard errors are clustered at the procedure level. $^{+}$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

C Details on the Estimation of the Effect of Coverage on Utilization

Here I provide more details on stacked regression and show results from various windows. Both of these methods can be used to overcome some of the problems with the convention two-way fixed effects estimator raised by Callaway and Sant'Anna (2021) and Goodman-Bacon (2021), among others. Stacked regression method, from Cengiz et al. (2019), works by constructing appropriate controls groups for each transition between different coverage level for each procedure for each jurisdiction. To implement this method, I create separate datasets for each change of coverage w (for wave) consisting of the jurisdiction-procedure pair that changes coverage at time g and control jurisdiction-procedure pairs of the same procedure for which coverage does not change during the event window. Each of these datasets is appended (or "stacked") such that each jurisdiction-procedure pair for which coverage changes appears once while each jurisdiction-procedure pair may appear as a control multiple times (although with different time values). To obtain estimates of the dynamic treatment effect of a coverage change, I can then estimate

$$Y_{pjtw} = \sum_{e=-K}^{-2} \beta_e T_{pjtw}(e) + \sum_{e=0}^{L} \beta_e T_{pjtw}(e) + \alpha_{pjw} + \alpha_{ptw} + \varepsilon_{pjtw}, \tag{9}$$

where K gives the size of the treatment window, $T_{jtw}(e)$ is an indicator for being the jurisdiction-procedure pair that changes coverage e months from transition (where e denotes event time: $e \equiv t - w$), α_{pjw} and α_{ptw} are procedure-by-jurisdiction-by-wave and procedure-by-time-by-wave fixed effects. These fixed effects account for the fact that control observations may appear more than once in this stacked dataset. Similarly, I can aggregate the pre- and post-coverage change periods and in this stacked dataset to estimate the average treatment effect of a coverage change over the L months following the change:

$$Y_{pjtw} = \beta \sum_{e=0}^{L} T_{pjtw}(e) + \alpha_{pjw} + \alpha_{ptw} + \varepsilon_{pjtw}, \tag{10}$$

Analogously, I can aggregate the treatment effect estimate using the traditional two-way fixed effects estimator to obtain an estimate of the treatment effect of a coverage change. To this, I use the non-stacked data set to estimate

$$Y_{pjt} = \beta \sum_{e=0}^{L} T_{pjt}(e) + \gamma_{pj} + \gamma_{pt} + \varepsilon_{pjt}. \tag{11}$$

Table A3: Effect of Coverage on Utilization

	(1) Use Rate	(2) Use Rate	(3) Use Rate	(4) Any Uses	(5) Any Uses	(6) Any Uses
Coverage Extended	7.499***	4.781**	3.505*	0.0728***	0.0512***	0.481***
	(1.355)	(1.637)	(1.434)	(0.00451)	(0.00457)	(0.00471)
Jurisdiction FEs	1	1	1	1	1	1
Time FEs	0	1	0	1	0	1
Stacked?	0	0	1	0	0	1
Dep. Var. Mean	9.795	2.708	10.54	0.135	0.0473	0.122
Observations	58,454	1,004,226	2,127	58,454	1,004,226	2,217

Notes: An observation is a procedure-jurisdiction-month. Sample consists of monthly observations from 2002-2017. Use rate is the number of uses of the procedure per million beneficiaries. Regressions include jurisdiction-by-procedure and procedure-by-month fixed effects where indicated. Standard errors are clustered at the jurisdiction-by-procedure level. $^+$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

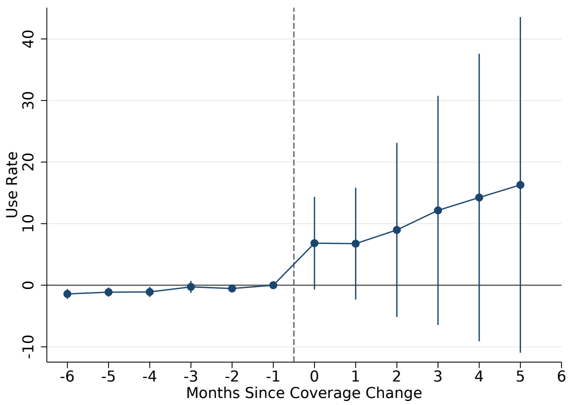
Notice that for both Equations (10) and (11), the β gives the differential change for the L months after coverage changes in the treated jurisdictions from to the K months before, relative to the appropriate control jurisdictions.

In Table A3, I present estimates of the treatment effect of extending coverage to a new procedure using these estimators for the treatment window from $K = -6$ to $L = 24$. Because to estimate Equation (10) I restrict each treatment-control wave w to be a balanced panel and to estimate Equation (11) I keep only treated jurisdiction-code pairs in the data for the entire event window, the number of observations differs by the model estimated and the length of the event window.

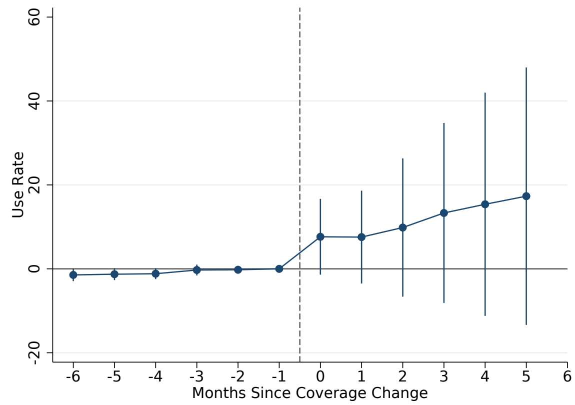
Figures A4, A5, A6, and A7 present estimates of Equation (2) with and without time-by-procedure fixed effects and Equation (9) using treatment windows of $L = K = 6$, $L = K = 12$, $L = K = 18$, and $L = K = 24$, respectively.

Figure A4: Change in Utilization at Coverage Change

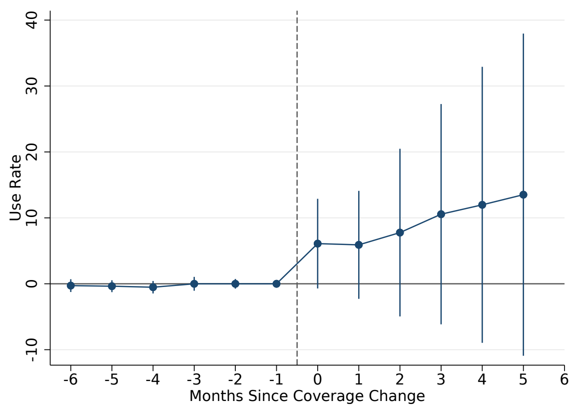
(a) No Time Fixed Effects



(b) Two-Way Fixed Effects



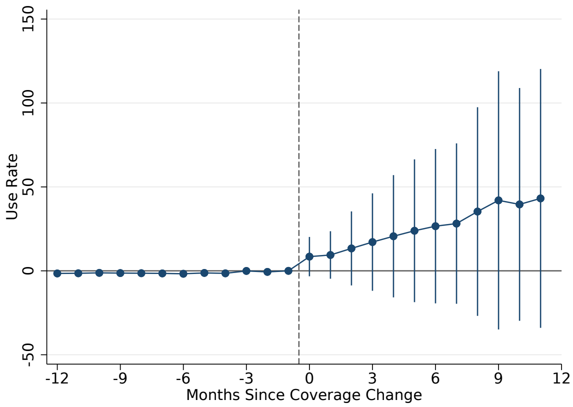
(c) Stacked Regression



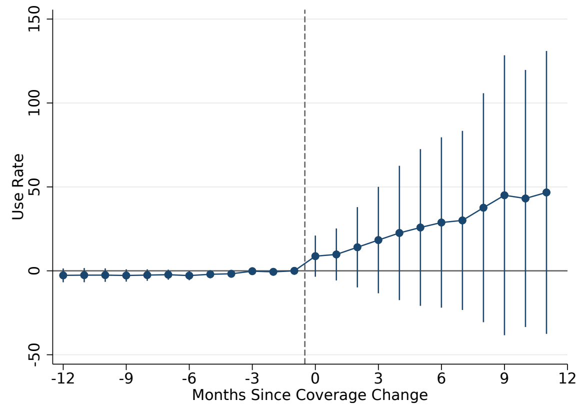
Notes: The figures report estimates of β_e from Equation (2) and (9) for $e \in \{-6, \ldots, 6\}$. An observation is a jurisdiction-procedure-month tuple. Panel (a) includes estimates without time fixed effects. Panel (b) uses the two-way fixed effects estimator. Panel (c) uses the stacked regression estimator. For all three specifications, only treated jurisdiction-procedure pairs in the data for the entire treatment window are included along with untreated observations. Error bars give the pointwise 95% confidence intervals. Standard errors are clustered at the jurisdiction-procedure level.

Figure A5: Change in Utilization at Coverage Change

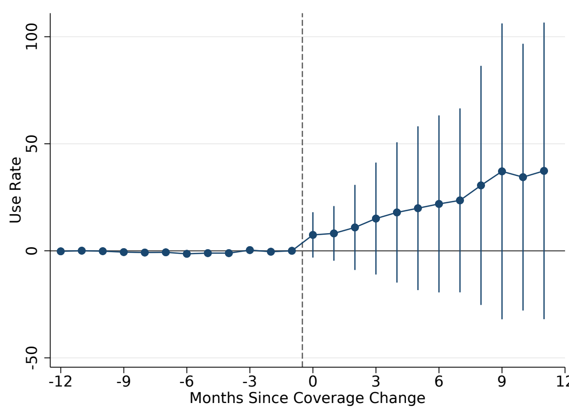
(a) No Time Fixed Effects



(b) Two-Way Fixed Effects



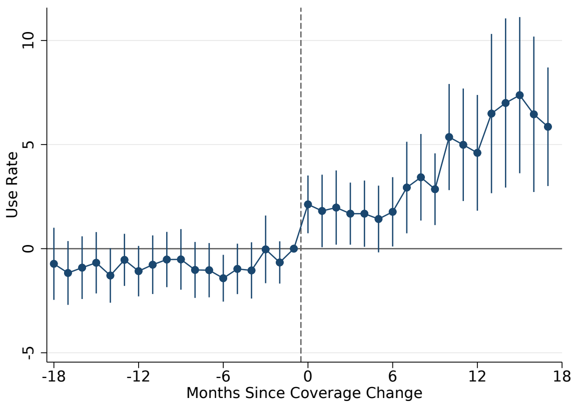
(c) Stacked Regression



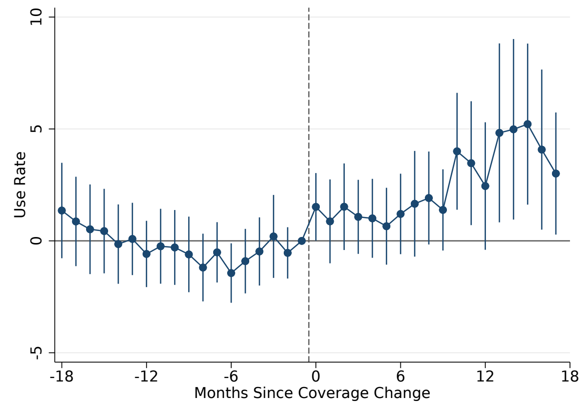
Notes: The figures report estimates of β_e from Equation (2) and (9) for $e \in \{-12, \ldots, 12\}$. An observation is a jurisdiction-procedure-month tuple. Panel (a) includes estimates without time fixed effects. Panel (b) uses the two-way fixed effects estimator. Panel (c) uses the stacked regression estimator. For all three specifications, only treated jurisdiction-procedure pairs in the data for the entire treatment window are included along with untreated observations. Error bars give the pointwise 95% confidence intervals. Standard errors are clustered at the jurisdiction-procedure level.

Figure A6: Change in Utilization at Coverage Change

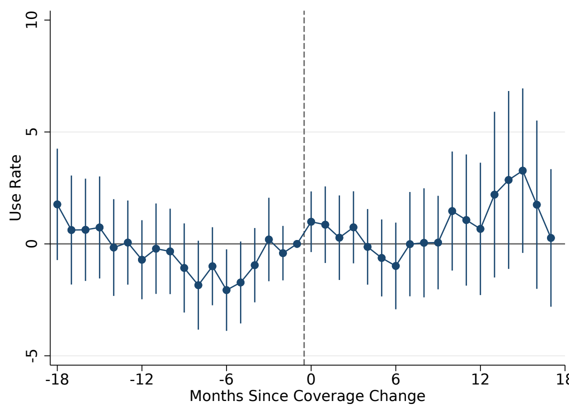
(a) No Time Fixed Effects



(b) Two-Way Fixed Effects



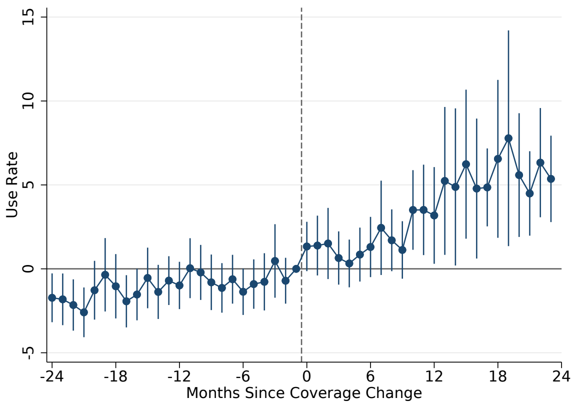
(c) Stacked Regression



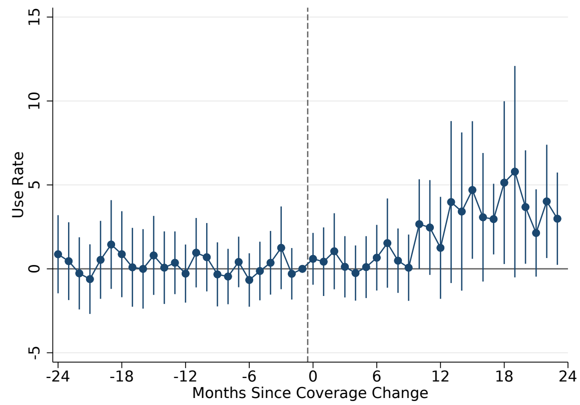
Notes: The figures report estimates of β_e from Equation (2) and (9) for $e \in \{-18, \ldots, 18\}$. An observation is a jurisdiction-procedure-month tuple. Panel (a) includes estimates without time fixed effects. Panel (b) uses the two-way fixed effects estimator. Panel (c) uses the stacked regression estimator. For all three specifications, only treated jurisdiction-procedure pairs in the data for the entire treatment window are included along with untreated observations. Error bars give the pointwise 95% confidence intervals. Standard errors are clustered at the jurisdiction-procedure level.

Figure A7: Change in Utilization at Coverage Change

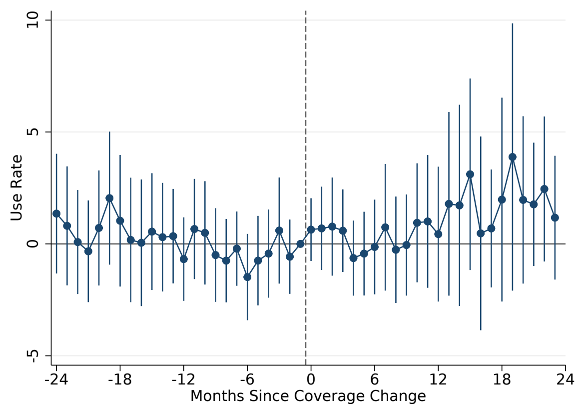
(a) No Time Fixed Effects



(b) Two-Way Fixed Effects



(c) Stacked Regression



Notes: The figures report estimates of β_e from Equation (2) and (9) for $e \in \{-24, \ldots, 24\}$. An observation is a jurisdiction-procedure-month tuple. Panel (a) includes estimates without time fixed effects. Panel (b) uses the two-way fixed effects estimator. Panel (c) uses the stacked regression estimator. For all three specifications, only treated jurisdiction-procedure pairs in the data for the entire treatment window are included along with untreated observations. Error bars give the pointwise 95% confidence intervals. Standard errors are clustered at the jurisdiction-procedure level.

D Using Administrator Decisions to Infer Evidence

In this appendix, I justify my use of the share of administrators covering a procedure as indicating the strength of the evidence surrounding the procedure's efficacy. I first do this by proving that the model of MAC behavior in Section 6 implies that the share of administrators coverage a procedure is increasing the strength of the evidence:

Proposition 1 *The expected share of MACs covering a procedure at any given time is weakly increasing in the procedure-specific quality of evidence, i.e.* $e_{t_2p_2} - \mu_{p_2} > e_{t_1p_1} - \mu_{p_1}$ *implies* $\int \frac{1}{I}\sum_i \mathbb{I}[e_{t_2p_2} > s_{ip_2}]d\varepsilon \geq \int \frac{1}{I}\sum_i \mathbb{I}[e_{t_1p_1} > s_{ip_1}]d\varepsilon$.

Proof. Suppose the proposition is false. Then there exists a pair of procedure-months such that $e_{t_2p_2} - \mu_{p_2} > e_{t_1p_1} - \mu_{p_1}$ while

$$\int \frac{1}{I}\sum_i \mathbb{I}[e_{t_2p_2} - \mu_{p_2} > \theta_i + \varepsilon_{ip_2}]d\varepsilon < \int \frac{1}{I}\sum_i \mathbb{I}[e_{t_1p_1} - \mu_{p_1} > \theta_i + \varepsilon_{ip_1}]d\varepsilon,$$

or by the linearity of expectation,

$$\int \frac{1}{I}\sum_i \left(\mathbb{I}[e_{t_2p_2} - \mu_{p_2} > \theta_i + \varepsilon_{ip_2}] - \mathbb{I}[e_{t_1p_1} - \mu_{p_1} > \theta_i + \varepsilon_{ip_1}]\right)d\varepsilon < 0.$$

Because $e_{t_2p_2} - \mu_{p_2} > e_{t_1p_1} - \mu_{p_1}$, this implies that

$$\int \frac{1}{I}\sum_i \left(\mathbb{I}[e_{t_2p_2} - \mu_{p_2} > \theta_i + \varepsilon_{ip_2}] - \mathbb{I}[e_{t_2p_2} - \mu_{p_2} > \theta_i + \varepsilon_{ip_1}]\right)d\varepsilon < 0,$$

or

$$\int \mathbb{I}[\varepsilon_{ip_1} > \varepsilon_{ip_2}]d\varepsilon > 0.5,$$

or $0.5 > 0.5$, a contradiction. ■

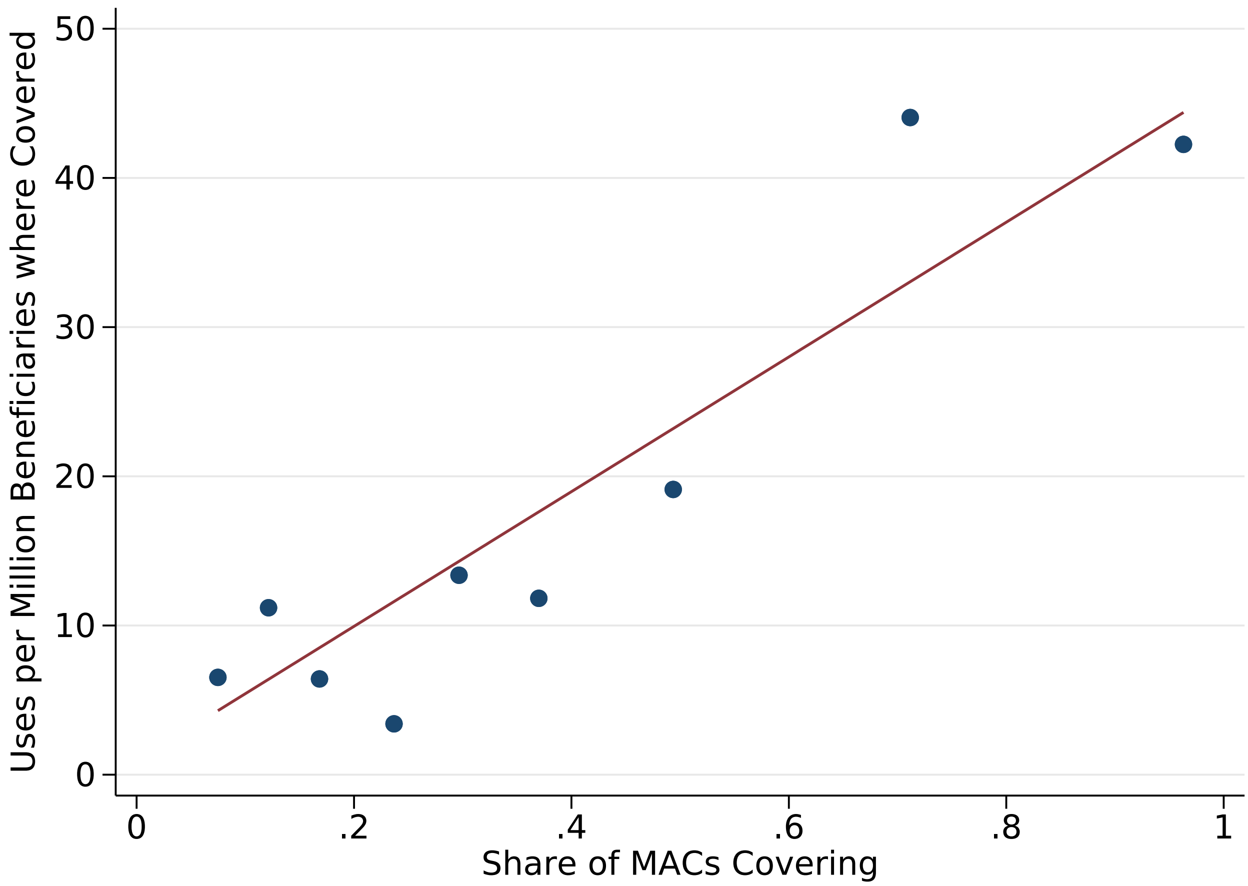
Next, I provide empirical evidence that the beliefs about procedures' values on the part of providers correspond to those of administrators. I do this by showing that in jurisdictions in which a procedure is covered, it's utilization is increasing in the share of administrators that have determined it meets Medicare's coverage standards. By limiting the sample to jurisdictions in which the procedure is covered, I am eliminating the direct effect of coverage on utilization and isolating the differences in selection of procedures that are widely covered compared to those that are not. Figure A8 plots a binned scatterplot of the share of administrators covering a procedure against its utilization in jurisdictions in which it is covered. We see that procedures that more administrators believe meet Medicare's coverage standards are also utilized more by providers.

Estimating the equation

$$Y_{pt} = \beta_0 + \beta_1 ShareCov_{pt} + \varepsilon_{pt} \tag{12}$$

on this same sample, I estimate β_1 to be 269.0, meaning that on average a 10 percentage point increase in the share of MACs covering a procedure is associated with an increase in utilization of 26.9 uses per million beneficiaries.[14]

Figure A8: Relationship Between Utilization and the Share of Administrators Covering a Procedure



Notes: The figure reports the average monthly utilization per million beneficiaries for each of nine quantiles of the share of administrators that cover a procedure fully or on a case-by-case basis. An observation is a procedure-month pair. Sample is limited to jurisdictions in which the procedure is covered fully or on a case-by-case basis. The red line gives the predicted values from estimates of Equation 12.

[14]The standard error of this estimate clustered at the procedure level is 59.92, meaning that the estimate of β_1 is statistically different from zero at the 0.1% significance level.

E Derivation of Likelihood Function

In this appendix, I derive the likelihood function I use to estimate the model presented in Section 6.

First, I derive the distribution of agents' beliefs after receiving a given number of independent signals from a common distribution. I assume that the initial belief of physician i about procedure p, δ_{0ip}, is drawn from the distribution $\mathcal{N}(\delta_{0p}, \sigma_{\delta p})$ and that all n subsequent signals about procedure p received by provider i, ν_{ipk}, are drawn from the distribution $\mathcal{N}(\delta_p^*, \sigma_{\nu p})$ for $k \in \{1, \ldots, n\}$ independently of the signal received by other physicians. Using Bayes rule, this means that physician beliefs after receiving n signals $\{\nu_{ipk}\}_{k\in\{1,\ldots,n\}}$, μ_{ipn}, are distributed $\mathcal{N}(\mu_{pn}^m, \sigma_{pn}^m)$, where

$$\sigma_{pn}^m = \frac{\sigma_{\delta p}\sigma_{\nu p}}{\sqrt{\sigma_{\nu p}^2 + n\sigma_{\delta p}^2}} \qquad \text{and} \qquad \mu_{pn}^m = \left(\sigma_{pn}^m\right)^2 \left(\frac{\delta_{0p}}{\sigma_{\delta p}^2} + \frac{n\delta_p^*}{\sigma_{\nu p}^2}\right)$$

Given this distribution of beliefs, the share of physicians having adopted procedure p at time t (after n independent signals to each physician i), $\alpha_{cov,pt}$, given the coverage status in their jurisdiction is given by $\mathbb{P}(\mu_{ipn} - X_{ip} - \beta_{1p}Case_{ipt} - \beta_{2p}Noncov_{ipt} > 0 | Case_{ipt}, Noncov_{ipt})$. With the assumption that X_{ip} is distributed standard normally, the distribution of $\mu_{ipn} - X_{ip} - \beta_{1p}Case_{ipt} - \beta_{2p}Noncov_{ipt}$ is given by $\mathcal{N}(\mu_{diff,n}^m, \sigma_{diff,n}^m)$, where

$$\mu_{diff,pn}^m = \mu_{pn}^m - \beta_{1p}Case_{ipt} - \beta_{2p}Noncov_{ipt} \qquad \text{and} \qquad \sigma_{diff,pn}^m = \sqrt{\left(\sigma_{pn}^m\right)^2 + 1}.$$

Thus we have that

$$\alpha_{cov,pt} = 1 - \Phi\left(-\frac{\mu_{diff,pn}^m}{\sigma_{diff,pn}^m}\right) = 1 - \Phi\left(-\frac{\mu_{pn}^m - \beta_{1p}Case_{ipt} - \beta_{2p}Noncov_{ipt}}{\sqrt{\left(\sigma_{pn}^m\right)^2 + 1}}\right).$$

Notice that μ_{pn}^m does not depend on the values taken by any of the signals received by any of the agents and is instead only a function of model parameters.[15] This means that I do not need to address the differences between the conditional and unconditional (on the signals) distributions of the share of adopting providers and that $\alpha_{cov,pt}$ is (non-stochastically) defined by the model parameters.

The final complication is that I do not directly observe physicians' adoption decisions. Rather, I only observe physicians' treatment decisions which means number of patients treated using procedure p by physician i at time t is distributed $B(g_{ipt}, \alpha_{cond,pt}\eta_p)$, where g_{ipt} is the number of patients in the consideration set of physician i at time t. This is true because I assume the η_p

[15] This would not be the case were all agents to receive common or correlated signals.

is independent of the physician's adoption decision (i.e. patients for whom the new treatment is appropriate are not disproportionately likely to be treated by a physician who has adopted the new procedure). Furthermore, because I assume the number of patients in the consideration set is also independent of the physician's adoption decision (i.e. physicians who treat many patients are not differentially likely to adopt the procedure), the total number of patients treated by physicians with the same observable characteristics of i (namely the same coverage status) at time t (denoted $y_{cov,pt}$) is distributed $B(g_{cov,pt}, \alpha_{cond,pt}\eta_p)$, where $g_{cov,pt}$ denotes the total number of patients in the consideration set of all such physicians.

Thus, the conditional likelihood function of $\alpha_{cov,pt}$ and η_p given the observed data is

$$\mathcal{L}_{cov,pt}(\alpha_{cov,pt}, \eta_p | g_{cov,pt}, y_{cov,pt}) = \binom{g_{cov,pt}}{y_{cov,pt}} (\alpha_{cov,pt}\eta_p)^{y_{cov,pt}} (1 - \alpha_{cov,pt}\eta_p)^{g_{cov,pt} - y_{cov,pt}},$$

and because $\alpha_{cov,pt}$ is non-stochastic, the likelihood function of model parameters θ_p is the same:

$$\mathcal{L}_{cov,pt}(\theta_p | g_{cov,pt}, y_{cov,pt}) = \mathcal{L}_{cov,pt}(\alpha_{cov,pt}, \eta_p | g_{cov,pt}, y_{cov,pt}).$$

Combining the likelihood function across coverage regimes and time, we have that the log-likelihood function for procedure p is

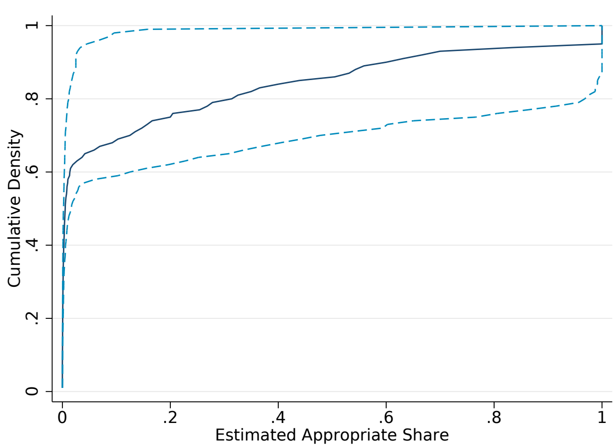
$$\mathcal{L}(\theta_p | X_{cov,pt}) = \sum_{t \in \{1,\dots,T\}} \sum_{cov \in \{-1,0,1\}} \log\left(\binom{g_{cov,pt}}{y_{cov,pt}}\right) + y_{cov,pt} \log(\alpha_{cov,pt}\eta_p) + (g_{cov,pt} - y_{cov,pt}) \log(1 - \alpha_{cov,pt}\eta_p).$$

F Alternative Estimation of η_p

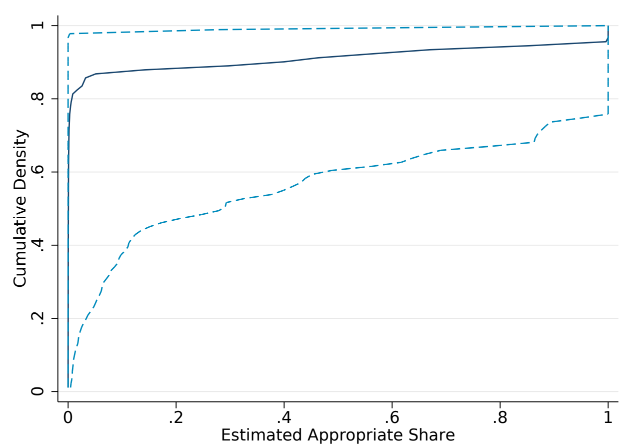
To check my estimates of η_p, I separately estimate the share of patients in the consideration set treated by physicians who have used the new procedure at least once in the past and will use it at least once in the future. Because the consideration set of patients is often small and the share of patients for whom the new procedure is appropriate is often smaller still, I cannot infer that a physician's lack of utilization of the new procedure implies that the physician has not adopted it as the physician may simply see no patients in that month for whom the treatment is appropriate. For this reason, I limit my estimation to the months strictly between a physician's first and last utilization. Under the assumption that providers do not de-adopt and re-adopt the procedure in this time, this provides an unbiased estimate of η_p.

Panel (a) of Figure A9 reports the distribution of these estimates of η_p while panel (b) reports the distributions estimated jointly with the other model parameters. Both estimation procedures indicate that the majority of new procedures are only appropriate for a very small fraction of the total patients with diagnoses indicating they could potentially benefit from the treatment. The two estimation procedures generally quite similar, although I find somewhat fewer procedures with very low values of η_p when estimating them separately from the other model parameters. That said, the jointly estimated distribution of η_p appears consistent with the 95% confidence intervals reported for the separately estimated distribution.

Figure A9: Distribution of Estimates for η_p



(a) Separately Estimated



(b) Jointly Estimated

Notes: The figure reports the empirical cumulative density function of estimates of η_p. An observation is a procedure. Dashed lines give the empirical CDF of the upper and lower bounds of the procedure-wise 95% confidence intervals. Standard errors for each procedure-level estimate are clustered at the provider level. Panel (a) reports separate estimates of η_p using the procedure described in Appendix F. Panel (b) reports estimates of η_p estimated jointly with the other model parameters using the procedure described in Section 6.1.