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

Riley League†

September 2025

The value of innovative technologies is often initially uncertain, forcing policymakers to weigh the potential benefit of promoting a valuable innovation against the cost of encouraging adoption of an ineffective one. This tradeoff is further complicated when wider adoption can reduce uncertainty by revealing information about the innovation's effectiveness. 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 contractors, I show that these rules significantly influence 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 increasing coverage for new procedures would result in large welfare gains.

*This research was supported by the National Institute on Aging, grant number T32-AG000186. I would like to thank Mingpei Li and Mohammad Mahmoudi Meymand for research assistance.

†University of Illinois Urbana-Champaign and NBER, league@illinois.edu

1 Introduction

Regulators often set the evidentiary threshold for allowing innovations of uncertain value to proliferate. 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 initially 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. Successful innovations can deliver valuable advances such as improved health, cost savings, and increased productivity, while unsuccessful ones can lead to uncertain health benefits, increased costs, and wasted resources (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 often issue conflicting coverage rules. As a result, Medicare reimbursement for new procedures varies widely by location and timing, often leading to a staggered rollout of coverage for new procedures. This variation gives incentives for providers in some jurisdictions to adopt the innovation early on, even when evidence on its efficacy is scarce, while others must wait and learn from the experience of providers elsewhere.

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 fivefold increase in utilization. This novel evidence of the ability of local administrators to impact the adoption of new innovations indicates that the evidentiary thresholds used by regulators has a large impact on welfare. Furthermore, since administrators influence utilization, knowledge spillovers across providers constitute an information externality: 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 more favorable coverage rules and updating their beliefs accordingly. Second, I show that the spillovers in utilization are larger when the procedure is covered in larger jurisdictions, consistent with additional utilization revealing additional information about the true value of the procedures. Third, I use idiosyncratic variation in the number of beneficiaries for which the procedure is covered when it is first introduced to show that additional past experience with a procedure causes more rapid adoption of successful procedures and speeds de-adoption of unsuccessful ones. Finally, I argue that other potential drivers of diffusion, including learning from clinical trials, learning-by-doing, 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 evidence coming from 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. 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 98% of the welfare gains of the static welfare-maximizing coverage policy. Simlutating welfare under the range of coverage policies observed in the data, I find that welfare is monotonically increasing in coverage generosity, indicating that allowing for additional experimentation and learning would be extremely valuable to patients.

The tradeoff between allowing for early experimentation and learning or waiting until the evidence is more certain is one commonly faced by regulators. There is substantial policy debate over the appropriate regulatory standard in contexts as diverse as pharmaceutical regulation (Jewett, 2022; Makary, 2021), AI safety (De Almeida et al., 2021), and regulation of cryptocurrencies (Davis and Kim, 2025). 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 (e.g., Grabowski and Wang, 2008; Olson, 2008; 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 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 journal 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). By

[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. By 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).

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 universal spillovers despite their likely importance.

More generally, research on global knowledge spillovers is scarce relative to their likely importance. Due to the endogeneity of innovation adoption decisions, to credibly identify knowledge 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., 2022; Moretti, 2011; Covert, 2015).[2] By contrast, I exploit exogenous variation in agents' abilities to adopt the innovations in order to identify learning spillovers.

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. While recent research has highlighted the potential for administratively determined prices (Clemens and Gottlieb, 2014), denials rates (Dunn et al., 2023; League, 2023), prior authorization policies (Brot-Goldberg et al., 2023; Eliason et al., 2025), and audits (Shi, 2024) to influence medical practice, few 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

[2] 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)).

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 the effect of counterfactual coverage policies. Finally, in Section 7, I conclude.

2 Institutional Context

2.1 Medicare Administrative Contractors and Coverage Rules

While often perceived as a single, centrally operated federal insurance program, Traditional Medicare relies heavily on private entities for its daily operations. Although the federal government sets federal Medicare policies and bears all actuarial risk, it delegates administrative responsibilities to private contractors known as Medicare Administrative Contractors (MACs).[3] These contractors handle essential functions such as processing medical claims and prior authorization requests, enrolling providers in the Medicare program, and determining coverage criteria for provider reimbursements 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 administrators 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.[4]

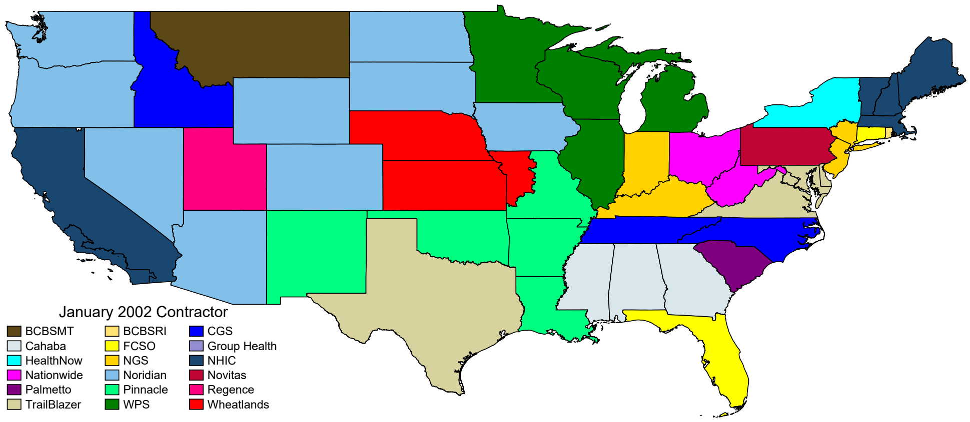
While there are statutory guidelines as to the type of medical services Medicare is intended to pay for, these administrators have wide discretion over how to implement these broad standards. The coverage standard the administrative contractors must implement is to avoid payment for services that "are 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 con-

[3]Before the Medicare Modernization Act of 2003, these contractors were referred to as "carriers" or "fiscal intermediaries." Despite the name change, their authority over local Medicare coverage decisions for new procedures remained intact. For consistency, I use the term "Medicare Administrative Contractors" across all periods. See League and Shi (2025) for more discussion about the transition from carriers and fiscal intermediaries to MACs.

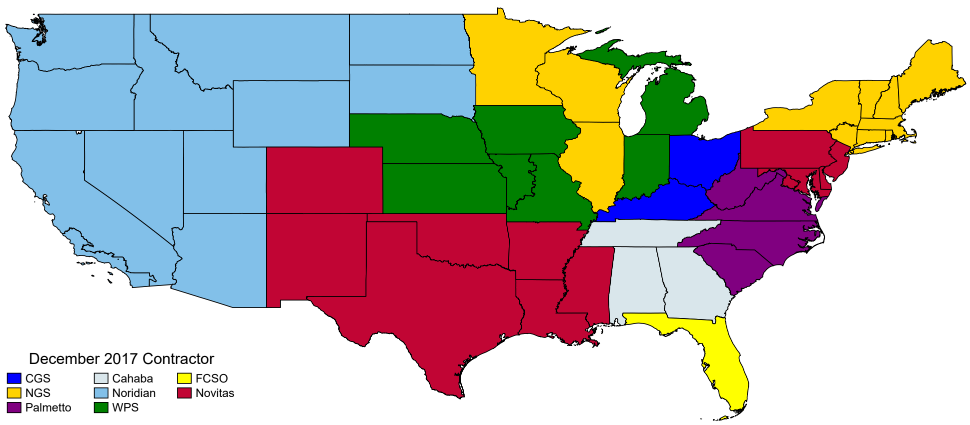
[4]A more detailed description of the jurisdiction combination process and its effects on the health care system is available in League (2023).

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.

tractors.[5] 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 pro-

[5]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).

cedures.[6] 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 with these restrictions rarely being common across all regions (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. This variation provides an opportunity to study how coverage rules influence procedure adoption and physician learning.

2.2 New Medical Procedures

New medical procedures go through a much lighter regulatory process than other medical innovations, such as pharmaceuticals or devices. After the procedure is created,[7] 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 the adoption of new procedures. Since the introduction of category III codes in 2002, there have been over 700 codes created for new procedures.

As these procedures represent the universe of new procedures added to the CPT codebook, they vary widely along multiple dimensions: the type of maladies they address, whether they are diagnostic or therapeudic, and whether they require a medical device, along with many other characteristics. Three procedures highlighted by Dranove, Garthwaite, Heard, and Wu (2022), henceforth DGHW (2022), are transcatheter aortic valve replacement (a cardiac surgical intervention requiring a newly created medical that was wildly successful), corneal incisions using laser (a novel surgical procedure that relied on well-established established technology), and applied behavioral therapy (a therapeudic technique meant to improve social and communication skills for patients with mental disorders). In my data, the most used procedures are corneal

[6]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."

[7]I do not consider the procedure creation process. This is because innovators are largely unable to capture the value of the procedures they create, such that Medicare coverage policies likely have little effect on the direction or amount of innovation. For example, Dranove, Garthwaite, and Wu (2022) note that for the set of new procedures the authors consider, all innovators are physicians practicing at hospitals and only a third of them hold device patents. Similarly, Dranove, Garthwaite, Heard, and Wu (2022) argue that the low private returns to procedure innovation have limited the amount of innovation that has occurred.

pachymetry (a test for measuring the thickness of the cornea), computer-aided evaluation of MRI breast imaging (a diagnostic tool to assist radiologists in diagnosing breast cancer), and intensity modulated radiation therapy (a radiation treatment meant to kill cancer cells). This small collection of examples of new procedures highlights the diversity of procedures covered by category III codes.

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.[8] Procedures reported using category I codes are the vast majority of codes and are much more likely to be paid by insurers, such that promotion to a category I code can be thought of as marking the innovation's success (DGHW, 2022).

This regulatory process has received little attention from academics. One notable exception is the work of DGHW (2022), 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 players studied in this paper: Medicare Administrative Contractors.

The 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, and many new procedures are deleted from the CPT system entirely by the American Medical Association. Indeed, DGHW (2022) 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 since been found to be of very limited utility.

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

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 Traditional Medicare.

In Table 1, I present procedure-level summary statistics on the utilization of new medical procedures among the population covered by my data. We see that there is wide variation in the overall level of utilization, as well as in the average reimbursement for the procedure and the share of claims denied.

Table 1: Summary Statistics

	Mean	Std. Dev.
Utilization Rate	6.520	33.95
Unique Patients Treated	1,447.24	6,833.74
Unique Providers Using	245.31	883.20
Procedure Price ($)	446.05	655.62
Percentage of Claims Denied (%)	76.79	30.02
Months Active in Data	63.61	39.90
Adopted?	0.608	0.489
AMA Promotion Status		
Promoted	0.427	0.495
Outstanding	0.295	0.457
Deleted	0.278	0.449
Distinct Procedures in Data	342	

Notes: Sample consists of all professional claims reporting a category III CPT code for a 20% sample of Traditional Medicare beneficiaries from 2002–2017. Utilization rate is the number of uses of the procedure per million beneficiary-months. Procedure price is the average paid amount for positively paid claims. An observation is a procedure.

In addition to utilization information, the data include 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. Coverage is classified as non-coverage (indicating claims for the procedure will be presumptively denied by the administrator), coverage on a case-by-case basis (indicating a heightened level of scrutiny relative to claims for most established, category I procedures), or full coverage (indicating presumptive claim payment in line with claims for established procedures). Appendix A gives more details on this classification process.

I find that administrators disagree quite regularly about whether these procedures meet this standard. 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 procedures almost a quarter of the time while the least generous never covers any procedure.[9] This variation results in 42% 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 1% of the time. Overall administrators are very hesitant to cover these procedures with procedures being non-covered over 85% of the time, while even the most generous administrator only fully covers less than 12% of procedure-months.

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.

Consistent with there being a high degree of uncertainty about the value of each new procedure, I find that nearly half of new 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.[10] Of the 342 procedures in my data, 208 see their use grow over time while the remaining 134 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 and trend 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.

Additionally, I supplement this data with hand-collected information on whether the Ameri-

[9] 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 6 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.

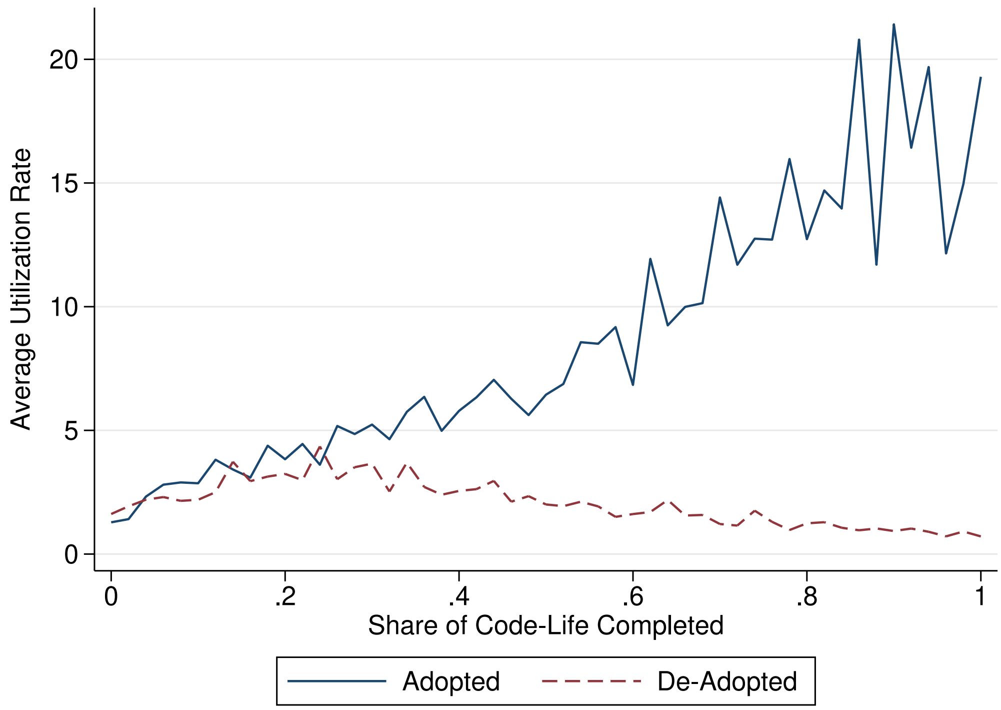
[10] Specifically, I estimate the model $Y_t = \beta_0 + \beta_1 T_t + \varepsilon_t$ separately for each procedure, where Y_t is the utilization rate of the procedure and T_t is the number of months since the introduction of the code covering the procedure. I classify each procedure as adopted if the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

Table 2: Coverage by Administrative Contractor

Administrator	Percentage of MAC-Procedure-Months Non-Covered	Case-by-Case	Covered
Novitas	76.9%	12.6%	10.5%
Cahaba	81.1%	13.5%	5.3%
NHIC	81.9%	12.9%	5.2%
Noridian	82.2%	6.3%	11.5%
Palmetto	82.3%	7.1%	10.6%
TrailBlazer	84.3%	10.5%	5.2%
HealthNow	86.2%	5.7%	8.1%
NGS	86.9%	7.4%	5.7%
FCSO	87.1%	8.4%	4.5%
Pinnacle	87.3%	7.6%	5.1%
WPS	87.4%	8.0%	4.6%
CGS	88.2%	7.1%	4.6%
Wheatlands	93.8%	3.2%	3.0%
Regence	98.0%	1.5%	0.5%
Group Health	98.3%	0.5%	1.2%
Triple-S	98.5%	0.8%	0.6%
BCBSMT	98.9%	0.1%	1.0%
Nationwide	99.0%	0.0%	1.0%
BCBSRI	100.0%	0.0%	0.0%
Overall	85.3%	8.3%	6.3%

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.

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.

can Medical Association has promoted or deleted the code covering each procedure. This complementary 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 made a decision by January 2022, 95 have been deleted while 146 have been promoted. This success rate of 60.6% is nearly identical to the 60.8% success rate implied by the more comprehensive measure based on utilization trends.[11]

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

[11]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.

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, estimating 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 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 28 more times per million beneficiaries each month relative to when it is non-covered, representing an increase over 5 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 from a less-generous coverage level to a more generous one. That is, I consider changes from non-coverage to coverage on a case-by-case basis or full coverage and changes from case-by-case coverage to full coverage.

Figure 3 reports estimates of β_e for specifications with and without time fixed effects in

Table 3: Effect of Coverage on Utilization

	(1) Utilization Rate	(2) Utilization Rate	(3) Any Uses	(4) Any Uses
Case-by-Case	6.704***	3.695*	0.0599***	0.0369***
	(1.939)	(1.635)	(0.00337)	(0.00320)
Covered	28.29**	18.44	0.122***	0.0561***
	(10.47)	(12.97)	(0.00705)	(0.00498)
Jurisdiction FEs	1	1	1	1
Time FEs	0	1	0	1
Dep. Var. Mean	4.870	4.870	0.0623	0.0623
Observations	1,240,092	1,240,092	1,240,092	1,240,092

Notes: An observation is a procedure-jurisdiction-month. Sample consists of monthly observations from 2002–2017. Utilization rate is the number of uses of the procedure per million beneficiary-months. 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.

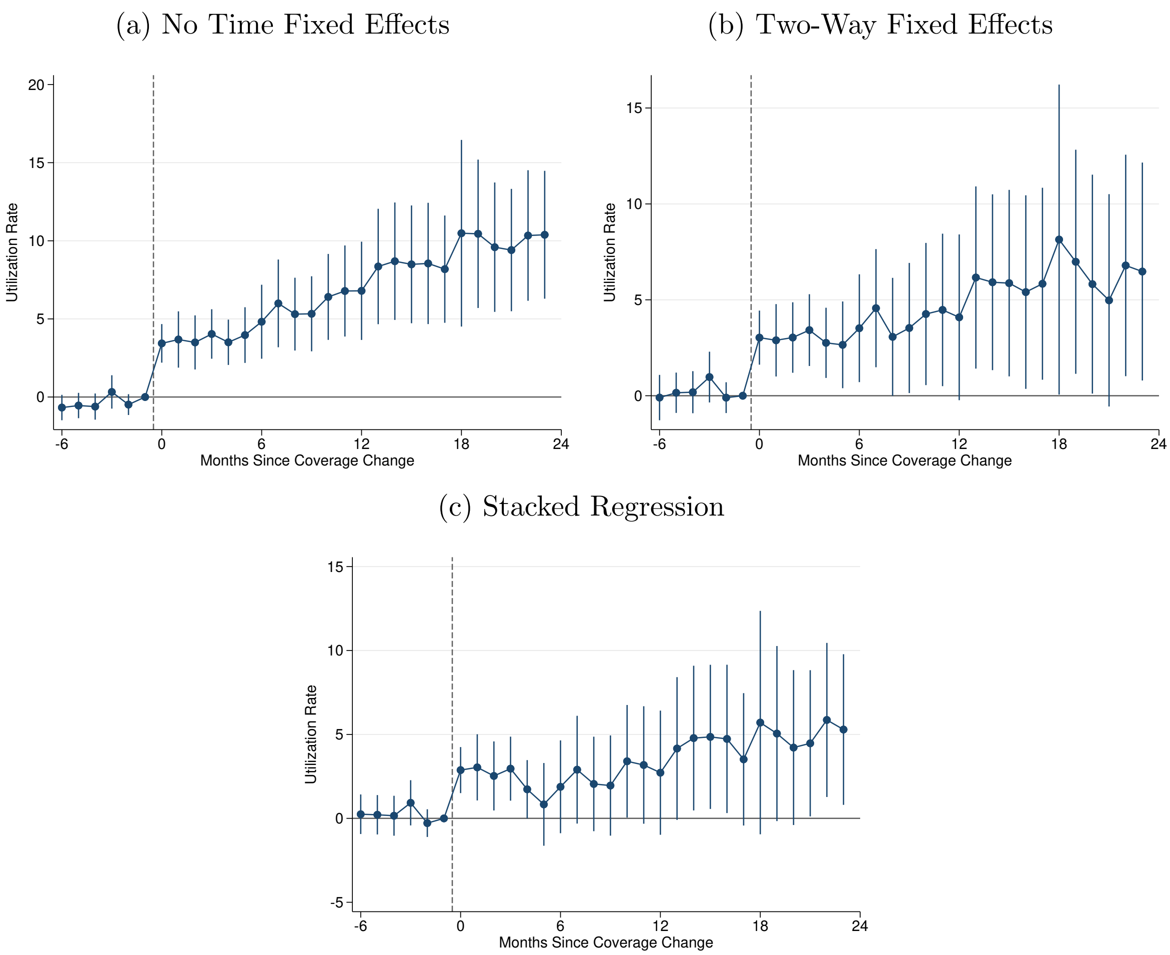
the model as well as estimated using stacked regression. Across all three specifications, we see consistent results indicating a lack of differential trends in utilization prior to coverage being extended followed by utilization increasing sharply immediately upon coverage and then gradually continuing 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.[12] In the next section of the paper, I will more closely examine this and other potential causes of the growth in utilization following coverage.

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

[12]Note that the attenuation the growth in the dynamic effects when including time fixed effects is consistent with the model presented in Section 6, with the knowledge created by the procedure's use being immediately incorporated into utilization decisions in both the jurisdiction where the procedure is performed and all other jursdictions.

Figure 3: Change in Utilization at Coverage Change



Notes: The figures report estimates of β_e from Equation (2) for $e \in \{-6, \ldots, 24\}$. An observation is a jurisdiction-procedure-month tuple. Utilization rate is the number of uses of the procedure per million beneficiary-months. 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.

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 not a control group without access to the common pool of information. To overcome this issue, 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. By contrast, social learning makes opposite predictions for the trends of adoption for innovations that are good or bad.[13] 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 the adoption of valuable procedures (and the de-adoption of low value ones) is more rapid the larger the jurisdictions in which the procedures are covered. This is consistent with coverage in larger jurisdictions providing the opportunity for more utilization and knowledge generation. The third piece of evidence is that additional past utilization causes more contemporary utilization for successful innovations but less utilization for unsuccessful ones. Here, I instrument for past utilization using the size of jurisidictions covering the procedure in the past, conditional on past administrator beliefs about the value of the procedure. In all cases, I find that additional utilization elsewhere (in time or space) leads to more rapid adoption of high value and de-adoption of low value innovations, consistent with the medical community learning about the value of these innovations through their use.

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 widely believed, for low value procedures the additional utilization will reveal that the innovation is worse

[13] 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.

than many of the providers already using the procedure believe, leading utilization to fall. Thus, for successful procedures, social learning predicts that granting Medicare 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

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

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 coverage 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 beyond 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 de-adopted procedures. 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.[14]

More parsimoniously, I can regress the utilization on a linear trend and the number of bene-

[14] In Appendix D, I show there is no differential response to coverage changes of providers in nearby jurisdictions, indicating that the knowledge spillovers are global in nature.

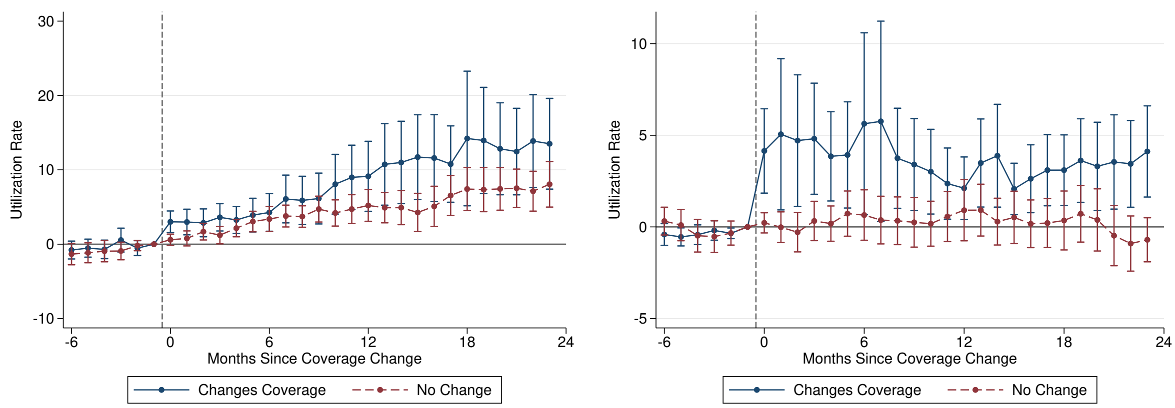
ficiaries for which the procedure is covered on at least a case-by-case basis, limiting the sample to jurisdictions in which the procedure is never covered:

$$Y_{pjt} = \beta_0 + \beta_1 t + \beta_2 t \times TotCov_{pt} + \gamma_{pj} + \varepsilon_{pjt}, \tag{4}$$

where t is the time variable and $TotCov_{pt}$ gives the extent of coverage of the procedure. β_2 is the coefficient of interest and reports the change in the trend of utilization in places that do not change coverage when coverage becomes more generous elsewhere. Table 4 reports estimates of β_2 separately for adopted and de-adopted procedures. As expected, we see that for adopted procedures, adoption becomes more rapid when coverage elsewhere becomes more generous while the opposite is true for de-adopted procedures.

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 adopted procedures, while panel (b) presents estimates for de-adopted procedures. Error bars give the pointwise 95% confidence intervals. Standard errors are clustered at the jurisdiction-by-procedure level.

5.2 Heterogeneity by Jurisdiction Size

The second piece of evidence of social learning I present is that the rate of adoption or de-adoption depends on the size of the jurisdictions in which the procedure is covered. In particular, I show that for adopted procedures, being covered in jurisdictions with more beneficiaries leads to more rapid increases in adoption, while for de-adopted procedures, coverage in larger jurisdictions results in more rapid de-adoption. This phenomenon can be explained by social learning as larger jurisdictions create more signals of a procedure's quality.

Table 4: Change in Trend of Utilization

	Adopted		De-Adopted	
	(1) Utilization Rate	(2) Any Uses	(3) Utilization Rate	(4) Any Uses
Trend $\times$ Total Coverage	0.00777*	0.0000243	-0.00489*	-0.000100***
	(0.00304)	(0.0000249)	(0.00239)	(0.0000249)
Jurisdiction FEs	1	1	1	1
Time FEs	0	0	0	0
Dep. Var. Mean	1.738	0.0398	0.611	0.0164
Observations	468,726	468,726	431,862	431,862

Notes: An observation is a procedure-jurisdiction-month. Sample consists of monthly observations from 2002–2017 for jurisdictions in which the procedure is never covered. Utilization rate is the number of uses of the procedure per million beneficiary-months. Total coverage is the total number of beneficiaries in millions for which the procedure is covered fully or on a case-by-case basis. Regressions include jurisdiction-by-procedure fixed effects. In columns (1) and (2), the sample is limited to promoted procedures, while in columns (3) and (4) it is limited to deleted procedures. Standard errors are clustered at the jurisdiction-by-procedure level. $^{+}$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

I estimate the equation

$$(5)\quad Y_{pjt} = \sum_{e=-6}^{-2} \beta_e T_{pjt}(e) + \sum_{e=0}^{24} \beta_e T_{pjt}(e) + \sum_{e=-6}^{-2} \phi_e T_{pjt}(e) \times Size_{pt} + \sum_{e=0}^{24} \phi_e T_{pjt}(e) \times Size_{pt} + \gamma_{pj} + \varepsilon_{pjt},$$

where $Size_{pt}$ is the total number of beneficiaries living in jurisdictions where procedure p is covered in month t. 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. I limit the sample to jurisidiction-procedure pairs for which the procedure is eventually covered, and I do not include time fixed effects in the specification. Both of these restrictions mean that this specification 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.

For a more parsimonious specification, I also estimate

$$(6) \quad Y_{pjt} = \beta_0 + \beta_1 Covered_{pjt} + \beta_2 MonthsToCoverage_{pjt} + \beta_3 Size_{pt}$$
$$+ \beta_4 Covered_{pjt} \times MonthsToCoverage_{pjt} + \beta_5 Covered_{pjt} \times Size_{pt} + \beta_6 MonthsToCoverage_{pjt} \times Size_{pt}$$
$$+ \beta_7 Covered_{pjt} \times MonthsToCoverage_{pjt} \times Size_{pt} + \gamma_{pj} + \varepsilon_{pjt},$$

where $MonthsToCoverage_{pjt}$ is a continuous variable reporting the number of months until procedure p is first covered on at least a case-by-case basis in jurisdiction j. Here, the coefficient of interest is β_7, which gives the differential change in the trend in utilization by the number of beneficiaries for which the procedure is covered. If social learning is a key driver of adoption patterns, we should see that the rate of adoption should increase more rapidly in larger jurisdictions (in which more signals of the procedure's value are produced) for successful procedures but not for unsuccessful ones. That is, because coverage in larger jurisdictions leads to greater utilization, larger jurisdictions should facilitate more rapid adoption of successful procedures and more rapid de-adoption of unsuccessful ones.

Table 5 presents estimates of β_7 in Equation (6) 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, while the opposite is true for de-adopted procedures. This result is corroborated by Figure 5, which presents estimates of ϕ_e in Equation (5). We against see that for adopted procedures, utilization grows more for larger jurisdictions than smaller ones. 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 no such pattern. For these procedures, there is a larger jump in utilization when the procedure is covered in larger jurisdictions, but see that utilization grows no more quickly thereafter. 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 over time. Overall, these results indicate that adoption and de-adoption are more rapid when there is greater opportunity to generate signals of the procedure's true value, consistent with social learning.

5.3 Effect of Past Use

Beyond examining spillovers of utilization across *space*, I can also investigate the presence of spillovers across *time*. Just as utilization in one jurisdiction may generate information accessible in others, more early utilization should create information that is incorporated into later utiliza-

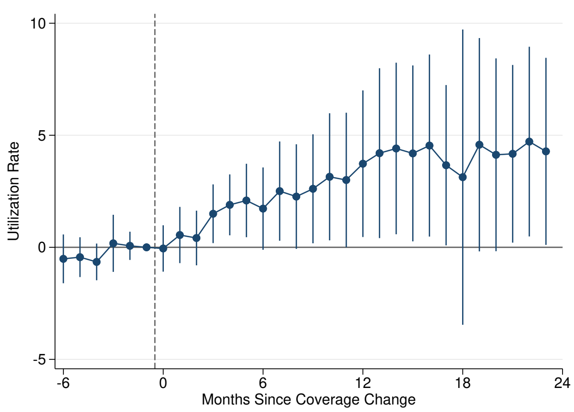
Table 5: Change in Trend of Utilization at Coverage Change, Heterogeneity by Size

	Adopted		De-Adopted	
	(1) Utilization Rate	(2) Any Uses	(3) Utilization Rate	(4) Any Uses
Trend Change × Size	0.204***	0.000797***	-0.0367***	-0.000499***
	(0.0453)	(0.0000753)	(0.0104)	(0.0000753)
Jurisdiction FEs	1	1	1	1
Time FEs	0	0	0	0
Dep. Var. Mean	19.14	0.181	5.615	0.0941
Observations	225,876	225,876	113,628	113,628

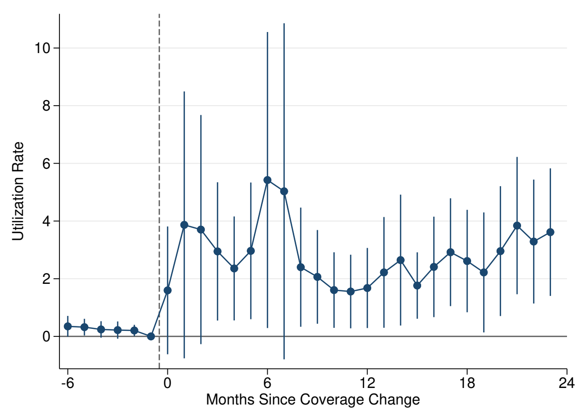
Notes: Estimates of β_7 in Equation (6). An observation is a procedure-jurisdiction-month. Sample consists of monthly observations from 2002–2017 for procedure-jurisdiction pairs in which the procedure is ever covered. Utilization rate is the number of uses of the procedure per million beneficiary-months. Size is the number of beneficiaries in millions for which the procedure is covered fully or on a case-by-case basis. Regressions include jurisdiction-by-procedure fixed effects. In columns (1) and (2), the sample is limited to adopted procedures, while in columns (3) and (4) it is limited to de-adopted procedures. Standard errors are clustered at the jurisdiction-by-procedure level. $^{+}$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

Figure 5: Effect of Coverage on Utilization, Heterogeneity by Size

(a) Adopted Procedures



(b) De-Adopted Procedures



Notes: The figures report estimates of ϕ_e from Equation (5) for $e \in \{-6, \ldots, 24\}$. An observation is a procedure-jurisdiction-month. Utilization rate is the number of uses of the procedure per million beneficiary-months. Panel (a) presents estimates for the sample limited to adopted procedures, while panel (b) presents estimates for de-adopted procedures. Error bars give the pointwise 95% confidence intervals. Standard errors are clustered at the jurisdiction-by-procedure level.

tion patterns. To do this, I estimate the effect of past utilization on current utilization, using past coverage policies as an instrumental variable. The idea is that past coverage rules have affected the experience the medical community has with a procedure, which will (potentially) reveal information about the procedure's true value that is reflected in later utilization.

This instrumental variables framework requires the instrument satisfy two key assumptions: exclusion and relevance. The exclusion restriction requires that past coverage policy can only be correlated with currect utilization through its affect on past utilization. The most obvious potential violation of this assumption would be if past coverage policy and current utilization are both correlated with the true value of the procedure. To ensure this is not the case, I parameterize past coverage policy as the number of beneficiaries living in jurisdictions where the procedure is covered *conditional on the share of contractors covering the procedure*. In this way, the instrument does not rely on variation in how favorably administrators treat the procedure, but only on variation in the size of the jurisdictions assigned to each administrator. By controlling for the share of contractors covering the procedure, I am controlling for contemporary beliefs among administrators about the value of the procedure, with the instrument relying only on variation in whether the procedure happens to be initially covered in larger or smaller jurisdictions.

Another potential violation of the exclusion restriction could be persistence of coverage rules over time such that past coverage affects current utilization through its effect on current coverage rules, regardless of the level of past utilization. To address this worry, I limit the sample to jurisdictions in which the procedure is not covered. In Appendix E, I discuss this potential exclusion restriction violation in more detail and demonstrate the robustness of my results to including jurisdictions where the procedure is covered. In the same appendix, I discuss other potential worries about the exclusion restriction as well, including concerns that the instrument may be correlated with changing average procedure quality over time.

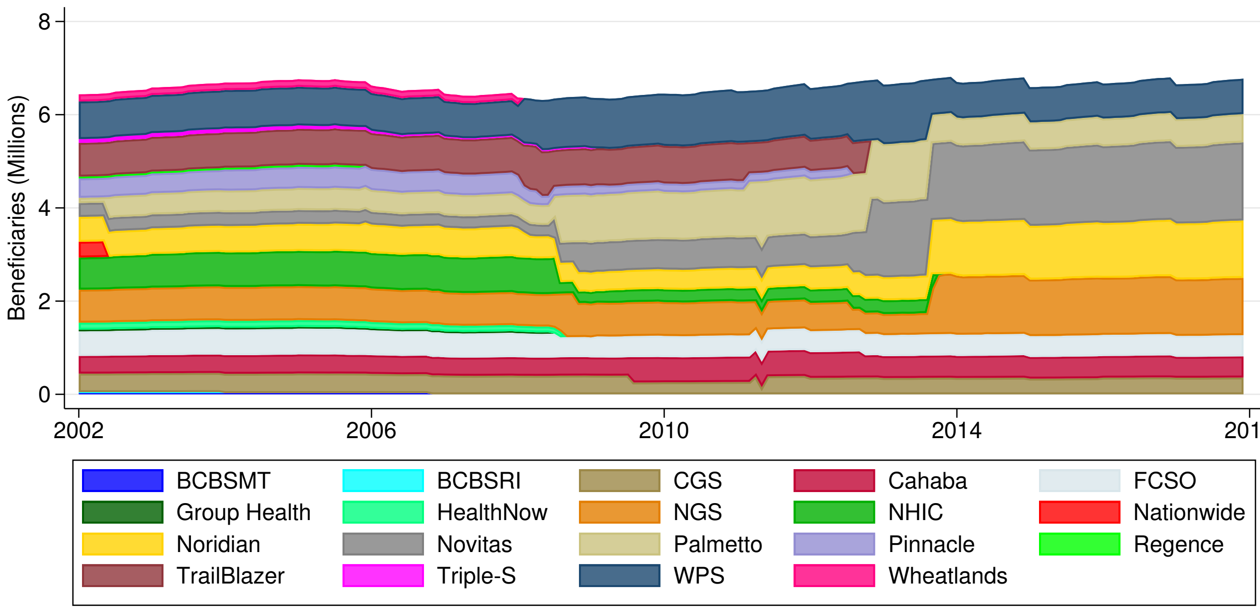
Figure 6 shows the variation I use as an instrument. The figure shows the size of the jurisdictions covered by each administrator, with wide variation in all time periods, particularly toward the beginning of my sample period. I use these differences in the number of providers subject to the coverage determinations of each administrator as an instrument for early utilization of a procedure.

Next, I present evidence that the instrumental variable is relevant: past utilization is strongly correlated with past coverage policy. To do this, I estimate the first-stage equation

$$(7)\quad Y^c_{pjt} = \alpha_0 + \alpha_1 Size^{c,cov}_{pt} + \alpha_2 Size^{c,case}_{pt} + \alpha_3 Share^{c,cov}_{pt} + \alpha_4 Share^{c,case}_{pt} + \alpha_5 Benes^c_{pt} + \eta_{pj} + \nu_{pjt},$$

where Y^c_{pt} is the number of cumulative past uses of procedure p before time t, $Size^{c,case}_{pt}$ ($Size^{c,cov}_{pt}$) is the cumulative number of beneficiary-months for which the procedure has been covered on a case-by-case basis (fully), $Share^{c,case}_{pt}$ ($Share^{c,cov}_{pt}$) is the share of total administrator-months in which the procedure has been covered on a case-by-case basis (fully), $Benes^c_{pt}$ is the cumulative number of beneficiary-months for which the procedure could have been covered, η_{pj} is a set of procedure fixed effects, and ν_{pt} is the econometric error term, which I allow to be correlated

Figure 6: Population Covered by Each Administrative Contractor



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

within jurisdiction-procedure pair.[15] The coefficient of interest is α_1, which gives the strength of the relationship between the instrumental and instrumented variables. Table 6 reports this coefficient separately for adopted and de-adopted procedures, measuring both the instrumental and instrumented variables in levels and subject to the inverse hyperbolic sine transformation. We can see that across all specifications, the instrument is extremely strong, with greater past coverage leading to greater past utilization.

The second-stage equation is

$$Y_{pjt} = \beta_0 + \beta_1 Y^c_{pt} + \beta_2 Share^{c,cov}_{pt} + \beta_3 Share^{c,case}_{pt} + \beta_4 Benes^c_{pt} + \gamma_{pj} + \varepsilon_{pjt}, \tag{8}$$

where Y_{pjt} is the contemporary utilization rate. β_1, the coefficient of interest, gives the effect of past utilization on current utilization. Table 7 presents estimates of this parameter using the same specifications as Table 6. Consistent with social learning, we see that for successful innovations, more early utilization positively affects later utilization, while the opposite is true for unsuccessful procedures. Variation in past use is isolated to that caused by the size of the jurisdictions in which it is covered, meaning that we see that for successful innovations, more

[15] By including procedure fixed effects, I am isolating variation over time within a procedure. This controls for the prevalence of the disease treated by the procedure, among other time-invariant procedure characteristics. By including the total past beneficiary-month count, I'm controlling for changes over time in the total number of traditional Medicare beneficaries as well as progression throughout the life of the code. In Appendix E, I demonstrate robustness of my results to these choices.

Table 6: Effect of Past Coverage on Past Utilization

	Adopted		De-Adopted	
	(1) Utilization Rate	(2) Utilization Rate (Asinh)	(3) Utilization Rate	(4) Utilization Rate (Asinh)
Past Case-by-Case Coverage	34.23***	0.685***	-1.515**	0.521***
	(8.677)	(0.0109)	(0.549)	(0.00863)
Past Full Coverage	107.0***	0.569***	24.36***	0.497***
	(7.468)	(0.0218)	(3.060)	(0.0107)
Dep. Var. Mean	650.9	3.001	342.1	3.570
Observations	569,006	569,006	472,945	472,9456

Notes: Ordinary least-squares estimates of α_1 and α_2 from Equation (7). An observation is a jurisdiction-procedure-month tuple in which the procedure is not covered. Utilization rate is the number of uses of the procedure per million beneficiary-months. In columns (1) and (2), the sample is limited to adopted procedures, while the opposite is true for columns (3) and (4). In columns (1) and (3), past coverage is measured in millions of beneficiary-months and past uses in levels, while in columns (2) and (4), past coverage is measured as the inverse hyperbolic sine of beneficiary-months and past uses and the utilization rate are similarly transformed. All specifications include controls for the share of past MAC-months in which the procedure was covered fully, the share for which it was covered on a case-by-case basis, the total past beneficiary-months the procedure has been available, and jurisdiction-by-procedure fixed effects. Standard errors are clustered at the jurisdiction-procedure level. $^{+}$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

favorable coverage in the past leads to more past utilization, which leads to more contemporary utilization, while the opposite is true for de-adopted procedures. In terms of magnitudes, I find that a 1% increase in past utilization causes 0.02% more contemporary use for adopted procedures but 0.02% less use for de-adopted ones. Again, the heterogeneity in the effect of past utilization by whether the procedure is eventually found to be effective is key evidence that the mechanism is social learning: for procedures that are found ex post to be of low value, more generous early coverage leads to *less* utilization later on.

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

Table 7: Effect of Past Utilization on Current Utilization

	Adopted		De-Adopted	
	(1)	(2)	(3)	(4)
	Utilization Rate	Utilization Rate (Asinh)	Utilization Rate	Utilization Rate (Asinh)
Past Uses	0.000847***	0.0243**	-0.000589	-0.0175**
	(0.000105)	(0.00878)	(0.000404)	(0.00670)
First-Stage F Stat.	106.0	3981.3	36.23	3641.1
Dep. Var. Mean	2.071	0.172	0.636	0.0611
Observations	569,006	569,006	472,945	472,945

Notes: Two-stage least-squares estimates of β_1 and from Equation (8). An observation is a jurisdiction-procedure-month tuple in which the procedure is not covered. Utilization rate is the number of uses of the procedure per million beneficiary-months. In columns (1) and (2), the sample is limited to adopted procedures, while the opposite is true for columns (3) and (4). In columns (1) and (3), past coverage is measured in millions of beneficiary-months and past uses in levels, while in columns (2) and (4), past coverage is measured as the inverse hyperbolic sine of beneficiary-months and past uses and the utilization rate are similarly transformed. All specifications include controls for the share of past MAC-months in which the procedure was covered fully, the share for which it was covered on a case-by-case basis, the total past beneficiary-months the procedure has been available, and jurisdiction-by-procedure fixed effects. Standard errors are clustered at the jurisdiction-procedure level. $^{+}$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

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.[16] 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.[17] These points cast doubt on change in the actual value of the innovations being first-order.[18]

Another set of alternative explanations is various types of non-social learning. The first of these is learning from one's own experience. In many areas of medicine, including the use of some new medical procedures, providers have been shown to learn from their own experience, both in terms of achieving better patient health outcomes (Gowrisankaran et al., 2006; Hockenberry and Helmchen, 2014; Gong, 2017; Yang, 2022) as well as acquiring more accurate beliefs about the treatment (Coscelli and Shum, 2004; Crawford and Shum, 2003; Ferreyra and Kosenok, 2011). However, in this context, I see little evidence that, on average, providers learn from their own

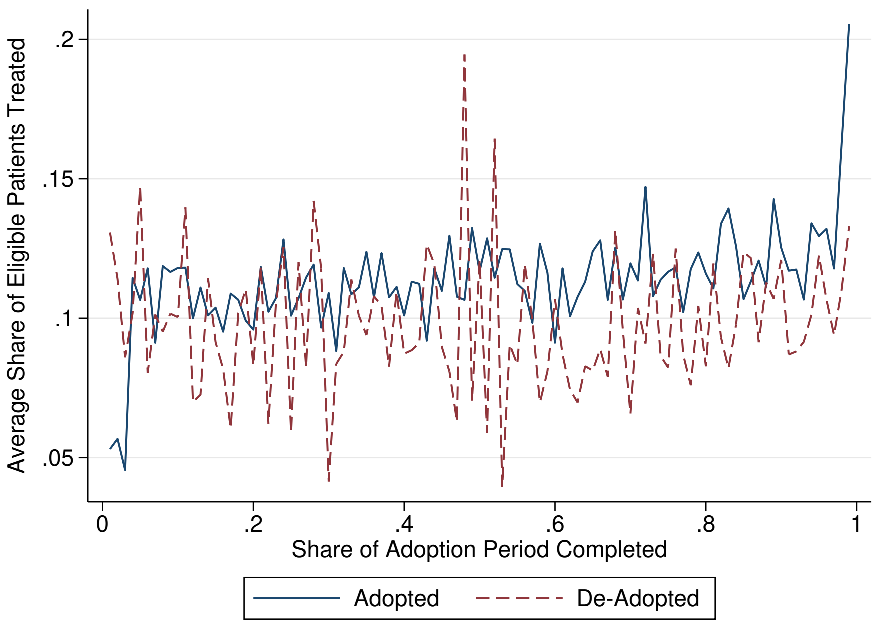
[16] DGHW (2022) find that just over half of approved category III CPT applications from 2008 and 2017 were associated with new devices.

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

[18] I should also 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.

experience. As shown by Figure 7, there is no within-provider growth in use as the provider gains more experience with the procedure. If providers were gaining ability as they used the procedure, then they would likely increase their use of the procedure. Furthermore, the (lack of) within-provider trend in utilization does not differ by the adoption status of the procedure, indicating that learning about procedure quality from one's own experience alone is unlikely to explain the patterns of adoption I observe.

Figure 7: Within-Provider Change in Use



Notes: Figure presents the share of patients in the consideration set treated using the new procedure among providers that have previously used the procedure and will use the procedure again in the future. The time from the provider's first to last use of the procedure is normalized to one. The consideration set is all patients with a primary diagnosis shared by any patient ever treated with the procedure who are treated by a physician whose specialty is that of any physician who ever uses the procedure. The horizontal axis gives the share of months between the provider's first and last use of the procedure. The solid blue line is limited to adopted procedures, while the dashed red line is for de-adopted procedures.

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., 2024). 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 DGHW (2022), from 2008 to 2017 only 20% of the procedures approved by the American Medical Association were supported by randomized controlled trials. Futhermore, in my data less than 1% of claims for the procedures I study are for care administered as part of a clinical trial. 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 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). These studies form an important foundation for the model of physicians as Bayesian learners updating their beliefs about the value of new innovations 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 information spillovers from each 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 for 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 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 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.

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.

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 patients for whom it is appropriate.[19] For procedure p, the share of appropriate patients is given by η_p.

The initial beliefs of physicians are distributed normally (independent of physician type) with a mean of δ_{0p} and a standard deviation of $\sigma_{\delta p}$.[20] Note that δ_{0p} may or may not equal δ_p^* (physicians may be optimistic or pessimistic) and $\sigma_{\delta p}$ captures the level of initial uncertainty about the efficacy of the new procedure.

Each time a physician uses 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 to 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 on the precision of the signals generated through the procedure's use such that the more precise 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

[19]This assumption is consistent with the results in Figure 7 that shows no change the share of patients for which a provider uses the new procedure as their experience with the procedure grows.

[20]In terms of notation, i denotes each physician such that $\delta_{0ip} \sim \mathcal{N}(\delta_{0p}, \sigma_{\delta p})$.

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 the n_{pt} signals $\{\nu_{ipk}\}_{k\in\{1,\ldots,n_t\}}$ generated by time t, the beliefs of physician i are distributed normally with mean μ_{ipt} and standard deviation σ_{ipt}, where

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

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, $\sigma_{ipt} = \sigma_{agg,pt}$ for all i, while μ_{ipt} is distributed normally with standard deviation $\sigma_{agg,pt}$ and mean

$$\mu_{agg,pt} = \sigma_{agg,pt}^2 \left(\frac{\delta_{0p}}{\sigma_{\delta p}^2} + \frac{n_{pt}\delta_p^*}{\sigma_{\nu p}^2} \right).$$

Proofs that physician beliefs are so distributed are given in Appendix F.

Administrators are also initially uncertain about the value of the new procedure and make coverage decisions based on their beliefs about its value. The utility of administrator a of covering procedure p at time t is given by

$$A_{apt} = \mathbb{E}_{apt}\left[\delta_p^*\right] - s_{apt}, \tag{9}$$

where s_{apt} is administrator a's evidentiary standard for procedure p. While I do not model how administrators choose their coverage standards s_{apt}, administrators may fail to cover a procedure physicians believe to be better than the outside option or decide to cover one physicians believe to be worse for any number of reasons, including placing different weights on patient health outcomes, the cost of the procedure, or how easy it is for providers to perform the procedure, for example. Like for physicians, I assume administrators make their coverage decisions statically, ignoring the role that their coverage decisions may play in facilitiating the creation of knowledge about the value of the procedure. Thus, each administrator will cover the procedure in months in which the administrator's beliefs about the value of the procedure exceed their coverage threshold.

Patient welfare is given by

$$W_{iptj} = \eta_{pj}(\delta_p^* - X_{ip}), \tag{10}$$

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. Notice that welfare losses arise because of uncertainty about the true value of δ_p^*: Physicians may make different adoption decisions under uncertainty than they would if they were correct about the procedure's value. Were there no uncertainty about the value of the procedure, the only providers that would adopt the procedure would be the ones that *should* adopt it, even if the procedure is much worse than the outside option for the average physician. However, as the variance of beliefs rises or physicians become more optimistic about a procedure's value, patient welfare per use of the new procedure falls and can be negative. This uncertainty will result in physicians unable to effectively employ the procedure performing it and physicians who are able foregoing the opportunity to use it, potentially creating a role for the administrator in limiting coverage.[21]

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. Proof of model identification is given in Appendix F.

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 make a different coverage decision. I also restrict the sample to procedures used at least 30 times during my sample period. Finally, I restrict the sample to the procedures for which I can separately estimate η_p using repeated utilization among adopting physicians. This restriction results in a sample of 136 procedures.

I estimate the parameters of the physician adoption decision other than η_p by maximum likelihood estimation. I separately estimate η_p using the share of patients treated by providers that have adopted the procedure, as discussed in Appendix F. The log-likelihood function for all

[21] In Appendix F, I prove the claims about patient welfare made in this paragraph.

other parameters is given by

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

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

where

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

Note that $g_{cov,pt}$ and $y_{cov,pt}$ are the total number of patients in the consideration set for, and the total number of patients being treated with, procedure p in jurisdictions with coverage level *cov* at time t.[22] 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 are treated by a physician whose specialty is that of any physician who ever uses the procedure.

With physician beliefs in hand, the model of administrator behavior can be estimated using variation in coverage decisions. I parameterize administrator a's evidentiary standards for procedure p as

$$s_{Case,apt} = \theta_a + \xi_p + \varepsilon_{apt}, \qquad s_{Full,apt} = s_{Case,apt} + \kappa \tag{11}$$

where θ_a and ξ_p are administrator and procedure fixed effects and ε_{apt} is a normally distributed, mean-zero error term. $s_{Case,apt}$ is the administrator's evidentiary standard for case-by-case coverage while $s_{Full,apt}$ is the standard for full coverage. With this parameterization, Equation (9) can be estimated by ordered probit where the dependent variable is an ordinal categorical variable for the level of coverage and the independent variables are the mean contemporary belief about the value of the procedure and procedure and administrator fixed effects, with the coefficient on beliefs normalized to 1. θ_a and ξ_p are identified by the likelihood of coverage by each administrator and for each procedure given beliefs, and κ is identified by the difference in the likelihood of full coverage relative to case-by-case coverage.[23]

With the parameters relating to physician learning and administrator coverage decisions 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 $\mathcal{W}_{cov,pt}$ from procedure p under coverage regime *cov* at time

[22]The derivation of this likelihood function is presented in Appendix F.

[23]Note that in the parameterization, administrators have coverage standards that are subject to idiosyncratic shocks and all have beliefs that match the mean contemporary beliefs among physicians. In terms of identification and estimation, this model is identical to one in which the econometric error attaches to administrator beliefs rather than to the coverage standards.

t is given by

$$\mathcal{W}_{cov,pt} = y_{cov,pt}\mathbb{E}[W_{iptj}|U_{ipt} > 0, \eta_{pj} = 1] = y_{cov,pt}\left(\delta_p^* + \frac{1}{\sqrt{\sigma_{agg,pt}^2 + 1}}\left(\frac{\phi(\theta_{cov,pt})}{\Phi(\theta_{cov,pt})}\right)\right)$$

where $\theta_{cov,pt} \equiv \frac{\mu_{agg,pn} - \beta_{1p}Case_{ipt} - \beta_{2p}Noncov_{ipt}}{\sqrt{\sigma_{agg,pn}^2+1}}$.[24] Total welfare is given by the sum of $\mathcal{W}_{cov,pt}$ over coverage policies, procedures, and time periods.

6.2 Estimation Results

Table 8 reports summary statistics of the estimated parameters for each procedure as well as estimated administrator risk aversion and the MAC-by-procedure coverage thesholds.[25] 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 that β_{2p} is positive for 80% of procedures and is statistically significantly negative at the 95% confidence level for only 4.5% of procedures. These results are consistent with the reduced form evidence presented in Section 4 that more generous Medicare coverage increases utilization.

The estimates of $\sigma_{\delta p}$ and $\sigma_{\nu p}$ indicate significant uncertainty about the value of each procedure. When a procedure is introduced, the distribution of beliefs across providers is 94 times wider than the distribution of ability to perform the procedure, on average. And while additional utilization does generate meaningful information about the procedure's value, the standard deviation of this signal is over 20 times greater than that of physicians' priors. This means that it takes 438 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 for 6% of them on average.

Interpreting the δ parameters, we see that on average providers consider these new procedures to be worse than the outside option, although they initially believe they are much worse than they actually are. The estimate of δ_p^* is positive for only 9.6% of procedures and statistically significantly negative for 55.2% of procedures, meaning that most new procedures are worse than the incumbent procedure for the median physician. Importantly, as discussed above and proven in Appendix F, a negative estimate of δ_p^* does not alone mean that there are not welfare gains from the availability of the procedure, just that the median physician should not adopt it.

The full distributions of providers' initial beliefs and the uncertainty-free value of each procedure are presented in Figure 8. The distributions are quite similar, although physician priors are much more likely to be very negative, and the true values are more likely to be positive.

[24] I prove that the second equality holds in Appendix F.

[25] I present evidence that the model with these parameters fits the data well in Appendix G.

Table 8: Summary Statistics of Parameter Estimates

Parameter	Mean	Std. Dev.
Physician Learning Model		
β_{1p}	11.72	46.06
β_{2p}	39.58	145.47
$\sigma_{\delta p}$	93.92	317.71
$\sigma_{\nu p}$	1965.51	7846.74
δ_p^*	-43.02	483.64
δ_{0p}	-288.45	954.20
η_p	0.060	0.101
Coverage Model		
$s_{Case,apt}$	3395.35	4899.58
$s_{Full,apt}$	7028.05	4899.58

Notes: Estimates of structural model parameters. An observation for the physician learning model panel is a procedure, with the sample limited to the 136 procedures for the model is identified. The coverage model panel reports summary statistics for estimates of $s_{1,apt}$ and $s_{2,apt}$ where an observation is an administrator-procedure pair.

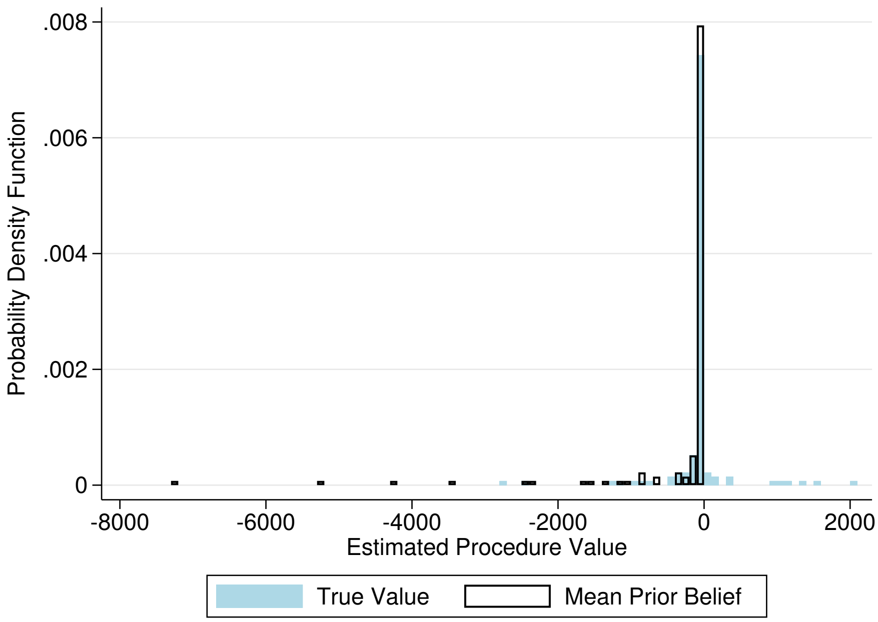
The distribution of priors being below that of the true values is consistent with physicians being pessimistic about the value of new procedures, on average. I estimate that the median physician is pessimistic about 89.7% of procedures and only significantly optimistic for 1.5 percent.[26] Furthermore, I estimate that physicians are overconfident in these beliefs. For example, for only 40.4%, 18.4%, and 6.6% of procedures does the median physician have the true procedure value in their 99%, 95%, and 90% confidence intervals of beliefs, respectively. Furthermore, the correlation between physicians' prior beliefs and the true value of the procedure is only 0.31, indicating that physicians are initially not very well informed about each procedure's value.

Finally, turning to the model of administrators' coverage decisions, we first see that administrators have a much higher threshold for covering a new procedure that its being better than the incumbent procedure for the median physician, with the average coverage threshold being well above 0. That said, there is substantial disagreement among administrators about the value of a given procedure, with the standard deviation of the coverage standard being larger than the mean threshold for case-by-case coverage. Furthermore, we see that the standard for full coverage is roughly twice as high, on average, as the standard for case-by-case coverage.

Furthermore, there is meaningful variation across administrators in perceived procedure value.

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

Figure 8: Estimated Distribution of Physician Priors and True Procedure Values



Notes: The light blue line bars report the estimated probability density of δ_p^*, while the white bars report the estimated probability density of δ_{0p}. An observation is a procedure. The sample is limited to the 136 procedures for which the model is identified.

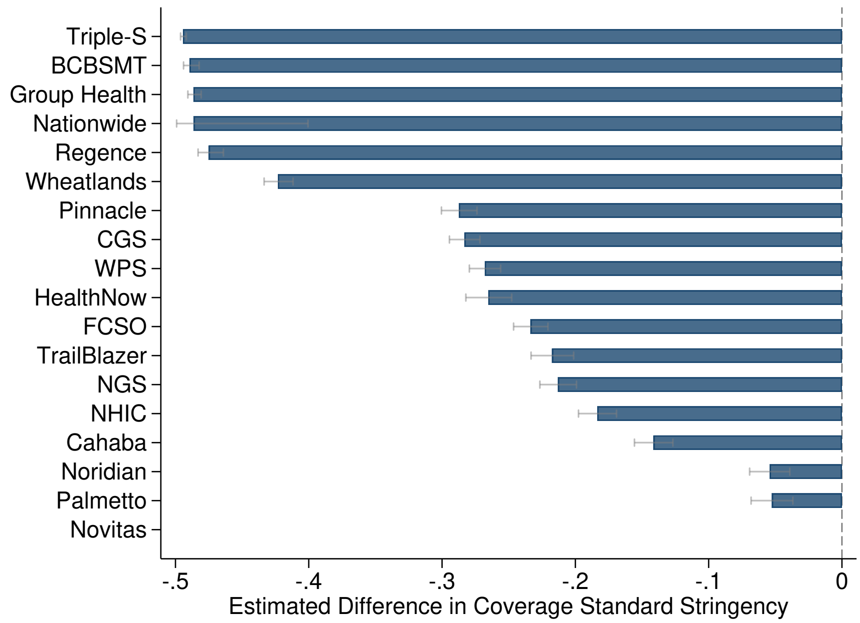
Figure 9 reports the estimated differences in the probability of coverage for each administrator for a marginal new procedure relative to Novitas, the administrator most generous with coverage of new procedures. I find wide heterogeneity in the propensity of each administrator to cover new procedures. 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.[27]

6.3 Welfare

I consider welfare under a range of potential coverage levels. First, I estimate the level of welfare under the observed average administrator coverage threshold. Then I consider welfare under counterfactual coverage rules, including universal coverage and universal non-coverage of new procedures. I then trace out counterfactual welfare under each administrator's coverage standards were that administrator's standard universally adopted. I compare welfare under these scenarios to the possible welfare gains from contractors implementing the static welfare-maximizing and welfare-minimizing coverage policies, which would require administrators to

[27]Note that the coverage thresholds for Blue Cross and Blue Shield of Rhode Island (BCBSRI) are not identified as this administrator is never observed covering a new procedure.

Figure 9: Estimated Differences in Coverage Probabilities Across Administrators



Notes: The figure reports estimates of θ_a from Equation (11), transformed to marginal effects. An observation is a MAC-procedure-month tuple. Dependent variable is an indicator for coverage of the procedure. Error bars give the 95% confidence intervals. Standard errors are clustered at the procedure-month level.

know the true value of each procedure while internalizing physician beliefs in each period.[28] These infeasible coverage policies serve as benchmarks against which more achievable coverage policies can be compared.

Table 9 reports welfare estimates under the status quo coverage policy and universal coverage or non-coverage, scaled such that welfare under the static welfare-maximizing policy is normalized to one and welfare-minimizing policy to zero. I find that welfare under the current coverage rules acheives only 31.9% of the possible welfare gains of moving from the welfare-minimizing to welfare-maximizing policy. I find that universal non-coverage would lower welfare to 2.2% of the welfare-maximizing policy, while universal coverage of the procedures in my sample would achieve 97.7% of the gains.[29] These results indicate that a policy of more generous coverage would achieve greater welfare than the status quo coverage policies.

That said, universal coverage is quite far out of sample, so there may be worries that the model is not suitable for such an extrapolation. Furthermore, we may worry that such a dramatic

[28]The static welfare-maximizing policy is to choose the coverage level (non-coverage, case-by-case coverage, or full coverage) that generates the highest total welfare in each period. I present total welfare in terms of model parameters in Appendix F.

[29]Recall that the sample is limited to procedures for which there was some coverage, meaning my counterfactuals consider changes to coverage only among the sample of plausibly covered procedures.

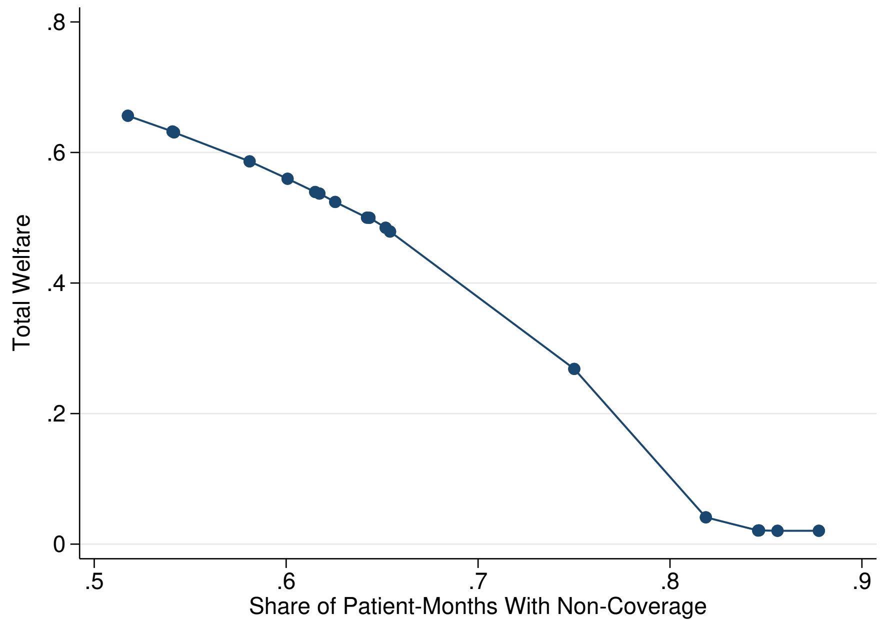
Table 9: Welfare Estimates

	Welfare
Welfare-Minimizing Policy	0.000
Status Quo	0.319
Universal Non-Coverage	0.022
Universal Coverage	0.977
Welfare-Maximizing Policy	1.000

Notes: Estimates of total welfare from simulations of the model with current and counterfactual coverage rules. Sample is limited to the 136 procedures for which the model is identified. Welfare is scaled such that welfare under the welfare-minimizing coverage policy is zero and welfare under the welfare-maximizing coverage policy is one.

change in coverage policy would alter the quality or rate of arrival of new procedures, which I do not model. Given the lack of property rights encouraging the development of new medical procedures (DGHW 2022) and my limitation of the sample to procedures that administrators current give strong consideration to covering, loosening coverage rules to cover all the procedures in my sample would be unlikely to dramatically alter upstream procedure innovation. That said, in order to assess counterfactuals closer to the observed coverage behavior, I also estimate welfare were the coverage threshold of each administrator observed in the data adopted universally. Given that these coverage thresholds have been employed by administrators, they represent a range of thresholds that are certainly within the range policymakers would consider adopting. Figure 10 gives estimated welfare under each of these coverage thresholds. The figure shows that welfare is dramatically increasing in the leniency of the threshold, with the most generous coverage threshold achieving 66% of the static welfare-maximizing policy. Adopting this threshold rather than the current average among administrators would acheive 50% of the welfare gains of moving to the infeasble, optimal coverage policy, indicating that lowering coverage thresholds and increasing coverage would improve welfare.

Figure 10: Welfare Under Difference Coverage Thresholds



Notes: The figure reports estimates of total welfare under counterfactual coverage thresholds. An observation corresponds to the estimated threshold of an administrator. Sample is limited to the 136 procedures for which the model is identified. Welfare is scaled such that welfare under the welfare-minimizing coverage policy is zero and welfare under the welfare-maximizing coverage policy is one.

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 impact the diffusion of these procedures, I present 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 regulators must consider when determining whether to allow the spread of innovations. In order to quantify this externality and determine the welfare effects of counterfactual coverage policies in this context, I estimate a structural model of provider learning. The results of this model indicate that reductions in administrator coverage standards and more generous coverage of new medical procedures could lead to large welfare gains.

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. von Hinke, B. Lagerqvist, C. Propper, and J. Vikström (2024). Do responses to news matter? evidence from interventional cardiology. *Journal of Health Economics 94*, 102846.

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.

Brot-Goldberg, Z. C., S. Burn, T. Layton, and B. Vabson (2023). Rationing medicine through bureaucracy: authorization restrictions in medicare. Technical report, National Bureau of Economic Research.

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.

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.

Davis, D. J. and A. C. Kim (2025, April). White house embraces crypto as sec opens door for banks. *The National Law Review XV*(111). Accessed: 2025-04-21.

De Almeida, P. G. R., C. D. dos Santos, and J. S. Farias (2021). Artificial intelligence regulation: a framework for governance. *Ethics and Information Technology 23*(3), 505–525.

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 (2022). The economics of medical procedure innovation. *Journal of Health Economics 81*, 102549.

Dranove, D., C. Garthwaite, and B. Wu (2022). The returns to medical inventions. *The Journal of Law and Economics 65*(S2), S389–S417.

Dunn, A., J. D. Gottlieb, A. H. Shapiro, D. J. Sonnenstuhl, and P. Tebaldi (2023, 06). A Denial a Day Keeps the Doctor Away. *The Quarterly Journal of Economics*, qjad035.

Eliason, P., R. League, J. Leder-Luis, R. C. McDevitt, and J. W. Roberts (2025). Ambulance taxis: The impact of regulation and litigation on health-care fraud. *Journal of Political Economy 133*(5), 1661–1702.

Fafchamps, M., M. Söderbom, and M. van den Boogart (2022). Adoption with social learning and network externalities. *Oxford Bulletin of Economics and Statistics 84*(6), 1259–1282.

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. and M. Shi (2025). Administrative fragmentation in health care. Technical report, National Bureau of Economic Research.

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.

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

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. (2024). Monitoring for waste: Evidence from medicare audits. *The quarterly journal of economics 139*(2), 993–1049.

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.

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.

Online Appendix

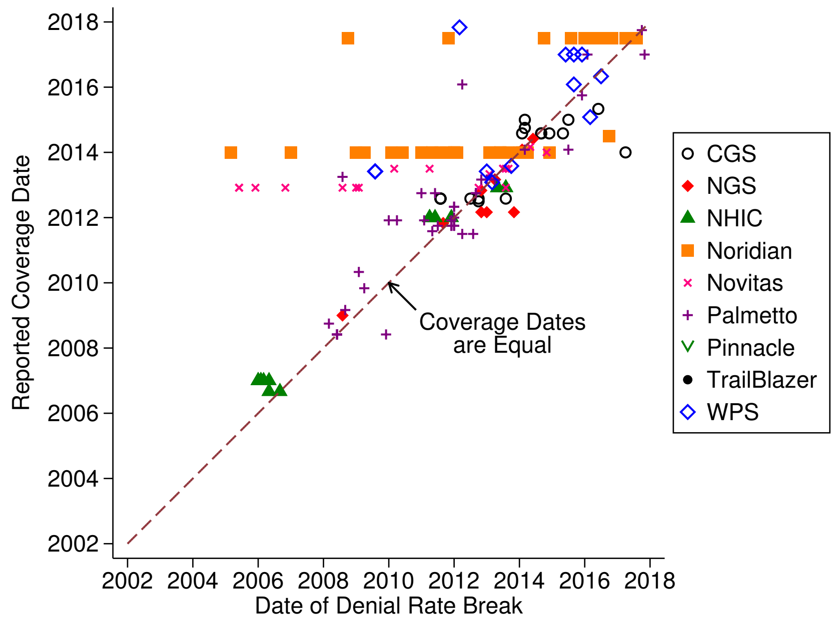
A Determining Coverage Rules

Medicare Administrative Contractors (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 apparatus is highly systemized, with claims usually 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 determinations. 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 19 administrative companies were active during my sample period, only 9 of them have posted publicly available coverage rules for category III codes. 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 the level and trend of denial rates by MAC for each procedure with at least 50 total uses in my sample. 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 0.05. Next, I limit the potential coverage changes to those that represent a change of at least 33 percentage points in the denial rate and for which the difference in the level of the denial rate before and after the potential break is statistically different at the 5% level. I also add the first date a MAC pays for a procedure as a potential change in coverage status and limit potential coverage changes to be at least 6 months apart. Figure A1 shows that the coverage changes detected by this method 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 30 percentage points. By contrast, League (2023) reports that in 2017 Medicare's denial rate for medical procedures was only 10 percent. I classify

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.

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 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 Table A1 and Figure A2, the changes in coverage I detect correspond to large and immediate changes in the denial rate. Using this process, I find that 85% of MAC-procedure-month tuples correspond to non-coverage, 8% correspond to coverage on a case-by-case basis, and 6% correspond to full coverage.

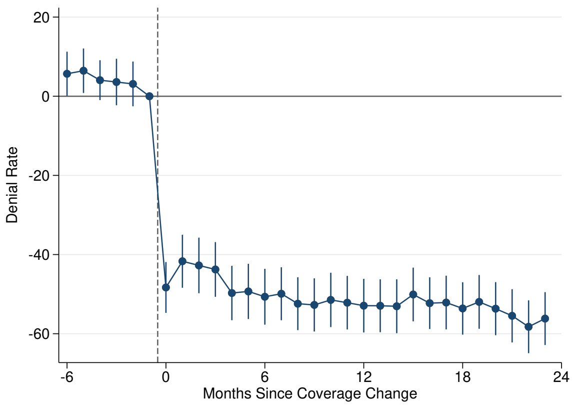
Table A1: Effect of Coverage on Denial Rate

	(1) Denial Rate	(2) Denial Rate
Case-by-Case	-45.46***	-42.74***
	(1.840)	(2.077)
Covered	-81.01***	-77.57***
	(1.264)	(1.486)
Jurisdiction FEs	1	1
Time FEs	0	1
Dep. Var. Mean	54.89	53.87
Observations	75,876	72,648

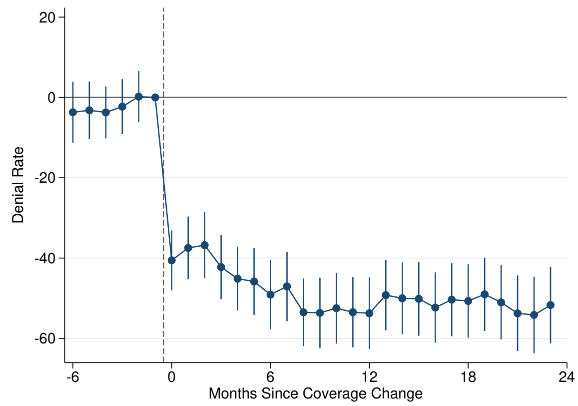
Notes: An observation is a procedure-jurisdiction-month. Sample consists of monthly observations from 2002–2017. Denial rate is the percentage of claims denied. Regressions include jurisdiction-by-procedure and procedure-by-month fixed effects where indicated. Standard errors are clustered at the jurisdiction-by-procedure level. $^{+}$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

Figure A2: Change in Denial Rate at Coverage Change

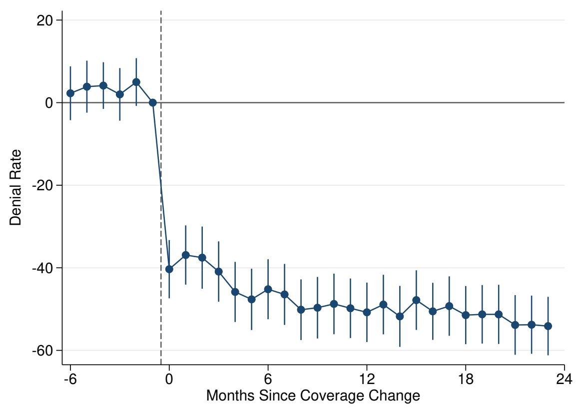
(a) No Time Fixed Effects



(b) Two-Way Fixed Effects



(c) Stacked Regression



Notes: The figures report estimates of β_e from Equation (2) for $e \in \{-6, \ldots, 24\}$. An observation is a jurisdiction-procedure-month tuple. Denial rate is the percentage of claims denied. 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.

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 as the primary measure of procedure quality 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 generally robust to using either measure, which isn't surprising given their rate of concordance, as shown in Table A2.

Next, I investigate the robustness of all the results in the main text to using this measure instead of my main one. Table A3 and Figure A3 recreate Table 4 and Figure 4 showing that coverage changes affect utilization in other jurisdictions, although here the effects on deleted procedures do not mirror the results for de-adopted ones. Table A4 and Figure A4 mirror Table 5 and Figure 5 in showing that adoption is more rapid for promoted procedures but not for deleted procedures the larger the jurisdictions in which the procedure is covered. Table A5 shows that more generous past coverage leads to more past utilization, as shown in Table 6, while Table A6 shows that additional past utilization causes more contemporary utilization for promoted procedures, replicating the result from Table 7, while the negative effect of past utilization on current use is less clear for deleted procedures, although the estimates are certainly much more negative than for promoted procedures. Finally, Figure A5 shows that as in Figure 7, the share of patients treated by a physician does not change as the provider gains more experience with a procedure. Overall, while there are some instances of results not being as strong when using this definition, I interpret these results as indicating broad agreement between the two measures.

Table A2: Concordance of Success Measures

	AMA Code Status			
Adoption Status	Deleted	Outstanding	Promoted	Total
Adopted	42 20.2%	55 26.4%	111 53.4%	208 60.8%
De-Adopted	53 39.6%	46 34.3%	35 26.1%	134 39.2%
Total	95 27.8%	101 29.5%	146 42.7%	342 100.0%

Notes: An observation is a procedure. Adoption status is whether the use of the procedure grows (adopted) or shrinks (de-adopted) over time. AMA code status is whether the procedure has been assigned a category I code (promoted), deleted from the codebook (deleted), or remains as a category III code as of January 2022. Percentages in the middle three columns of the table give the percentage of procedures in the row that fall in the cell, while percentages in the final column give the percentage out of all procedures.

Table A3: Change in Trend of Utilization

	Promoted		Deleted	
	(1) Utilization Rate	(2) Any Uses	(3) Utilization Rate	(4) Any Uses
Trend × Total Coverage	0.0411 (0.0379)	0.000198*** (0.0000360)	0.151* (0.0632)	0.0000353 (0.0000613)
Jurisdiction FEs	1	1	1	1
Time FEs	0	0	0	0
Dep. Var. Mean	21.47	0.208	9.193	0.0799
Observations	191,028	191,028	54,288	54,288

Notes: An observation is a procedure-jurisdiction-month. Sample consists of monthly observations from 2002–2017 for jurisdictions in which the procedure is never covered. Utilization rate is the number of uses of the procedure per million beneficiary-months. Total coverage is the total number of beneficiaries in millions for which the procedure is covered fully or on a case-by-case basis. Regressions include jurisdiction-by-procedure fixed effects. In columns (1) and (2), the sample is limited to promoted procedures, while in columns (3) and (4) it is limited to deleted procedures. Standard errors are clustered at the jurisdiction-by-procedure level. $^{+}$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

Table A4: Change in Trend of Utilization at Coverage Change, Heterogeneity by Size

	Promoted		Deleted	
	(1) Utilization Rate	(2) Any Uses	(3) Utilization Rate	(4) Any Uses
Trend Change × Size	0.0343	0.0000395	0.24^{*}	0.000293^{+}
	(0.0321)	(0.0000968)	(0.106)	(0.000152)
Jurisdiction FEs	1	1	1	1
Time FEs	0	0	0	0
Dep. Var. Mean	21.47	0.208	9.193	0.0799
Observations	191,028	191,028	54,288	54,288

Notes: Estimates of β_7 in Equation (6). An observation is a procedure-jurisdiction-month. Sample consists of monthly observations from 2002–2017 for procedure-jurisdiction pairs in which the procedure is ever covered. Utilization rate is the number of uses of the procedure per million beneficiary-months. Size is the number of beneficiaries in millions for which the procedure is covered fully or on a case-by-case basis. Regressions include jurisdiction-by-procedure fixed effects. In columns (1) and (2), the sample is limited to promoted procedures, while in columns (3) and (4) it is limited to deleted procedures. Standard errors are clustered at the jurisdiction-by-procedure level. $^{+}$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

Table A5: Effect of Past Coverage on Past Utilization

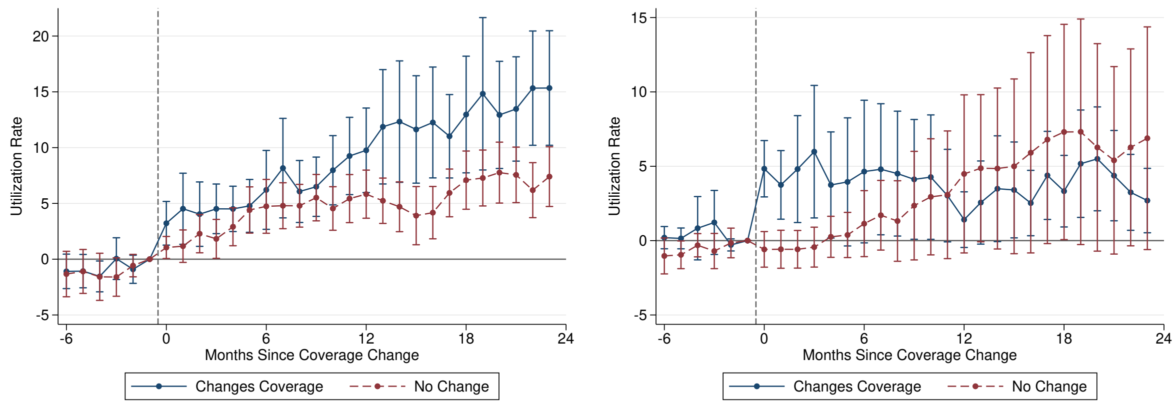
	Promoted		Deleted	
	(1) Utilization Rate	(2) Utilization Rate (Asinh)	(3) Utilization Rate	(4) Utilization Rate (Asinh)
Past Case-by-Case Coverage	74.15^{***}	0.657^{***}	-19.66^{***}	0.714^{***}
	(12.63)	(0.0109)	(2.348)	(0.0173)
Past Full Coverage	137.7^{***}	0.390^{***}	77.00^{***}	0.565^{***}
	(10.81)	(0.0102)	(6.406)	(0.0246)
Dep. Var. Mean	924.0	3.379	224.3	3.053
Observations	272,295	272,295	369,982	369,982

Notes: Ordinary least-squares estimates of α_1 and α_2 from Equation (7). An observation is a jurisdiction-procedure-month tuple in which the procedure is not covered. Utilization rate is the number of uses of the procedure per million beneficiary-months. In columns (1) and (2), the sample is limited to promoted procedures, while the sample is limited to deleted procedures in columns (3) and (4). In columns (1) and (3), past coverage is measured in millions of beneficiary-months and past uses in levels, while in columns (2) and (4), past coverage is measured as the inverse hyperbolic sine of beneficiary-months and past uses and the utilization rate are similarly transformed. All specifications include controls for the share of past MAC-months in which the procedure was covered fully, the share for which it was covered on a case-by-case basis, the total past beneficiary-months the procedure has been available, and jurisdiction-by-procedure fixed effects. Standard errors are clustered at the jurisdiction-procedure level. $^{+}$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

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

(a) Promoted Procedures

(b) Deleted Procedures

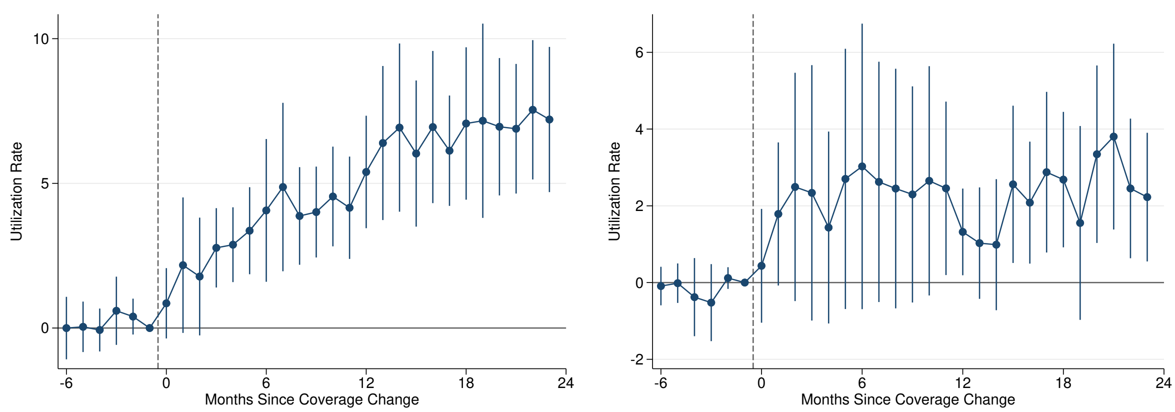


Notes: The figures report estimates of τ_e (in red) and $\beta_e + \tau_e$ (in blue) from Equation (3) for $e \in \{-6, \ldots, 24\}$. An observation is a jurisdiction-procedure-month-group tuple, where groups are defined by the stacked regression procedure defined in Appendix C. Panel (a) presents estimates for the sample limited to promoted procedures, while panel (b) presents estimates for deleted procedures. Error bars give the pointwise 95% confidence intervals. Standard errors are clustered at the jurisdiction-by-procedure level.

Figure A4: Effect of Coverage on Utilization, Heterogeneity by Size

(a) Promoted Procedures

(b) Deleted Procedures



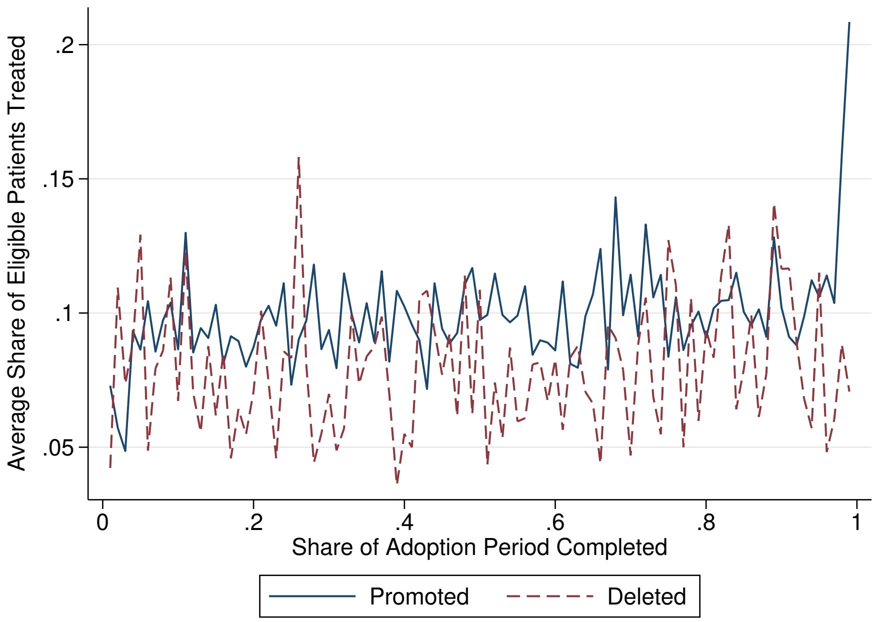
Notes: The figures report estimates of ϕ_e from Equation (5) for $e \in \{-6, \ldots, 24\}$. An observation is a procedure-jurisdiction-month. Utilization rate is the number of uses of the procedure per million beneficiary-months. Panel (a) presents estimates for the sample limited to promoted procedures, while panel (b) presents estimates for deleted procedures. Error bars give the pointwise 95% confidence intervals. Standard errors are clustered at the jurisdiction-by-procedure level.

Table A6: Effect of Past Utilization on Current Utilization

	Promoted		Deleted	
	(1)	(2)	(3)	(4)
	Utilization Rate	Utilization Rate (Asinh)	Utilization Rate	Utilization Rate (Asinh)
Past Uses	0.000532***	0.0769***	-0.00000764	0.0194**
	(0.0000856)	(0.00858)	(0.0000851)	(0.00712)
First-Stage F Stat.	81.52	3649.6	72.64	1708.9
Dep. Var. Mean	3.072	0.221	0.375	0.0460
Observations	272,295	272,295	369,982	369,982

Notes: Two-stage least-squares estimates of β_1 and from Equation (8). An observation is a jurisdiction-procedure-month tuple in which the procedure is not covered. Utilization rate is the number of uses of the procedure per million beneficiary-months. In columns (1) and (2), the sample is limited to promoted procedures, while the sample is limited to deleted procedures in columns (3) and (4). In columns (1) and (3), past coverage is measured in millions of beneficiary-months and past uses in levels, while in columns (2) and (4), past coverage is measured as the inverse hyperbolic sine of beneficiary-months and past uses and the utilization rate are similarly transformed. All specifications include controls for the share of past MAC-months in which the procedure was covered fully, the share for which it was covered on a case-by-case basis, the total past beneficiary-months the procedure has been available, and jurisdiction-by-procedure fixed effects. Standard errors are clustered at the jurisdiction-procedure level. $^{+}$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

Figure A5: Within-Provider Change in Use



Notes: Figure presents the share of patients in the consideration set treated using the new procedure among providers that have previously used the procedure and will use the procedure again in the future. The time from the provider's first to last use of the procedure is normalized to one. The consideration set is all patients with a primary diagnosis shared by any patient ever treated with the procedure who are treated by a physician whose specialty is that of any physician who ever uses the procedure. The horizontal axis gives the share of months between the provider's first and last use of the procedure. The solid blue line is limited to promoted procedures, while the dashed red line is for deleted procedures.

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. The stacked regression method, from Cengiz et al. (2019), works by constructing appropriate controls groups for each transition between different coverage levels 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 estimate

$$(12) \qquad Y_{pjtw} = \sum_{e=-K}^{-2} \beta_e T_{pjtw}(e) + \sum_{e=0}^{L} \beta_e T_{pjtw}(e) + \alpha_{pjw} + \alpha_{ptw} + \varepsilon_{pjtw},$$

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 the stacked dataset. Similarly, I can aggregate the pre- and post-coverage change periods in this stacked dataset to estimate the average treatment effect of a coverage change over the L months following the change:

$$(13) \qquad Y_{pjtw} = \beta \sum_{e=0}^{L} T_{pjtw}(e) + \alpha_{pjw} + \alpha_{ptw} + \varepsilon_{pjtw},$$

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 over the L month treatment window. To this, I use the non-stacked data set to estimate

$$(14) \qquad Y_{pjt} = \beta \sum_{e=0}^{L} T_{pjt}(e) + \gamma_{pj} + \gamma_{pt} + \varepsilon_{pjt}.$$

Table A7: 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**	3.320*	0.0728***	0.0512***	0.0486***
	(1.355)	(1.637)	(1.540)	(0.00451)	(0.00457)	(0.00458)
Jurisdiction-Procedure FEs	1	1	1	1	1	1
Procedure-Month FEs	0	1	1	0	1	1
Stacked?	0	0	1	0	0	1
Dep. Var. Mean	9.795	2.708	12.54	0.135	0.0473	0.139
Observations	58,454	1,003,890	619,500	58,454	1,003,890	619,500

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. Columns (1) and (4) report estimates of β from Equation (14) with the sample limited to a balanced panel of jurisdiction-months of the 6 months before coverage is extended to the procedure in the jurisdiction and 24 months after. Columns (2) and (5) report estimates of β from Equation (14) with the sample limited to the same balanced panel of treated jurisdiction-procedure pairs plus control jurisdiction-procedure pairs that are not subject to treatment. Columns (3) and (6) report estimates of β from Equation (13). Standard errors are clustered at the jurisdiction-by-procedure level. $^{+}$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

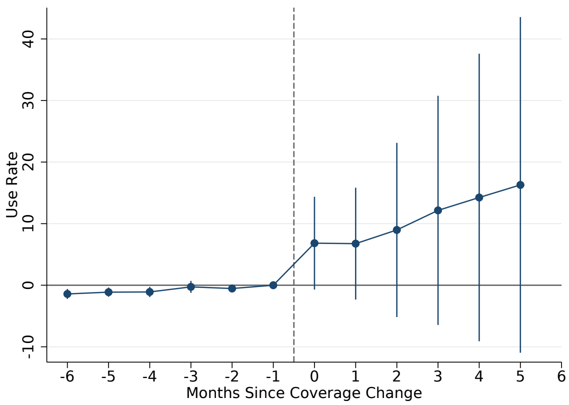
Notice that for both Equations (13) and (14), the β gives the differential change for the L months after coverage changes in the treated jurisdictions relative to the K months before, relative to the appropriate control jurisdictions.

In Table A7, 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 (13) I restrict each treatment-control wave w to be a balanced panel and to estimate Equation (14) 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.

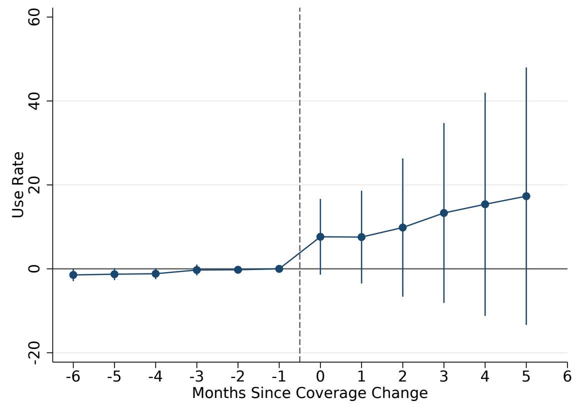
I use $K = -6$ in the main text because many coverage changes occur relatively shortly after the introduction of a code and a longer pre-period requires limiting the sample. By contrast, using a longer post-period of $L = 24$ allows for the assessment of dynamic treatment effects over a longer time horizon. Figures A6, A7, A8, and A9 present estimates of Equation (2) with and without time-by-procedure fixed effects and Equation (12) using treatment windows of $L = K = 6$, $L = K = 12$, $L = K = 18$, and $L = K = 24$, respectively.

Figure A6: Change in Utilization at Coverage Change

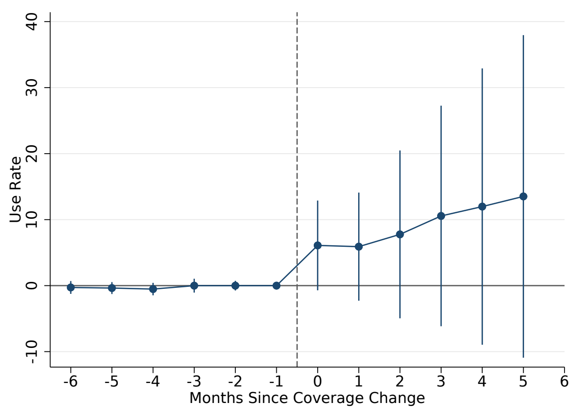
(a) No Time Fixed Effects



(b) Two-Way Fixed Effects



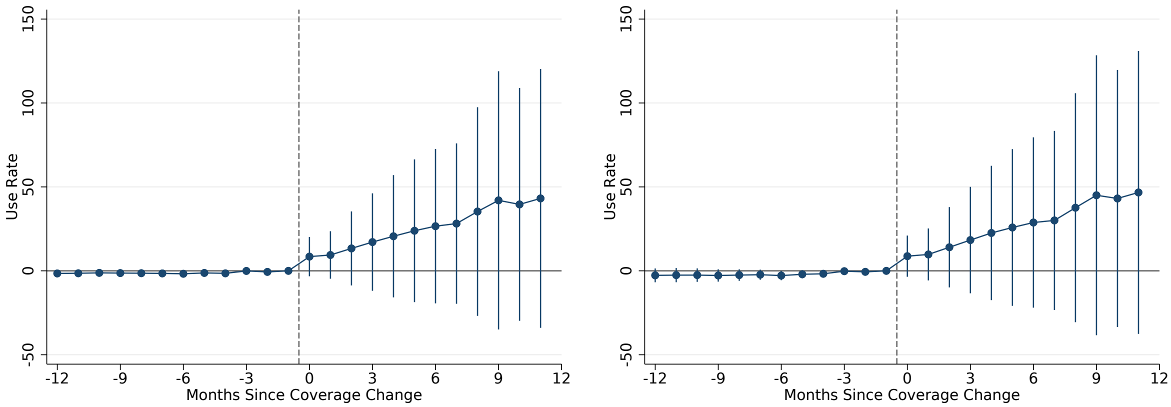
(c) Stacked Regression



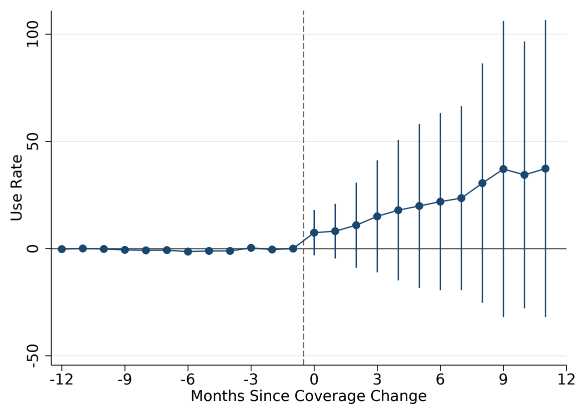
Notes: The figures report estimates of β_e from Equation (2) and (12) for $e \in \{-6, \ldots, 6\}$. An observation is a jurisdiction-procedure-month tuple. Panel (a) includes estimates without time fixed effects. Panel (b) uses the two-way fixed effects estimator. Panel (c) uses the stacked regression 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 A7: Change in Utilization at Coverage Change

(a) No Time Fixed Effects (b) Two-Way Fixed Effects



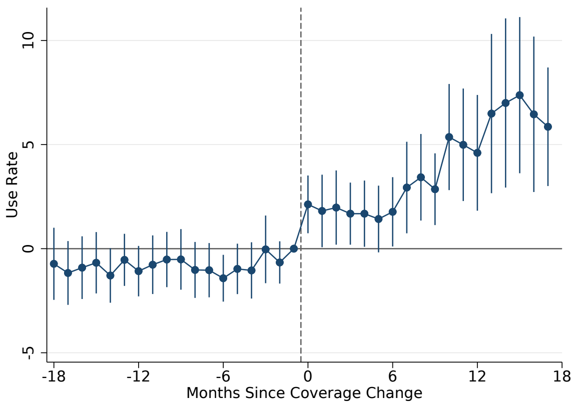
(c) Stacked Regression



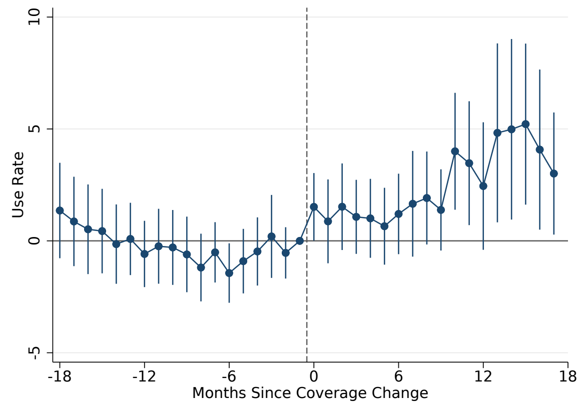
Notes: The figures report estimates of β_e from Equation (2) and (12) for $e \in \{-12, \ldots, 12\}$. An observation is a jurisdiction-procedure-month tuple. Panel (a) includes estimates without time fixed effects. Panel (b) uses the two-way fixed effects estimator. Panel (c) uses the stacked regression 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 A8: Change in Utilization at Coverage Change

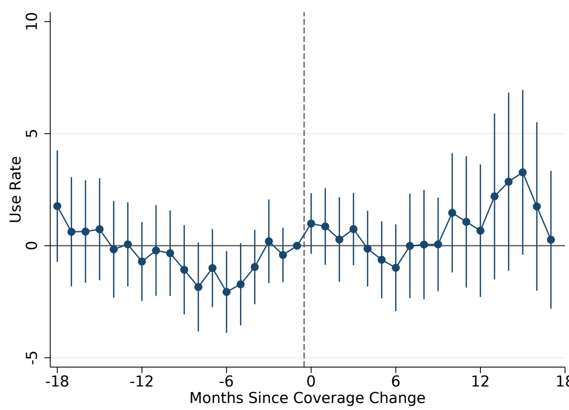
(a) No Time Fixed Effects



(b) Two-Way Fixed Effects



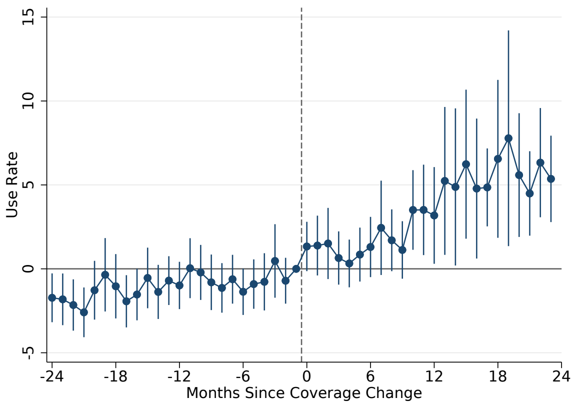
(c) Stacked Regression



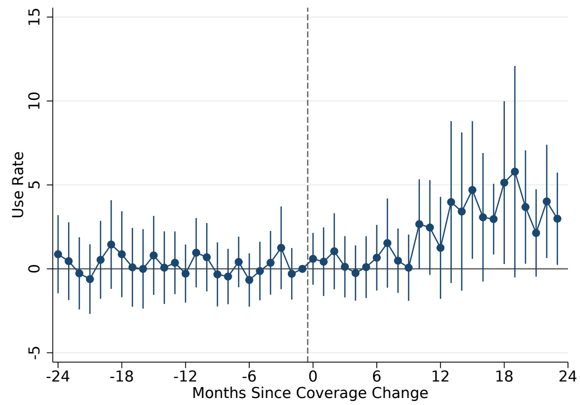
Notes: The figures report estimates of β_e from Equation (2) and (12) for $e \in \{-18, \ldots, 18\}$. An observation is a jurisdiction-procedure-month tuple. Panel (a) includes estimates without time fixed effects. Panel (b) uses the two-way fixed effects estimator. Panel (c) uses the stacked regression 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 A9: Change in Utilization at Coverage Change

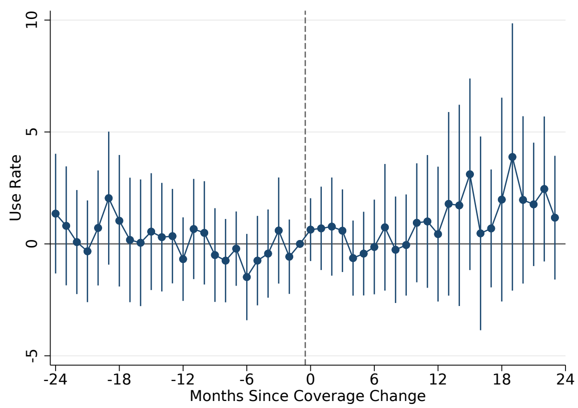
(a) No Time Fixed Effects



(b) Two-Way Fixed Effects



(c) Stacked Regression



Notes: The figures report estimates of β_e from Equation (2) and (12) 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 Local Spillovers

In this appendix, I investigate the geography of knowledge spillovers. In particular, I assess whether there are larger spillovers to nearby geographies than to the medical community writ large. To do this, I estimate whether utilization responds to coverage decisions in nearby, bordering jurisdictions. To do this, I augment Equation (1) by including an indicator for bordering a jurisdiction in which the procedure is covered. These coefficients give the differential change in utilization in bordering jurisdictions after the focal jurisdiction changes coverage of the procedure relative to the change in utilization elsewhere. In other words, they capture the spillovers to nearby jurisdictions that are not shared nationally. Table A8 shows that utilization in bordering jurisdictions responds no more than in other, farther-away jurisdictions, either overall or by procedure adoption status. These results indicate that any knowledge spillovers from utilization of the procedure accrue to the entire medical community rather than being geographically limited.

Table A8: Effect of Coverage on Utilization, Local Spillovers

	All		Adopted		De-Adopted	
	(1) Utilization Rate	(2) Any Uses	(3) Utilization Rate	(4) Any Uses	(5) Utilization Rate	(6) Any Uses
Case-by-Case	5.071	0.0372***	6.881	0.0360***	1.794*	0.0394***
	(3.330)	(0.00347)	(5.127)	(0.00420)	(0.894)	(0.00615)
Covered	19.03	0.0572***	26.52	0.0537***	4.531***	0.0641***
	(14.17)	(0.00544)	(21.46)	(0.00678)	(1.295)	(0.00908)
Borders Case-by-Case	-3.078	-0.000688	-4.620	0.000324	0.129	-0.00262
	(4.023)	(0.00286)	(6.062)	(0.00360)	(0.473)	(0.00471)
Borders Covered	-1.544	-0.00235	-2.078	-0.000181	-0.0180	-0.00635
	(3.172)	(0.00430)	(4.709)	(0.00568)	(0.684)	(0.00638)
Jurisdiction-Procedure FEs	1	1	1	1	1	1
Procedure-Month FEs	1	1	1	1	1	1
Dep. Var. Mean	4.870	0.0623	7.397	0.0857	1.653	0.0326
Observations	1,240,092	1,240,092	694,602	694,602	545,490	545,490

Notes: An observation is a procedure-jurisdiction-month. Sample consists of monthly observations from 2002–2017. Utilization rate is the number of uses of the procedure per million beneficiary-months. Regressions include jurisdiction-by-procedure and procedure-by-month fixed effects. In columns (3) and (4), the sample is limited to adopted procedures, while in columns (5) and (6) it is limited to de-adopted procedures. Standard errors are clustered at the jurisdiction-by-procedure level. $^{+}$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

E Investigating the Exclusion Restriction

In this appendix, I investigate additional threats to the exclusion restriction beyond those discussed in Section 5.3. The first threat to identification I consider is persistance in coverage rules. The worry here is that past coverage rules may be correlated with current coverage rules such that past coverage predicts utilization later not because of past utilization but because of contemporary coverage. In the main text, I avoid this possibility by limiting the sample to jurisdictions in which the procedure is not contemporaneously covered, such that contemporary coverage cannot have an effect on utilization. Alternatively, I could have included all jurisdictions and controlled for contemporary coverage directly. Doing this, I find in Table A9 even stronger results than those reported in the main text in Table 7. The results indicate that by including jurisdictions in which the procedure is covered, the magnitudes of the effect of past coverage are larger. Furthermore, by controlling for contemporary coverage, I am shutting down the channel of past utilization affecting current coverage rules (because of additional information about the procedure). In other words, controlling for contemporary coverage rules (or controlling for them as I do here) is conservative because one channel through which past utilization affects current utilization is through its effect on administrators revising their coverage rules in light of the additional information created by the utilization.

The next concern I address is correlation between the instrument and time. That is, utilization may grow or fall over time due to changing average value of the mix of procedures over time or due to different trends in utilization between adopted and de-adopted procedures. If the generosity of coverage rises or falls over time, then this would induce correlation between the error term and the instrument. I address this in the main text by controlling for the cumulative beneficiary-months the procedure has been available and including procedure-jurisdiction fixed effects to isolate within-procedure changes in utilization. Here, I show that the results are robust to these choices. Table A10 replicates Table 7 dropping the control for cumulative beneficiary-months. We see a stronger first stage and similar estimates. Similarly, adding a linear control for time weakens the first stage and makes the estimates somewhat less precise but results in similar estimates, as shown in Table A11. Finally, dropping the procedure-jurisdiction fixed effect results in noisier estimates that demonstrate a similar pattern, as shown by Table A12.

More generally, the heterogeneity in the effect of past utilization by whether the procedure is eventually found to be effective means that a potential exclusion restriction violation would have to have a very specific and unlikely form: it would have to be the case that the number of beneficiaries covered in the past is *positively* correlated with later utilization for adopted procedures but *negatively* correlated for de-adopted ones. Even the concerns discussed here would generally not be able to produce this heterogeneity.

Table A9: Effect of Past Utilization on Current Utilization, All Coverage Levels

	Adopted		De-Adopted	
	(1) Utilization Rate	(2) Utilization Rate (Asinh)	(3) Utilization Rate	(4) Utilization Rate (Asinh)
Past Uses	0.00277***	0.0384***	-0.000693**	-0.0502***
	(0.000305)	(0.00799)	(0.000269)	(0.0103)
First-Stage F Stat.	789.8	6138.7	561.7	4072.6
Dep. Var. Mean	7.397	0.356	1.653	0.118
Observations	694,602	694,602	545,490	545,490

Notes: Two-stage least-squares estimates of β_1 and from Equation (8). An observation is a jurisdiction-procedure-month tuple. Utilization rate is the number of uses of the procedure per million beneficiary-months. In columns (1) and (2), the sample is limited to adopted procedures, while the opposite is true for columns (3) and (4). In columns (1) and (3), past coverage is measured in millions of beneficiary-months and past uses in levels, while in columns (2) and (4), past coverage is measured as the inverse hyperbolic sine of beneficiary-months and past uses and the utilization rate are similarly transformed. All specifications include controls for the share of past MAC-months in which the procedure was covered fully, the share for which it was covered on a case-by-case basis, the total past beneficiary-months the procedure has been available, jurisdiction-by-procedure fixed effects, and indicators for being contemporaneously covered fully or on a case-by-case basis. Standard errors are clustered at the jurisdiction-procedure level. $^+$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

Table A10: Effect of Past Utilization on Current Utilization, No Control for Population

	Adopted		De-Adopted	
	(1) Utilization Rate	(2) Utilization Rate (Asinh)	(3) Utilization Rate	(4) Utilization Rate (Asinh)
Past Uses	0.000983***	0.106***	-0.000627+	-0.0190***
	(0.000118)	(0.00823)	(0.000360)	(0.00425)
First-Stage F Stat.	117.0	4931.7	110.0	8213.6
Dep. Var. Mean	2.071	0.172	0.636	0.0611
Observations	569,006	569,006	472,945	472,945

Notes: Two-stage least-squares estimates of β_1 and from Equation (8). An observation is a jurisdiction-procedure-month tuple in which the procedure is not covered. Utilization rate is the number of uses of the procedure per million beneficiary-months. In columns (1) and (2), the sample is limited to adopted procedures, while the opposite is true for columns (3) and (4). In columns (1) and (3), past coverage is measured in millions of beneficiary-months and past uses in levels, while in columns (2) and (4), past coverage is measured as the inverse hyperbolic sine of beneficiary-months and past uses and the utilization rate are similarly transformed. All specifications include controls for the share of past MAC-months in which the procedure was covered fully, the share for which it was covered on a case-by-case basis, and jurisdiction-by-procedure fixed effects. Standard errors are clustered at the jurisdiction-procedure level. $^+$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

Table A11: Effect of Past Utilization on Current Utilization, Linear Time Control

	Adopted		De-Adopted	
	(1) Utilization Rate	(2) Utilization Rate (Asinh)	(3) Utilization Rate	(4) Utilization Rate (Asinh)
Past Uses	0.000840***	0.113**	-0.000574	-0.0273**
	(0.000104)	(0.000965)	(0.000401)	(0.00875)
First-Stage F Stat.	104.0	1001.3	37.68	1039.6
Dep. Var. Mean	2.071	0.172	0.636	0.0611
Observations	569,006	569,006	472,945	472,945

Notes: Two-stage least-squares estimates of β_1 and from Equation (8). An observation is a jurisdiction-procedure-month tuple in which the procedure is not covered. Utilization rate is the number of uses of the procedure per million beneficiary-months. In columns (1) and (2), the sample is limited to adopted procedures, while the opposite is true for columns (3) and (4). In columns (1) and (3), past coverage is measured in millions of beneficiary-months and past uses in levels, while in columns (2) and (4), past coverage is measured as the inverse hyperbolic sine of beneficiary-months and past uses and the utilization rate are similarly transformed. All specifications include controls for the share of past MAC-months in which the procedure was covered fully, the share for which it was covered on a case-by-case basis, the total past beneficiary-months the procedure has been available, a continuous measure of monthly time, and jurisdiction-by-procedure fixed effects. Standard errors are clustered at the jurisdiction-procedure level. $^+$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

Table A12: Effect of Past Utilization on Current Utilization, No Fixed Effects

	Adopted		De-Adopted	
	(1) Utilization Rate	(2) Utilization Rate (Asinh)	(3) Utilization Rate	(4) Utilization Rate (Asinh)
Past Uses	0.00136***	0.0170	0.000335	-0.00896
	(0.000143)	(0.0144)	(0.000204)	(0.0124)
First-Stage F Stat.	78.88	888.4	48.73	234.0
Dep. Var. Mean	2.071	0.172	0.636	0.0612
Observations	569,229	569,229	472,978	472,978

Notes: Two-stage least-squares estimates of β_1 and from Equation (8). An observation is a jurisdiction-procedure-month tuple in which the procedure is not covered. Utilization rate is the number of uses of the procedure per million beneficiary-months. In columns (1) and (2), the sample is limited to adopted procedures, while the opposite is true for columns (3) and (4). In columns (1) and (3), past coverage is measured in millions of beneficiary-months and past uses in levels, while in columns (2) and (4), past coverage is measured as the inverse hyperbolic sine of beneficiary-months and past uses and the utilization rate are similarly transformed. All specifications include controls for the share of past MAC-months in which the procedure was covered fully, the share for which it was covered on a case-by-case basis, and the total past beneficiary-months the procedure has been available. Standard errors are clustered at the jurisdiction-procedure level. $^+$, *, ** and *** indicate significance at the 10%, 5%, 1% and 0.1% level, respectively.

F Model Details

In this appendix, I provide mathematical details on the model presented in Section 6, including derivation of the likelihood function and proofs of model identification and claims about patient welfare.

F.1 Derivation of Likelihood Function

In this subsection, 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_{pt} subsequent signals about procedure p received by provider i by time t, ν_{ipk}, are drawn from the distribution $\mathcal{N}(\delta_p^*, \sigma_{\nu p})$ for $k \in \{1, \ldots, n_{pt}\}$ independently of the signal received by other physicians. Using Bayes rule, this means that physician i's beliefs after receiving n_{pt} signals $\{\nu_{ipk}\}_{k \in \{1,\ldots,n_{pt}\}}$, are distributed $\mathcal{N}(\mu_{ipt}, \sigma_{ipt})$, where

$$\sigma_{ipt} = \frac{\sigma_{\delta p}\sigma_{\nu p}}{\sqrt{\sigma_{\nu p}^2 + n_{pt}\sigma_{\delta p}^2}} \qquad \text{and} \qquad \mu_{ipt} = (\sigma_{ipt})^2 \left(\frac{\delta_{0ip}}{\sigma_{\delta p}^2} + \frac{n_{pt}\delta_p^*}{\sigma_{\nu p}^2} \right).$$

Because the only stochastic term in the formula for μ_{ipt} is δ_{0ip}, which is normally distributed, physician's mean beliefs μ_{ipt} will be distributed $\mathcal{N}(\mu_{agg,pn}, \sigma_{agg,pn})$, where

$$\sigma_{agg,pt} = \sqrt{\left(\frac{\sigma_{\delta p}\sigma_{\nu p}}{\sqrt{\sigma_{\nu p}^2 + n_{pt}\sigma_{\delta p}^2}} \right)^2 \cdot \frac{\sigma_{\delta p}^2}{\sigma_{\delta p}^2}} = \frac{\sigma_{\delta p}\sigma_{\nu p}}{\sqrt{\sigma_{\nu p}^2 + n_{pt}\sigma_{\delta p}^2}}$$

and

$$\mu_{agg,pt} = \left(\frac{\sigma_{\delta p}\sigma_{\nu p}}{\sqrt{\sigma_{\nu p}^2 + n_{pt}\sigma_{\delta p}^2}} \right)^2 \left(\frac{\delta_{0p}}{\sigma_{\delta p}^2} + \frac{n_{pt}\delta_p^*}{\sigma_{\nu p}^2} \right) = (\sigma_{agg,pt})^2 \left(\frac{\delta_{0p}}{\sigma_{\delta p}^2} + \frac{n_{pt}\delta_p^*}{\sigma_{\nu p}^2} \right).$$

Notice that the variance of the beliefs of any physician i is equal to the variance of mean beliefs across physicians at any point in time.

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

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

Thus we have that

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

Notice that $\mu_{agg,pt}$ 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.[1] 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 $B(g_{ipt}, \alpha_{cov,pt}\eta_p)$, 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 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).[2] 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 $B(g_{cov,pt}, \alpha_{cov,pt}\eta_p)$, 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

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

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

[2] In the data, I find no evidence that the number of patients in a physician's consideration set is positively correlated with the share of eligible patients or, equivalently, the share treated. In particular, I estimate the provider-level correlation between the share of patients treated with the new procedure and the number of patients in the consideration set among providers that have adopted the procedure is -0.08. If providers better at performing the procedure attract additional patients seeking out the procedure, this correlation would be positive, so the very weak, even negative correlation I find lends support to my assumption that η_p and X_{ip} are independent.

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

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

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

F.2 Estimation of η_p

I do not estimate η_p jointly with the other parameters of the physician learning model because this would be identified only off of functional form assumptions. Instead, I separately estimate the share of patients for which the new procedure is appropriate by looking at the share of patients treated among physicians that have adopted the procedure. Under the model assumption that providers that have adopted the procedure treat all patients for which the new procedure is appropriate, this share identifies η_p.

The difficulty I must deal with is that I do not directly observe providers' adoption decisions. 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. To overcome this difficulty, I limit the estimation of η_p to physicians who have used the new procedure at least once in the past and will use it at least once in the future. In so doing, I limit my attention to physicians that I know have adopted (and not de-adopted) the procedure. To obtain an unbiased estimate of η_p, I estimate the share of patients in the consideration set that are treated in the months strictly between a physician's first and last utilization.[3]

Because of the change in diagnosis reporting with the switch from ICD-9 to ICD-10 in 2015, I estimate η_p separately under these two reporting regimes.

F.3 Proof of Model Identification

In this subsection, I prove that the model presented in Section 6 is identified. First, note that with η_p separately identified as explained in the previous subsection, $\alpha_{cov,pt}$ is identified by the

[3]Formally, this identifies η_p because $y_{cov,pt} = \alpha_{cov,pt}\eta_p g_{cov,pt}$ and I restrict my attention to cases where $\alpha_{cov,pt}$ is equal to one, so $\eta_p = \frac{y_{cov,pt}}{g_{cov,pt}}$.

share of patients in the consideration set that are treated with the new procedure:

$$\alpha_{cov,pt} = \frac{y_{cov,pt}}{g_{cov,pt}\eta_p}.$$

With $\alpha_{cov,pt}$ in hand, then $\theta_{cov,pt} \equiv \frac{\mu_{agg,pt}-\beta_{1p}Case_{ipt}-\beta_{2p}Noncov_{ipt}}{\sqrt{(\sigma_{agg,pt})^2+1}}$ is identified by

$$\theta_{cov,pt} = \Phi^{-1}\left(\alpha_{cov,pt}\right).$$

I will next write the parameters of the physician learning model in terms of $\theta_{cov,pt}$.

First, the true procedure value δ_p^* and the costs of coverage restrictions β_{1p} and β_{2p} are identified by projecting the trend in utilization to the limit. This is because we have $\lim_{n_{pt}\to\infty}\sigma_{agg,pt} = 0$ and

$$\lim_{n_{pt}\to\infty}\mu_{agg,pt} = \frac{\lim_{n_{pt}\to\infty}\delta_p^*\sigma_{\delta p}^2}{\lim_{n_{pt}\to\infty}\sigma_{\delta p}^2} = \delta_p^*,$$

where the first equality applies L'Hôpital's rule. This implies

$$\lim_{n_{pt}\to\infty}\theta_{cov,pt} = \frac{\delta_p^* - \beta_{1p}Case_{ipt} - \beta_{2p}Noncov_{ipt}}{\sqrt{(0)^2+1}} = \delta_p^* - \beta_{1p}Case_{ipt} - \beta_{2p}Noncov_{ipt}.$$

Thus, we have $\delta^* = \lim_{n_{pt}\to\infty}\theta_{1,pt}$, $\beta_{1p} = \lim_{n_{pt}\to\infty}\theta_{1,pt} - \lim_{n_{pt}\to\infty}\theta_{0,pt}$, and $\beta_{2p} = \lim_{n_{pt}\to\infty}\theta_{1,pt} - \lim_{n_{pt}\to\infty}\theta_{-1,pt}$, which gives each of these parameters in terms of data.

With these parameters in hand, the mean δ_{0p} and standard deviation $\sigma_{\delta p}$ of physicians' priors are identified by the level of utilization in the first period, along with the differences by coverage level. To see this, note that at $n_{pt} = 0$ (which occurs at $t = 1$), we have $\sigma_{agg,p1} = \sigma_{\delta p}$ and $\mu_{agg,p1} = \delta_{0p}$, so

$$\theta_{cov,p1} = \frac{\delta_{0p} - \beta_{1p}Case_{ipt} - \beta_{2p}Noncov_{ipt}}{\sqrt{(\sigma_{\delta p})^2+1}}. \tag{15}$$

Comparing this value under different coverage regimes, we have

$$\theta_{1,p1} - \theta_{0,p1} = \frac{\beta_{1p}}{\sqrt{(\sigma_{\delta p})^2+1}},$$

or

$$\sigma_{\delta p} = \sqrt{\left(\frac{\beta_{1p}}{\theta_{1,p1}-\theta_{0,p1}}\right)^2 - 1},$$

which because β_{1p} is identified, means that $\sigma_{\delta p}$ is as well. Because the same comparison can be done for $cov = -1$, this parameter is over-identified. Plugging this back into Equation (15), we have

$$\theta_{cov,p1} = \frac{\delta_{0p} - \beta_{1p} Case_{ipt} - \beta_{2p} Noncov_{ipt}}{\frac{\beta_{1p}}{\theta_{1,p1} - \theta_{0,p1}}},$$

or

$$\delta_{0p} = \frac{\beta_{1p}\theta_{cov,p1}}{\theta_{1,p1} - \theta_{0,p1}} + \beta_{1p} Case_{ipt} + \beta_{2p} Noncov_{ipt},$$

which because β_{1p} and β_{2p} are identified, is (over-)identified.

This leaves only the standard deviation of the signals $\sigma_{\nu p}$ left to be identified. This parameter is identified by the response of adoption to additional signals under each coverage regime. This can be seen by noting that $\theta_{cov,pt}$ is a function of $\mu_{agg,pt}$ and $\sigma_{agg,pt}$, both of which depend only on previously identified parameters, n_{pt}, and $\sigma^2_{\nu p}$. Therefore, all parameters of the structural model of physician learning are identified.

F.4 Proofs of Claims About Patient Welfare

In this subsection, I prove various propositions about welfare in the model presented in Section 6. First, I will derive patient welfare per use of the new procedure $\mathbb{E}[W_{iptj}|U_{ipt} > 0, \eta_{pj} = 1]$ in terms of model parameters. First, I substitute out W_{iptj} and U_{ipt} and make use of the independence of X_{ip} and U_{ipt} to get

$$\mathbb{E}[W_{iptj}|U_{ipt} > 0, \eta_{pj} = 1] = \mathbb{E}[\delta^*_p - X_{ip}|\mathbb{E}_{ipt}[\delta^*_p] - \beta_{1p} Case_{ipt} - \beta_{2p} Noncov_{ipt} - X_{ip} > 0].$$

Next, I define $Z_{ipt} \equiv X_{ip} - \mathbb{E}_{ipt}[\delta^*_p] + \mu_{agg,pt}$ and make use of the fact that δ^*_p is a constant to rewrite this as

$$\mathbb{E}[W_{iptj}|U_{ipt} > 0, \eta_{pj} = 1] = \delta^*_p - \mathbb{E}[X_{ip}|Z_{ipt} < \mu_{agg,pt} - \beta_{1p} Case_{ipt} - \beta_{2p} Noncov_{ipt}].$$

Because of the distributions of $\mathbb{E}_{ipt}[\delta^*_p]$ ($\mathbb{E}_{ipt}[\delta^*_p] \sim \mathcal{N}(\mu_{agg,pt}, \sigma^2_{agg,pt})$) and X_{ip} ($X_{ip} \sim \mathcal{N}(0,1)$), the joint distribution of X_{ip} and Z_{ipt} is given by $(X_{ip}, Z_{ipt}) \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & \sigma^2_{agg,pt} + 1 \end{bmatrix}\right)$. By the properties of the multivariate normal distribution, the conditional mean of X_{ip} is $\mathbb{E}[X_{ip}|Z_{ipt}] = \frac{Z_{ipt}}{\sigma^2_{agg,pt}+1}$. Thus, using the law of iterated expectations, welfare per use of the new procedure can be written

$$\begin{aligned}
\mathbb{E}[W_{iptj}|U_{ipt} > 0, \eta_{pj} = 1] &= \delta^*_p - \mathbb{E}[\mathbb{E}[X_{ip}|Z_{ipt}]|Z_{ipt} < \mu_{agg,pt} - \beta_{1p} Case_{ipt} - \beta_{2p} Noncov_{ipt}] \\
&= \delta^*_p - \frac{1}{\sigma^2_{agg,pt} + 1}\mathbb{E}\left[Z_{ipt}|Z_{ipt} < \mu_{agg,pt} - \beta_{1p} Case_{ipt} - \beta_{2p} Noncov_{ipt}\right] \\
&= \delta^*_p - \frac{1}{\sigma^2_{agg,pt} + 1}\left(-\frac{\phi(\theta_{cov,pt})}{\Phi(\theta_{cov,pt})} \times \sqrt{\sigma^2_{agg,pt} + 1}\right)
\end{aligned}$$

where $\theta_{cov,pt} \equiv \frac{\mu_{agg,pt}-\beta_{1p}Case_{ipt}-\beta_{2p}Noncov_{ipt}}{\sqrt{\sigma^2_{agg,pt}+1}}$ and the last equality follows by the properties of the trucated normal distribution. Simplifying, average welfare per use of procedure p is given by

$$\mathbb{E}[W_{iptj}|U_{ipt}>0,\eta_{pj}=1]=\delta^*_p+\frac{1}{\sqrt{\sigma^2_{agg,pt}+1}}\left(\frac{\phi(\theta_{cov,pt})}{\Phi(\theta_{cov,pt})}\right). \tag{16}$$

Total welfare is given by

$$\begin{aligned}\mathcal{W}_{cov,pt} &= g_{cov,pt}\eta_p\Phi(\theta_{cov,pt})\mathbb{E}[W_{iptj}|U_{ipt}>0,\eta_{pj}=1]\\ &= g_{cov,pt}\eta_p\Phi(\theta_{cov,pt})\left(\delta^*_p+\frac{1}{\sqrt{\sigma^2_{agg,pt}+1}}\left(\frac{\phi(\theta_{cov,pt})}{\Phi(\theta_{cov,pt})}\right)\right)\\ &= g_{cov,pt}\eta_p\left(\delta^*_p\Phi(\theta_{cov,pt})+\frac{\phi(\theta_{cov,pt})}{\sqrt{\sigma^2_{agg,pt}+1}}\right),\end{aligned}$$

meaning the static welfare-maximizing coverage policy is to choose the level of procedure coverage among the three options (full coverage, case-by-case coverage, or non-coverage) that maximizes the value of this expression.

With welfare per use of the new procedure written in terms of model parameters, I next show a few interesting properties of this object. First, welfare per use is strictly greater than the true average value of the procedure δ^*_p. This is because the physicians that adopt the procedure will be those that believe they are well-suited to perform the procedure, a group with higher average ability than the entire physician population. To see this mathematically, note that in Equation (16), $\sigma^2_{agg,pt}$ is strictly positive and $\phi(\cdot)$ and $\Phi(\cdot)$ are everywhere strictly positive, so $\mathbb{E}[W_{iptj}|U_{ipt}>0,\eta_{pj}=1]>\delta^*_p$.

Next, I show that with no uncertainty around the procedure's value, the welfare from each use of the procedure is strictly positive under full coverage. That is, I show that $\mu_{agg,pt}=0$, $Case_{ipt}=0$, $Noncov_{ipt}=0$ implies that $\lim_{\sigma_{agg,pt}\to 0}\mathbb{E}[W_{iptj}|U_{ipt}>0,\eta_{pj}=1]>0$,[4] or rewriting this in terms of model parameters, $\delta^*_p+\frac{\phi(\delta^*_p)}{\Phi(\delta^*_p)}>0$. To see that this is true, note that the function $g(x)=\frac{x^2}{2}+\log\Phi(x)$ is strictly increasing in x, with the second term being a strictly increasing transformation of a strictly increasing function of x, meaning $g'(x)=x+\frac{\phi(x)}{\Phi(x)}>0$ for all x.

Next, I show that welfare per use is decreasing in belief about the efficacy of the procedure. The intuition is that as provider beliefs become more negative, the providers that nevertheless adopt the procedure will be more positively selected. Formally, I show that $\frac{\partial\mathbb{E}[W_{iptj}|U_{ipt}>0,\eta_{pj}=1]}{\partial\mu_{agg,pt}}<$

[4] Rather than $Case_{ipt}=0$ and $Noncov_{ipt}=0$, $\beta_{1p}\geq 0$ and $\beta_{2p}\geq 0$ would also deliver the same result.

0. To do this, I define $g(x) \equiv \frac{\phi(x)}{\Phi(x)}$ and note that the derivative of $g(x)$ is given by

$$g'(x) = \frac{\Phi(x)(-x\phi(x)) - \phi(x)\phi(x)}{\Phi(x)\Phi(x)} = -\frac{\phi(x)}{\Phi(x)^2}\left(x\Phi(x) + \phi(x)\right),$$

which is always negative because

$$xF(x) + f(x) = x\int_{-\infty}^{x} f(t)dt + \int_{-\infty}^{x} -tf(t)dt = \int_{-\infty}^{x} (x-t)f(t)dt \geq 0,$$

where $f(x)$ is the derivative of $F(x)$ with respect to x, with the inequality coming from t being less than or equal to x, with the inequality being strict when $f(t) > 0$ for some $t < x$, which it is in the case of $\phi(x)$ for all x. Thus the derivative of welfare per use with respect to beliefs is given by

$$\frac{\partial \mathbb{E}[W_{iptj}|U_{ipt} > 0, \eta_{pj} = 1]}{\partial \mu_{agg,pt}} = \frac{g'\left(\theta_{cov,pt}\right)}{\sigma_{agg,pt}^2 + 1} < 0.$$

Finally, I show that welfare per use is decreasing in uncertainty about the efficacy of the procedure, or $\frac{\partial \mathbb{E}[W_{iptj}|U_{ipt}>0,\eta_{pj}=1]}{\partial \sigma_{agg,pt}} < 0$. Defining $g(x) \equiv \frac{\phi(x)}{\Phi(x)}$, we have that

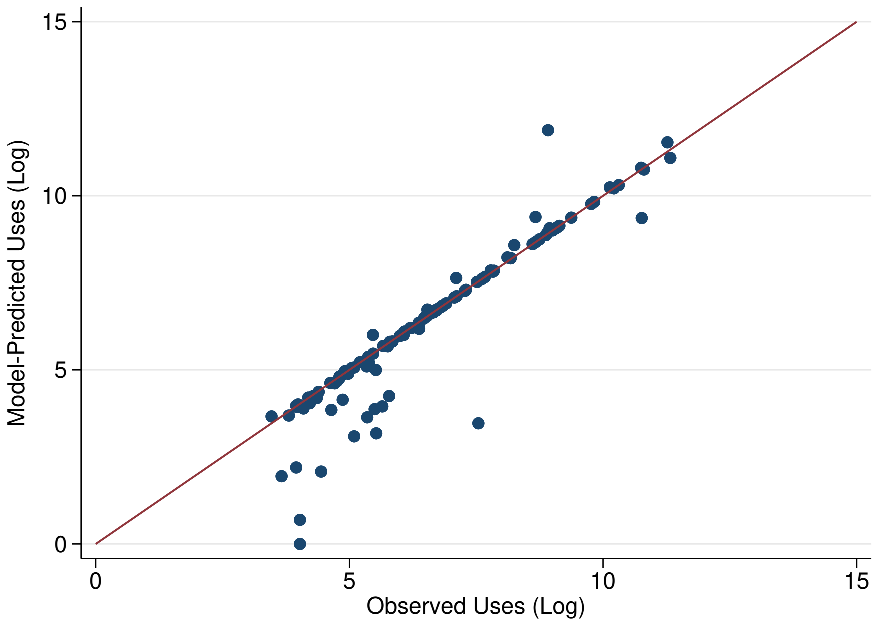
$$\begin{aligned}\frac{\partial \mathbb{E}[W_{iptj}|U_{ipt} > 0, \eta_{pj} = 1]}{\partial \mu_{agg,pt}} &= -\frac{g'\left(\theta_{cov,pt}\right)\theta_{cov,pt}\sigma_{agg,pt}}{\left(\sigma_{agg,pt}^2 + 1\right)^{\frac{3}{2}}} - \frac{\sigma_{agg,pt}\, g\left(\theta_{cov,pt}\right)}{\left(\sigma_{agg,pt}^2 + 1\right)^{\frac{3}{2}}} \\ &= -\frac{\sigma_{agg,pt}}{\left(\sigma_{agg,pt}^2 + 1\right)^{\frac{3}{2}}}\left(g'\left(\theta_{cov,pt}\right)\theta_{cov,pt} + g(\theta_{cov,pt})\right) < 0,\end{aligned}$$

where the inequality follows from the fact that $xg(x) + g'(x) > 0$ is less than zero for all x.

G Model Fit

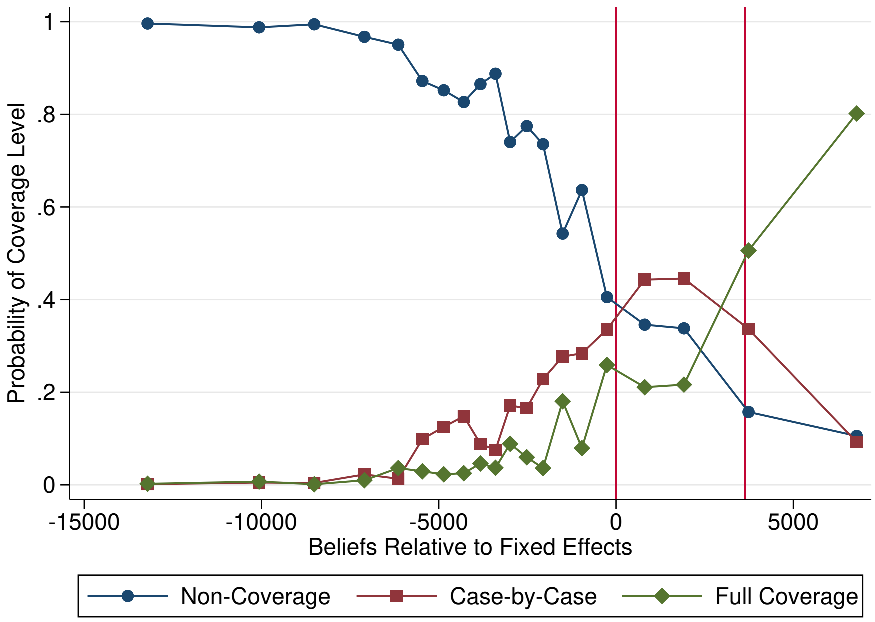
In this appendix, I assess the fit of my model to the data. Figure A10 below plots the observed and model-predicted level of utilization for each procedure in the sample. Note that with few exceptions, the predicted uses of the procedure are nearly identical to the observed level of utilization. Figure A11, I show that the observed likelihood of coverage matches the coverage levels predicted by my model of administrator coverage decisions, with clear changes as the estimated coverage thresholds.

Figure A10: Observed and Predicted Levels of Utilization



Notes: The figure reports the observed uses of each procedure along with the model-predicted number of uses (both in logs) for each procedure in the structural estimation sample. An observation is a procedure. Sample is limited to the 136 procedures for which the model is identified.

Figure A11: Observed and Predicted Coverage Levels



Notes: The figure reports a binned scatterplot of the probability of coverage at different levels based on beliefs relative to the estimated coverage threshold for each procedure-administrator-month tuple. The vertical red lines give the estimated thresholds for case-by-case (at zero) and full coverage.