NBER WORKING PAPER SERIES

REGULATING INNOVATION WITH UNCERTAIN QUALITY: INFORMATION, RISK, AND ACCESS IN MEDICAL DEVICES

Matthew Grennan
Robert Town

Working Paper 20981
http://www.nber.org/papers/w20981

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2015, Revised June 2019

Data for this research was provided by Millennium Research Group (www.mrg.net). We owe thanks to Jason Abaluck, Andres Aradillas-Lopez, Steve Berry, Dave Chan, Allan Collard-Wexler, Joe Harrington, Ali Hortaçsu, Robin Lee, Sam Peltzman, Scott Stern, Ashley Swanson, Mauricio Varela, and numerous seminar audiences for helpful comments. The University of Toronto, Wharton Dean's Research Fund and Mack Institute, and the Indian School of Business provided financial assistance. Alexandra Charbi, Gi Heung Kim, Ellie Prager, Mike Punzalan, and Emi Terasawa provided excellent research assistance. Any errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Regulating Innovation with Uncertain Quality: Information, Risk, and Access in Medical Devices

Matthew Grennan and Robert Town

NBER Working Paper No. 20981
February 2015, Revised June 2019
JEL No. I11,L11,L51

ABSTRACT

We study the impact of regulating product entry and quality information requirements on an oligopoly equilibrium and consumer welfare. Product testing can reduce consumer uncertainty, but also increase entry costs and delay entry. Using variation between EU and US medical device regulations, we document patterns consistent with valuable learning from more stringent US requirements. To derive welfare implications, we pair the data with a model of supply, demand, and testing regulation. US policy is indistinguishable from the policy that maximizes total surplus in our estimated model, while the EU could benefit from more testing. "Post-market surveillance" could further increase surplus.

Matthew Grennan
The Wharton School
University of Pennsylvania
3641 Locust Walk
Philadelphia, PA 19104
and NBER
grennan@wharton.upenn.edu

Robert Town
Department of Economics
The University of Texas at Austin
2225 Speedway
BRB 1.116, C3100
Austin, TX 78712
and NBER
robert.town@austin.utexas.edu

1 Introduction

Most innovative new products are brought to the market because their makers believe they create new value. However, with innovation often comes uncertainty, and once in the hands of consumers, there is some chance that the product will not operate as hoped. The consequences of this failure range from consumer regret to death. When this uncertainty matters for welfare, products often must go through pre-market testing and become approved and certified by a formal body before entering the marketplace. Especially in oligopolistic markets, where private and public incentives may diverge (Spence 1975), the standard that a regulatory body imposes has the potential to fundamentally alter market outcomes by requiring testing that firms would not undertake themselves. As first highlighted by Peltzman (1973) in the context of pharmaceuticals, higher testing standards can create value through generating information that decreases uncertainty, but this benefit comes with the potential cost of fewer available products due to delayed entry and higher entry costs from more testing. Today such certification processes are commonplace and a source of controversy in areas as diverse as airplanes, automobiles, electronics, finance, health care, and toys.[1]

In this paper, we use new, detailed data and exploit exogenous regulatory differences between the United States (US) and European Union (EU) to identify the impact of product testing requirements (and the information and costs they generate) on market outcomes for medical devices.[2] Among its many duties, the US Food and Drug Administration (FDA) oversees medical device regulation in the US, while in the EU medical device approval is performed by private organizations called "notified bodies." The FDA applies a "safe and effective" standard while EU notified bodies only certify the safety performance of the product. For the Class III medical devices we study, this difference is material. Meeting the "effectiveness" standard often requires manufacturers to generate product performance information through large, randomized clinical trials. These trials are costly in both time and expense. As a result, medical device manufactures (many of which are US based) typically introduce products in the EU well before they are granted FDA approval, if they enter the

[1]See, for example in airplanes "Boeing Acknowledges Tests Underestimated 787 Battery Risks", The New York Times, April 23, 2013; in automobiles "U.S. Sues Chrysler After Auto Maker Refuses to Recall Cars", The New York Times, June 5, 1996; in electronics "European Environmental Rules Propel Change in U.S.", The New York Times, July 06, 2004; in finance "An FDA for Securities Could Help Avert Crises", Bloomberg, April 2, 2012; in toys "Toy Makers Fight for Exemption From Rules", The New York Times, September 28, 2010.

[2]In particular, our analysis focuses on coronary stents 2004-13. We chose this segment as the coronary stent market is large and important with excellent market data and with constant innovations introduced over time. Coronary stents treat ischemic heart disease—the narrowing of the coronary artery caused by fatty deposits. Ischemic heart disease is the leading cause of global death accounting for 7 million fatalities in 2010 (Lozano 2012). In 2013, total worldwide sales of coronary stents exceeded $7.9 billion with the vast majority of those sales occurring in the US and the EU.

US at all.

The shared "safety" standard between the US and EU represents one of several limitations of our study for understanding the varied issues involved in product testing regulation in general. It means that we do not have variation to credibly examine differential safety standards or the impact of no safety regulation at all. In keeping with the variation available in our context, we focus on measuring the costs and benefits of additional effectiveness information, which is also where the current policy debate for Class III medical devices is centered.[3]

The differences between the US and the EU in the medical device approval process have led to calls for reform in both regions. In the US, the FDA has faced criticism from some commenters claiming that a slower, tougher approval process is crippling innovation. However, others have taken the opposite view claiming that the approval process is too lax.[4] Congress has responded to this debate by including measures in the 21st Century Cures Bill that would change the amount of information the FDA is allowed to require before market approval.[5] In April 2017, the EU amended the Medical Device Directive, increasing data collection on high risk devices both before and after they are allowed into the market.[6]

Despite its broad importance, empirical research on testing and information provision for innovative new products is scarce. One major challenge is finding exogenous variation in information provision regimes. To address this challenge, we exploit the fact that the EU approval process requires less intensive pre-market testing from manufacturers, and as a consequence it is both faster and less costly than the US approval process for any given Class III device. We describe this difference in detail and argue that it is due to historical political processes that are not correlated with market demand for Class III devices. From a research design perspective, this setting provides us with two key features: First, we are able to observe market outcomes for a number of new devices that are invented and enter the EU where regulatory requirements are less stringent. This provides us with a candidate population of products that might potentially be developed and enter (or not) under more

[3]For the reader more familiar with pharmaceutical clinical trials, EU Class III device requirements could be thought of as similar to Phase 1 and 2 pharmaceutical requirements, and the additional US Class III device requirements as similar to Phase 3 pharmaceutical requirements. Thus, in the context of pharmaceuticals, our analysis could be thought of as asking "How much Phase 3 testing should be required?" There are other features of our context, however, that may differ from some pharmaceutical markets. In the Conclusion, we provide more discussion on the extent to which our study of coronary stents might extrapolate to other product markets.

[4]For an example arguing the FDA is too lax "Report Criticized F.D.A. on Device Testing", The New York Times, January 15, 2009; and too tight "FDA Seeks to Toughen Defibrillator Regulations", The New York Times, March 22, 2013.

[5]See "How Not to Fix the FDA", The New York Times, July 20, 2015.

[6]See http://ec.europa.eu/growth/sectors/medical-devices/regulatory-framework/revision˙en.

stringent requirements. Second, we observe market outcomes for some devices under two regulatory regimes with different pre-market testing requirements. Most importantly, we observe EU market outcomes for devices that are concurrently undergoing US trials as well as for those devices that are not undergoing US trials. This allows us to examine the EU market response to the information generated by additional US trials. The key additional identifying assumption for this comparison (which we verify in the data) is that selection into US trials is based on the *level* of expected US profits, not uncertainty about product quality at the time of EU entry.

A further challenge is assembling a dataset of sufficient detail and scope to credibly identify the impact of different regulatory information regimes. We assemble monthly data on prices and quantities for all coronary stents implanted at a large number of hospitals in the US and the EU from 2004-13. Paired with product-month variation in participation in US clinical trials, revealed preference arguments imply that such data capture the state of market knowledge over the expected performance of a device, the uncertainty over that expected performance for these devices, and, in turn, consumer choice patterns and welfare.[7] We augment the market data with hand-collected data on clinical trials, which help to more precisely demonstrate the differences in US and EU requirements, and also allow a validation of our revealed preference estimates of product quality.

We begin by documenting multiple patterns in the data. The EU enjoys greater access to the newest medical technologies. On average, US physicians have 11 stents available to implant while their EU counterparts have 39 from which to choose – 81 percent of products (accounting for 23 percent of stents used) in the EU never enter the US. Conditional on the product entering the US, EU physicians have access to the product 10 months earlier.

However, EU consumers also face greater performance uncertainty by allowing entry with less evidence on product efficacy. A clinical trial has been published for only 20 percent of EU-only available devices. In contrast, 85 percent of FDA approved devices have undergone a published clinical trail. Also, conditional on publishing a clinical trial, average sample sizes for the FDA approved devices are 1313 patients vs. 280 for the EU-only devices. This extra evidence comes at a cost as the additional subjects are associated with an extra 9 months in trials (due to recruitment time). This time is costly in terms of delayed access for patients

[7]In this sense, our approach contrasts with studies of the FDA using product introductions and withdrawals (e.g., Grabowski and Wang 2008; Olson 2008; Philipson et al. 2008). The EU does not record introductions or recalls of devices in a publicly available database. More importantly, our interest is in understanding whether further efficacy testing required by the US provides more precise information on product performance, on which negative tail events such as recalls provide little information. See Stern et al. (2017) and Nistor and Tucker (2015) for analyses of the correlations between FDA review time and adverse event reporting for cardiac devices and the benefits and limitations of using adverse event data to infer device safety.

as well as raising fixed costs of entry.[8]

To explore whether the information generated by additional testing for US approval is valuable to consumers, we look to the market usage data in the EU and compare products that undergo FDA trials to those that do not. As expected, in both the clinical trial and market usage data, the products that begin US trials appear *better on average* at the time of EU introduction. They are more clinically efficacious and are more likely to be implanted. However, both sets of products have *similar levels of uncertainty* in terms of the standard deviation of efficacy and usage upon introduction. Thus, selection into US trials appears to be on differences in expected outcomes, not uncertainty about those outcomes.[9]

After EU entry, the two sets of products display different usage dynamics. For those products in US trials, volatility in usage decreases over time, consistent with learning from the trials. Average usage also increases as volatility decreases, consistent with consumers valuing this additional information and facing decreasing risk. Products not in trials exhibit neither of these patterns. We consider alternative mechanisms such as non-learning models of product diffusion, learning from observational use versus learning from clinical trials, and signaling with asymmetric information. We conclude the evidence is strongest for EU market learning from information spillovers from US clinical trials, with firms and physicians facing symmetric uncertainty about these *additional* trials.[10]

In order to derive welfare measures and address policy questions regarding optimal regulation, we construct and estimate a structural model of an agent charged with regulating medical devices and medical device manufacturers and consumers optimally responding to

[8]See Makower et al. (2010) for industry estimates of $1.6M per month for a Class III medical device trial.

[9]Note that this assumption is key in the reduced form analysis (to rule out that the evidence of no learning in the sample of products not in US trials is not due to there simply being nothing to learn about them), but a weaker version is needed in the structural analysis, where we model and estimate the distribution of product qualities for each subsample. There the key to the credibility of our counterfactual analysis is that the same learning rate we estimate from the set of products in US trials would apply to the set of products not in trials, if they were to undertake them (a slightly milder "parallel trends" assumption).

[10]Manufacturers may indeed have private information about their device prior to undertaking in-human trials, but our analysis of the data is most consistent with a model where firms and physicians are symmetrically (un)informed after the results of trials required for EU market entry are released. This is a departure from the asymmetric information that is frequently the focus of discussion in regulation of pharmaceutical markets (Scott Morton and Kyle 2012) and in the broader literature on certification (Dranove and Jin 2010). Our institutional setting of coronary stents – where trials generate important information that could not be otherwise obtained by manufacturers and interventional cardiologists pay close attention to new technologies being developed – is a case where symmetric information seems like a reasonable approximation to the first-order forces at work. We believe that many markets with published testing results and informed consumers may also fit this model, and indeed symmetric information games of persuasion and information disclosure have recently received increased attention in the literature (e.g., Kamenica and Gentzkow 2011). In the conclusion, we discuss in more detail what we perceive as the boundaries to our analysis, in particular the ways in which other product markets may be similar or different to coronary stents 2004-13.

the agent's regulatory policy. In the model, products are invented with uncertain performance characteristics. EU and US regulators establish product performance statistical thresholds that the product must meet before it can be marketed to consumers in those regions. These performance thresholds are designed to limit the likelihood that harmful devices (or devices that provide limited health benefits) are marketed to consumers. The statistical thresholds determine the clinical trial sample size which, in turn, maps into the manufacturer's entry cost and time required to run the trial. Consumers learn about product performance through these trials and/or potentially through observational learning once a product is available in the marketplace.

We estimate the structural parameters of the demand model using detailed product-hospital-month price and quantity data and our hand collected data on the timing and results of clinical trials. Our demand system combines a model of utility over health outcomes (Cardon and Hendel 2001; Handel 2013) with a model of consumer learning (Roberts and Urban 1988; Erdem and Keane 1996; Ackerberg 2003; Crawford and Shum 2005; Ching 2010) and recent work by Quan and Williams (2017) that accounts for regional variation in tastes (and in our adaptation, hospital variation in learning processes). The model provides an internally consistent approach to estimate the perceived stent quality distribution, market and hospital level learning about product quality, consumer risk aversion over health outcomes, and heterogeneity in preferences over stent attributes across hospitals and patients/doctors.

The demand model generates sensible parameter estimates which we also validate using outside data sources. Consistent with the reduced form evidence, they imply that FDA required clinical trials generate useful information, and there is practically no hospital or market level observational learning via market usage experience in the EU marketplace.[11]

Combined with product quality estimates that indicate significant variation in stent quality, this implies the returns to early product testing are large for stents. Without any EU testing, the market for stents would shrink significantly. Further, the estimates suggest that required US testing in excess of EU requirements substantially decreases the uncertainty of using an inferior product and thus significantly increases consumer surplus. It also implies that the EU enjoys positive spillovers from US testing – if US testing were equal to the EU, welfare in the EU would decline by 6.4 percent. Our demand estimates also allow us to calculate technological change in the EU stent market where we find that (from 2004 to 2013) consumer surplus increased by 10 percent, and this increase was driven by an increase in product variety and not from increases in the mean quality of newly introduced products.

[11]The estimate of no observational learning in the EU for coronary stents is not surprising, given that there is currently no systematic data collected that links stents used to clinical outcomes. It is exactly this lack of data that has prompted calls for more "post-market surveillance" that we examine in some of our counterfactuals.

We then consider optimal regulatory policy that balances risk from efficacy uncertainty versus access to new devices. A full model of supply requires work at or beyond the methodological frontiers of the buyer-supplier contracting and dynamic entry game literatures, which we leave for future research. Instead we consider simple-to-compute cases that approximate larger and smaller sets of firms that are expected to enter as a function of regulatory policy and firm behavior, and we explore welfare outcomes and optimal regulatory policy under these cases. Our estimates imply that EU surplus could increase by as much as 3-7 percent by requiring more pre-market testing for stents. Indeed, total surplus is maximized when the premarket trials are at least 16-17 months longer than current EU requirements. Thus, for stents 2004-13, US regulatory policy is statistically equivalent to the policy that maximizes surplus in our estimated model. We explore the factors that affect the optimal trial length calculation and find that the optimal trial length is: decreasing in costs of trials, increasing in the quality of existing technology, and non-monotonic in the precision of clinical trial information.

Our final piece of analysis examines optimal policy under counterfactual regimes with greater "post-market surveillance." This idea, which is a centerpiece of the 21st Century Cures legislation, has a straightforward logic. Increased post-market learning could maintain risk reduction while lowering pre-market requirements, thus decreasing entry lags (and potentially costs). We find that if post-approval learning rates could approach those we observe from clinical trials, the benefits from such a policy change are substantial.[12] An extreme case, where post-approval learning is as informative as pre-market trials, at zero incremental cost, would yield an estimated welfare increase of 15-18 percent.[13]

Our focus on information and market structure is complementary to recent empirical research on other regulatory tools that affect late-stage product development and entry incentives, such as patent breadth and length (Budish et al. 2015), price regulations (Kyle 2007; Filson 2012), and regulatory uncertainty and innovation incentives (Stern 2017). Whereas the focus of that literature is on the extent and timing with which products undergo late stage R&D/testing and eventually enter the market, we show that, in addition to these innovation and entry implications, the welfare impact of the product performance information generated can also be large. New medical technologies with uncertain quality can only achieve their welfare potential if the necessary clinical trial studies are performed to inform

[12]The FDA recently introduced a Unique Device Identifier system that could facilitate post-market data collection. However, there is currently debate regarding if/when UDIs may be added to patient claims data.

[13]It is unclear how extreme this case is. On one hand, post-market learning would likely lack the clean randomization and blinding of a clinical trial, decreasing learning. However, real-world usage patterns might be the policy effect of interest, and real-world use might see more cases and uncover rare events a clinical trial could not.

their proper use. This also points to one of the limitations our analysis shares with much of the prior research – we do not consider the impact of changes in regulatory requirements on earlier stage innovation.

More broadly, our work relates to recent empirical research that estimates model primitives without imposing optimality of the regulatory environment, and thus can use the estimated model to study optimal regulation itself (Timmins 2002; Seim and Waldfogel 2013; Miravete et al. 2014; Hamilton et al. 2018). Combining this literature with recent developments in modeling consumer demand with learning is essential in allowing us to build upon the work of Peltzman (1973) in measuring the impact of regulatory information requirements. As we build on established models, we provide an approach that others with similar data and variation in information regimes might find useful for studying regulation in other markets, such as the value of price information studied in Brown (2017). Our work also relates to the literature measuring the value of new products in general (Petrin 2002; Quan and Williams 2017; Aguiar and Waldfogel 2018), where our integration of quality uncertainty seems like an important component to account for in many industries, as referenced in the opening paragraph.

Beyond the economics of information and product quality regulation, our analysis also speaks to an active and contentious policy debate with potentially large welfare consequences. The amount of economic activity regulated by the FDA and the notified bodies is significant. As of 2010, medical device sales exceeded $150B or 6 percent of total national health expenditures in the US and $130B (7.5 percent) in the EU.[14] Further, the introduction of new medical technologies are responsible for significant reductions in mortality; and in so far as different regulatory regimes affect the availability of these technologies, their welfare impact extends beyond the direct costs of the devices themselves (Murphy and Topel 2006).

2 Medical Device Regulation in the US and the EU

The term medical device applies to a broad set of product categories, ranging from crutches to pacemakers to CT scanners. In this study we focus on coronary stents, a blockbuster device in terms of sales and health impact, but also typical of implantable devices that are deemed "necessary for the sustainment of life" and thus regulated as Class III devices in the US and EU. It is for Class III devices that EU and US regulatory approaches diverge most widely, creating the variation we leverage in our study.[15] Coronary stents are a small metal

[14] Donahoe and King (2012) and Medtech Europe (2013).

[15] Class I devices are low risk devices such as elastic bandages are subject to 'general controls' and do not require pre-market approval. Class II devices are higher risk in which general controls alone cannot assure safety and effectiveness (e.g. infusion pumps).

mesh tube that is inserted into the coronary artery to treat atherosclerosis (the build up of lesions of plaque that narrow the arteries).

Before detailing these regulatory differences, it is useful to keep in mind some basic facts about the structure of decision making and the players in the market. First, hospitals generate revenue by performing a procedure (such as an angioplasty with stent), and the price for purchasing the device is an input cost the hospital incurs. The physician who performs the procedure will typically be compensated either as a salaried employee of the hospital, or on a fee-for-service basis for the procedure, where in either case the financial benefits to the physician are unrelated to the specific brand of device used. Physicians typically have strong preferences over which specific product is best to use for a given patient/lesion type (devices in this class are often referred to as "physician preference items") because devices are differentiated in physical characteristics of the implanted device itself (for a stent examples are shape, strength, flexibility, and type of drug/polymer) and also characteristics that affect ease of implantation (for stents: unexpanded size and flexibility, and controls and capabilities of the catheters and balloons used in delivery). The supply-side of the market is thus a differentiated oligopoly, and prices are typically negotiated between manufacturers and hospitals, hospital systems, or regional purchasing authorities.

For the purposes of this study, the most important features of the stent market to note are the constant introduction of new products. These may differ from incumbent products by offering clinical performance improvements or by design modifications to address less common niche markets such as small vessel and bifurcated lesions. The two most common stent types are bare metal (BMS), first introduced to the US in 1994; and drug eluting (DES), first US introduction 2003, which are coated with a polymer and drug to inhibit scar tissue growth. Interventional cardiologists are a relatively small and technologically aware community who stay engaged through close relationships with manufacturers, journals, and several well-attended meetings each year (Transcatheter Cardiovascular Therapeutics each Fall, American College of Cardiology in March, and European Society of Cardiology in August each year, as well as numerous regional affiliated conferences throughout the year) at which the most recent results of in-progress clinical trials are reported.

2.1 Similarities and Differences in US and EU Regulation

Medical device regulation in the US began with the Medical Device Amendments Act of 1976, placing oversight authority within the FDA. The criteria the FDA is mandated to use is "safe and effective." The Act established three classifications of devices (I, II and III), based on perceived health risk. Class III devices are defined as those used in "supporting or sustaining human life, of substantial importance in preventing impairment of human health,

or presenting a potential unreasonable risk of illness or injury."

In the US, the approval process for a Class III device generally requires data from randomized clinical trials, involving thousands of patients and costing tens of millions of dollars to complete.[16] The FDA plays a significant role in determining the design, statistical power, clinical endpoints and duration of the trial (Kaplan and Stern 2018). The FDA also insures that the proper clinical trial best practices are used (e.g. data management, audits, core laboratory review), while clinical studies performed outside of the context of obtaining FDA approval typically lack many of these best practices (Kaplan and Stern 2018). For stents, the FDA generally requires the trial to demonstrate efficacy on a number of clinically meaningful end points including target lesion revascularization (TLR), death, and major adverse cardiac events (MACE) which is a composite of death, myocardial infarction (heart attack), stent thrombosis, and target lesion revascularization.

In the EU, the regulatory process is quite different, governed primarily by the Medical Devices Directive of June, 1993, which has been adopted by each EU member state. A medical device is approved for marketing in the EU once it receives a 'CE mark' of conformity. The CE mark system relies heavily on third parties know as "notified bodies", which are independent, commercial organizations that are designated, monitored, and audited by the member states via "competent authorities." As of this writing, there are more than 70 active notified bodies within the EU. A firm is free to choose any notified body designated to cover the particular type of device.[17] To obtain a CE mark, a Class III medical device needs to demonstrate safety and performance. Compliance with this standard can usually be demonstrated with simpler, cheaper clinical trials than required by the FDA. Once a device has been approved for use in one EU country, it can be marketed in any member country.[18]

Despite their differences, both regions require the submission of similar, detailed engineering and manufacturing process information to assess safety and some measures of perfor-

[16]There are two common pathways to bring a device to market: Pre-Market Approval (PMA) for Class III devices, and the 510(k) for Class II and some Class I devices. Under the 510(k) process the manufacturer demonstrates the device is 'substantially equivalent' to a predicate device. Bench testing and perhaps a small clinical study are all that are typically necessary. A straightforward 510(k) clearance can typically be obtained within months.

[17]See *Guidelines Relating to Medical Devices Directives, http://ec.europa.eu/health/medical-devices/documents/guidelines/.*

[18]In both the US and EU, new-to-the-world devices may face the additional hurdle of gaining payor reimbursement, but the devices we study are second (and later) generation products, so coverage and payment determination has already been made. Coverage decisions are generally based on cost-effectiveness and budgetary impact analysis performed at the national level. For the EU countries in our sample, hospitals are typically paid on a per procedure basis and the hospitals pay for devices used in the procedure as part of the cost of providing care (Schreyögg et al. 2006). The price of the device is determined through bilateral negotiations between the device manufacture and either a local or regional purchasing authority (Sorenson and Kanavos 2011).

mance. Thus, insofar as the EU testing requirements successfully prevent "unsafe" devices from reaching the market, we do not have in-sample variation allowing us to assess the value of these minimal "safety" standards shared by both regimes. This places our primary focus on the value of additional FDA "efficacy" testing, which is also the region of focus of the current policy debates.[19]

The difference between the two regulatory regimes implies that there will be variation in the information sets available to physicians on the performance of the stent across devices marketed in the EU. Devices undergoing FDA required trials in order to enter the US market will run large, costly, randomized clinical trials while those devices that won't enter the US will not. In our sample, all devices that are ultimately granted FDA approval are sold in the EU well before they are granted FDA clearance. This is the variation we will leverage to understand the extent to which the additional FDA "efficacy" testing in the human body generates information that the marketplace values.

It is also important to note that the differences between the EU and the US are largely a consequence of different histories that led to the passing of the primary medical device legislation in the two regions (Van Norman 2016). The Medical Device Directive, the centerpiece of the EU medical device regulatory framework, was passed in 1993 when there was keen interest in a new approach to harmonizing regulatory frameworks across the member states. The EU had just undertaken a long and frustrating harmonization process for food and drugs. This new approach sought to avoid detailed and bureaucratic government approval processes, particularly duplicative approvals. This framework was also applied to other products including toys, pressure vessels, and personal protective equipment. In contrast, the US medical device regulatory framework was established after the Dalkon Shield injured several thousand women, garnering significant public outcry. The FDA already had oversight on some aspects of medical devices, and expanding that role was the most viable political option. At that time, a non-governmental approach to device regulation was never seriously considered by the Congress.

The gap between the two regulatory systems is the focus of a number of consulting and government reports. For example, a series of Boston Consulting Group (BCG) reports shows

[19]It is common to view "safety and effectiveness" as separate concepts. In our context, they can best be thought of as lying on a single dimensional continuum. For example, a key endpoint for the FDA in assessing a stent is Target Vessel Revascularization (TVR), the need for a repeat procedure on the same lesion, but it is not obvious if a TVR rate of say 10 vs. 5 percent in one month should be categorized as a deficiency in safety, efficacy, or both. The FDA implicitly acknowledges this is as it does not distinguish different clinical endpoints for safety and effectiveness for cardiac stents – they categorize the clinical analysis as simply "safety and effectiveness". See, for example, http://www.accessdata.fda.gov/cdrh_docs/pdf15/P150003b.pdf. While we believe that the single quality/safety dimension assumption is a good approximation of the coronary stent setting, it might not be generalizable to other health technology environments.

that there is no difference in recalls between devices that are marketed in both the US and the EU. The FDA countered the BCG report with their own case study of 12 devices that were approved in the EU and were not approved by the FDA. They found that four of those devices caused significant adverse events in patients, and the other eight devices would not have met the FDA's efficacy standard. While there are highly publicized events in which a device clearly and obviously causes significant harm, those cases are rare. This is not surprising, given that both the EU and FDA require significant safety testing. Perhaps most importantly, by focusing on extreme, rare cases of recalls and adverse events, none of these studies address the primary difference inherent in FDA vs CE Mark requirements for Class III devices—more precise estimation of product efficacy.

It is important to note that while unsafe stents appear not to have been marketed in the EU, the clinical trial results suggest meaningful differences in the clinical efficacy of stents. For example, in Medtronic's FDA approval for its Endeavor stent, the summary reports that Endeavor's 9-month major adverse cardiac event (MACE) rate is equivalent to Boston Scientific's Taxus Express II and 20 percent less than Johnson and Johnson's Cypher stent. Its target vessel failure (TVF) rate was 8 percent less than the Taxus stent.[20] The impact of TVF is significant as it requires additional interventions to restore vessel function.

3 Data Summary and Reduced Form Analysis

The primary data set used in this study consists of quantities and prices at the product-hospital-month level, collected by Millennium Research Group's (MRG) *MarketTrack* international survey of hospitals from 2004-2013. This survey, covering approximately 10 percent of total market activity, is the main source of detailed market intelligence in the medical device sector. Its goal is to produce representative estimates of product market shares and prices by region. Importantly, MRG also tracks the number of diagnostic angiographies (a procedure that must be preformed before a stent can be inserted), providing the number of patients potentially eligible for a stent in each hospital-month. The countries in our sample are US, France, Germany, Italy, Spain, and the United Kingdom. These data are quite large with 494,304 product-hospital-month observations across 372 hospitals in the US and 416 hospitals in the EU.

We supplement the detailed market data with our own searches for product approval dates in the EU and US in order to validate data coverage within our sample and also to determine the time in market for products that enter outside of our sample period. In addition, we collected clinical trial data (when available) from various journal articles, conference

[20] http://www.accessdata.fda.gov/cdrh_docs/pdf6/P060033b.pdf

abstracts, press releases, and product catalogs. These provide further evidence regarding the size and length of trials required for US versus EU entry. They also provide clinical outcomes, which we use to validate our revealed preference estimates of product quality.

Figure 1 summarizes statistics on testing and market access in the EU vs. US. The top third of the table in panel (a) presents summary statistics for our clinical trial data, listing data on trials with primary endpoints completed prior to entry in each market. We were able to find such data for almost all of the products entering the US and 20 percent of the products that enter the EU. Conditional on publishing a clinical trial online, EU trials are shorter and enroll fewer patients. On average, by the time a product enters the US, it has undergone 2 clinical trials, enrolling over 1300 patients and lasting 28 months in total, while upon entering the EU, the typical product has completed only a single trial with 280 patients lasting 19 months. This large difference in trial patterns is not surprising given the testing requirement differentials across the two regions.

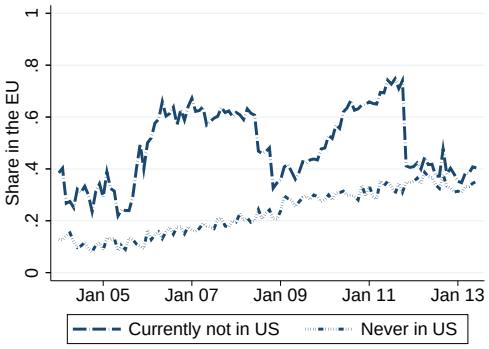
Interestingly, the modal/median follow up time for the primary trial endpoint across all of these trials is 12 months, so the additional time in US trials is driven primarily by the additional time required in patient recruitment for a larger trial.[21] This points to the primary cost of generating information through clinical trials – more certainty in performance estimates requires recruiting more patients, which takes more time (delays entry), and is more expensive (raises fixed costs of entry).[22]

Figure 1: Stent clinical trials and market structure in the US and EU.

(a) EU vs. US: Clinical and Market Data

	US	EU
Clinical trial data:		
Pct products w/ published trials pre-entry	85.7	20.1
Median number of trials	2	1
Median total trial size (patients)	1313	280
Median total trial time (months)	28	19
Market structure data:		
Mean manufacturers in market	4 (3)	21 (5)
Mean products in market	11 (5)	39 (8)
Total products in market 2004-13	21 (11)	109 (22)
Mean months EU to US entry	10	-
Mean months EU to US entry (DES)	17	-
(Usage within hospital in parentheses.)		

(b) EU usage of stents not in US



The bottom two thirds of the table, and the graph in panel (b), show how these pre-

[21] Total trial length equals recruitment time until the last patient is recruited, plus the 12 months until the primary follow up for the last patient.

[22] See Appendix A.2 for more detailed figures and regressions relating number of patients and trial recruiting time for EU and US trials.

market testing requirements are correlated with market structure and product usage in the US and the EU over our sample period. The EU has over three times as many manufacturers and products as the US (and still nearly two times as many when measured at the hospital rather than region level). For those products that eventually enter the US, the average lag between EU and US introduction is 10 months (17 months for the more technologically advanced DES).

Many of the products to which the EU has greater access are important and frequently used. In the average month, 49 percent of the stents used in the EU are unavailable in the US at that point in time, and 23 percent will never be available in the US.

These basic clinical trial and market structure data illustrate the tension between the two regulatory approaches. The EU enjoys greater access to a broader variety of devices, and these devices are available earlier than in the US. However, EU consumers have less testing on the health impact of these products. The goal of our analysis will be to determine, for our sample of coronary stents 2004-13, whether the extra US testing provides information that the market values in terms of decreasing uncertainty, the extent to which there is observational learning outside of clinical trials, and the value of access to more products earlier in the EU versus the value of any reductions in uncertainty.

3.1 Evidence: Information and Market Response

We next turn to examining the patterns in adoption and diffusion of stents by region and FDA trial status. Figure 2 illustrates the evolution of three different statistics plotted against product age (defined as time since introduction to the region) for three subsets of the data: the US, the EU for products running clinical trials to enter the US, and the EU for other products. The figures are constructed after controlling for product fixed effects, so that all patterns are driven by within-product variation over time.[23]

Panel (a) plots the empirical mean across products of a given age of $E_{j|a}\left[\widetilde{\ln(s_{jt}/s_{0t})}\right]$ where s_{jt} is the within region share of product j in month t and s_{0t} is the relevant outside good share based on the number of reported angiographies – this proxies for the mean perceived stent utility (which incorporates both the perceived uncertainty and clinical performance) at age a.[24] For the EU products undergoing US trials, this value is lower upon introduction

[23]Within product variation, re-centered at mean across all products for each sample to preserve level differences across the samples, so: $\widetilde{\ln(s_{jt}/s_{0t})} = \ln(s_{jt}/s_{0t}) - E_{t|j}\ln(s_{jt}/s_{0t}) + E_{jt}\ln(s_{jt}/s_{0t})$. See Appendix C.3 for proofs regarding the relationship between the patterns in the moments of the market share (as proxies for mean utility) distribution and age/information explored here.

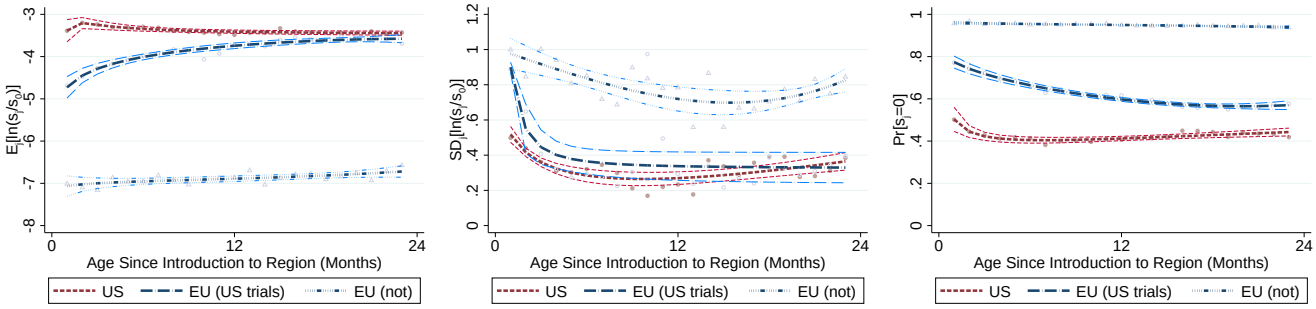
[24]We chose this measure to balance allowing for some basic controls on the data without putting too much structure on this exploratory analysis. This would be exactly mean utility in a logit model, so along with product fixed effects, this measure controls at least in part for competition and substitution. Our subsequent structural analyses control explicitly for several additional sources of variation.

and gradually increases with age, plateauing after approximately two years in the market. As we discuss in more detail in the next section, this trend is consistent with a model where consumers learn from US trials and increase average usage as uncertainty is resolved. However, it is also potentially consistent with observational learning by product experience in the market, or with drivers of diffusion other than learning. We will use the two other product subsets and two other statistics to examine these differing explanations.

Figure 2: Usage patterns after entry, by region and trial status.

Within product variation, re-centered at mean across all products, so: $\ln(\widetilde{s_{jt}/s_{0t}}) = \ln(s_{jt}/s_{0t}) - E_{t|j}\ln(s_{jt}/s_{0t}) + E_{jt}\ln(s_{jt}/s_{0t})$ for each of three samples. Lines fit using fractional polynomials with standard errors clustered by month. Coefficients in table below from regressions of each dependent variable (disaggregated except for SD) regressed on age fixed effects: $\widetilde{y}_{jt} = \theta_a + e_{jt}$. $\Delta\theta_a := \theta_{a=24} - \theta_{a=1}$.

(a) $E_{j|a}\left[\ln(\widetilde{s_{jt}/s_{0t}})\right]$ **(b)** $SD_{j|a}\left[\ln(\widetilde{s_{jt}/s_{0t}})\right]$ **(c)** $E_{j|a}\left[Pr(\widetilde{s_{jht}=0})\right]$



	$\theta_{a=1}$	$\theta_{a=24}$	$\Delta\theta_a$	$\Delta\theta_a^{EU\|_{USTrials}} - \Delta\theta_a^{row}$
$E_{j\|a}^{US}\ \ln(s_{jt}/s_{0t})$	-3.39	-3.43	-0.04	1.04***
	(0.12)	(0.14)	(0.18)	(0.29)
$E_{j\|a}^{EU\|_{USTrials}}\ \ln(s_{jt}/s_{0t})$	-4.69	-3.69	1.00***	
	(0.27)	(0.09)	(0.28)	
$E_{j\|a}^{EU\|not}\ \ln(s_{jt}/s_{0t})$	-7.01	-6.57	0.44*	0.56*
	(0.13)	(0.14)	(0.19)	(0.34)
$SD_{j\|a}^{US}\ \ln(s_{jt}/s_{0t})$	0.50	0.38	-0.11	-0.38***
	(0.05)	(0.06)	(0.07)	(0.20)
$SD_{j\|a}^{EU\|_{USTrials}}\ \ln(s_{jt}/s_{0t})$	0.89	0.39	-0.49***	
	(0.18)	(0.05)	(0.19)	
$SD_{j\|a}^{EU\|not}\ \ln(s_{jt}/s_{0t})$	0.99	0.84	-0.15	-0.34***
	(0.08)	(0.12)	(0.15)	(0.25)
$Pr_{j\|a}^{US}(s_{jht}=0)$	0.50	0.42	-0.08*	-0.11*
	(0.05)	(0.06)	(0.07)	(0.20)
$Pr_{j\|a}^{EU\|_{USTrials}}(s_{jht}=0)$	0.77	0.58	-0.19***	
	(0.18)	(0.05)	(0.19)	
$Pr_{j\|a}^{EU\|not}(s_{jht}=0)$	0.96	0.94	-0.02**	-0.17***
	(0.08)	(0.12)	(0.15)	(0.25)

$N^{US} = 317$, $N^{EU|_{USTrials}} = 380$, and $N^{EU|not} = 1050$ product-month observations. Standard errors clustered by month $N_t = 114$ in parentheses.

If product introduction exhibited a slow diffusion of usage due to timing of response to marketing, sales, or distribution post-launch, then one would expect the same products in

the US, or other products in the EU, to exhibit a similar pattern. However, neither the US data nor the EU data for products not undergoing US trials exhibit a meaningful upward trend in panel (a) – the mean usage patterns in those cases are flat over time after product introduction. This suggests neither market-specific nor product-specific factors alone drive the increased usage over time of products in the EU undergoing US trials.

To further examine the learning hypothesis, panel (b) plots the standard deviation $SD_{j|a}\left[\widetilde{\ln(s_{jt}/s_{0t})}\right]$ across products against age. Standard models of learning predict that this statistic will decrease toward the population's true quality standard deviation as uncertainty is resolved. As with the mean, this second moment changes over time for the EU sample of products concurrently in US trials – decreasing as we would expect with learning – but it does not change for the US sample or EU sample of products not in US trials. Importantly, both EU samples have the same level of volatility upon EU introduction, suggesting there is a similar amount to be learned about products regardless of US trial status. However, only those stents in US trials exhibit evidence of learning.

Finally, panel (c) shows how usage at the hospital level evolves with age as measured by the proportion of zero usage observations at the product-hospital level $E_{j|a}\left[\widetilde{Pr(s_{jht}=0)}\right]$. Similar to the aggregate results, the EU sample undergoing US trials begins with slightly more hospitals using each product on average, and this proportion grows with age, whereas it stays flat for products not undergoing trials.

The results above highlight that more information, on average, increases a product's share. However, information generated in a trial might not be positive and lead to a stent's share to decline. In Appendix B.1 we show some example paths for individual products and discuss one case in detail. CoStar was a new stent technology that was acquired by Johnson and Johnson prior to running the FDA pivotal trial. While early small sample evidence was promising, the final trial results on the full sample showed that the device was not as effective as other existing stents. The impact of that information caused CoStar's EU share to tumble. These results are also consistent with the notion, explored more systematically in Section 3.2, that device manufactures do not precisely know the efficacy of their device prior to running a large clinical trial of the type required by the FDA.

3.2 Robustness and Alternative Explanations

Placebo Test: PTCA Balloons. One alternative explanation for the above findings is that the set of manufacturers/products that undergo US trials promote their products differently than other products in the EU, and they may also market the same products upon US introduction differently. While we believe the evidence on decreasing variance and on the same products upon US launch make this unlikely, it is not impossible. To further

explore this possibility, we perform a placebo test using percutaneous transcatheter coronary artery (PTCA) balloons, which are FDA Class II devices and thus face similar regulatory requirements in both the EU and US. Thus, PTCA should not display the differential signs of learning we document for stents if our proposed mechanism is true. The results in Appendix B.2.1 show that we do see more total entry in the EU (presumably due to pre-existing complementary sales and distribution assets in the US for some manufacturers and not others); but the differences in amount of entry are smaller than in stents, there is no gap in time of entry on average, and usage patterns with age show no evidence of learning.

Alternative Explanation: Observational Learning with Different Initial Sample Size. Another potential explanation for the results in Figure 2 is that there is learning in the EU sample undergoing US trials, but this is learning is observational. The difference between the patterns in the two samples is then plausibly driven by the fact that those stents undergoing US trials enter with higher usage levels, which generate sufficient sample sizes for observational learning to occur, whereas the EU sample not undergoing trials contains too many products that do not gain enough early traction to enable learning.

We examine this hypothesis by reformulating the same figures and tests for a set of products with overlapping support on initial values of $\ln(s_{jt}/s_{0t})$ at $a_j = 1$, so they all have similar chances to generate early observational learning. The pattern in Appendix Figure A5 is essentially identical to that in Figure 2, suggesting that our results are not driven by selection on initial quality/usage levels.[25]

Alternative Explanation: Asymmetric Information and Signaling. Another potential explanation that could rationalize Figure 2 is manufacturer signaling. Under this hypothesis, after the release of EU trial data, manufacturers retain a sufficiently large degree of private information about expected product quality, so that undertaking costly US trials is a credible signal of expected product quality to physicians. To produce the observed data patterns, such a model also needs to include some combination of slow signal diffusion across hospitals and/or increasing signal strength as a trial continues. We explore this hypothesis by looking more closely at the shapes of the distribution of $\widetilde{\ln(s_{jt}/s_{0t})|_a}$ with age.

Appendix Figure A6 shows the evolution with age of different quantiles of the $\widetilde{\ln(s_{jt}/s_{0t})|_a}$ distribution. Under a model where manufacturers and physicians are similarly informed about quality after the release of trials for EU entry, and then learn similarly as data from US trials is released, the distribution of product quality estimates should converge symmetrically

[25]For this matched sample, selection into US trials must be based on level shifts in expected US profit due to the fact that those products that enter the US all have pre-existing complementary assets for sales and distribution (while those that don't enter do not). This is consistent with the challenges firms such as Biotronik have faced in developing US sales forces. See, "Tipping the Odds for a Maker of Heart Implants," *New York Times*, April 2, 2011.

to the true product quality distribution. In an asymmetric information setting, consumers do not receive direct information about quality, but instead infer quality must be above some threshold if a manufacturer is willing to continue with costly testing (see Appendix B.2.3 for more on this intuition). Learning in this way would cause the lower tail of the distribution for product in US trials to become truncated. In the Figure, the 25 and 75 percentiles appear to move symmetrically towards the median as information arrives. Below the figure, we present relevant test statistics. The change in the skewness of the distribution and the change in the ratio of the 75th-50th percentile to the 50th-25th are both insignificant.

Exploring Other US/EU Differences. We consider the evidence comparing the two samples within the EU to be the strongest regarding the risk-access tradeoff, and so our estimation and welfare analysis moving forward will focus on the EU sample only. However, we still find the comparison between the US and EU informative in considering the broader policy environment and the extent to which results from the EU sample can be extrapolated to consider US policy.

We have argued that historical political circumstances have led to greater testing requirements in the US than in the EU, and that the cost of these different testing requirements have led to more and earlier entry in the EU. Further, we have presented evidence from EU usage patterns that this differential testing has led to different amounts of information generation, and that the market values the resulting decreased uncertainty of products with more information. In theory, these differences in entry and usage patterns could be confounded with other differences in disease incidence, preferences for angioplasty and stents, or variation in price setting regimes between the US and EU over time. However, all the evidence that we have been able to gather (detailed in Appendix B.3 and summarized here) indicates that the patterns in the data described above are unlikely to be explained by other cross-region differences. Rates of ischemic heart disease, hospital diagnostic procedures, and prevalence of angioplasty with stenting are all similar between US and EU. Willingness-to-pay for new technology and prices tend to follow similar trends, but are on level lower in the EU, making the US a more attractive entry target, all else equal, and pushing in the opposite direction of the entry levels observed.

3.3 Summary of the Evidence

Our reading of the totality of the evidence we have assembled from stent entry and usage patterns aligns most closely with a model in which there is uncertainty about new product performance, learning occurs symmetrically to market players over time, and risk-averse decision makers factor uncertainty about quality into their product choice. The results imply

that there is significant learning from US clinical trials but very little learning observationally in the marketplace. This second finding is also consistent with institutional details regarding the lack of clinical follow-ups and systematic data collection on device clinical performance after market entry, which itself is part of the current policy debate.

We examine alternative plausible explanations, and no other model seems to fit the full set of patterns in the data. Specifically, the patterns we observe are not consistent with differential marketing/diffusion, differential demand side factors, differential prices and lags in reimbursement determination, selection into testing based on uncertainty, or residual asymmetric information (post EU testing) between manufacturers and regulators/consumers.

4 Model of Demand and Supply with Uncertainty and Regulation

In this section, we specify an empirical model of cardiac stent demand and supply. This framework incorporates the important institutional details of the industry and is flexible enough to capture the dynamic patterns in the data we documented in the previous section. The model allows us to: (1) decompose the various drivers of product utilization in a way that we could not in the analysis in Section 3; (2) translate the patterns in the data into measures of welfare; and (3) explore equilibrium outcomes under counterfactual scenarios related to the current policy debates in medical device regulation.

We begin our description of the model by characterizing the players, the timing and the information structure of the game. We then detail our demand model where consumers (physicians) have heterogeneous preferences over stent design characteristics and face uncertainty over the stent's performance. Physicians are able to learn about stent performance from clinical trials as well as market and hospital specific experience. We then turn to the supply side, where we develop a simple model of device manufacturers' behavior that enables us to consider the welfare impact of different regulatory policies. The section concludes by describing the role of the regulator.

4.1 Players, Timing and Information

1. There are two exogenously determined types of medical device manufacturers: $UStrial$ firms with sunk distribution networks in the EU and US and $notUStrial$ firms with an EU-only sunk distribution network. A sunk distribution network means that the marginal fixed cost of introducing a new product is given only by the cost of satisfying the regulatory approval process.[26] In each period t, there is a positive, exogenous

[26]There are three key implications of assuming no additional costs of product introduction beyond testing costs. First, the set of product qualities we estimate from the products who enter in the data $\{Q_j\}$ provides

probability that each manufacturer will innovate and produce a new device.

2. The mean performance of a new implanted device j is given by $Q_j^* \sim F_t^{UStrial}(Q)$. In order to keep the model tractable, we follow the consumer learning literature (Erdem and Keane 1996) in assuming this distribution is normal $F_t^{UStrial}(Q) := N(\mu_{Q_t^{UStrial}}, \sigma^2_{Q^{UStrial}})$. The dependence of F on t allows for the technology to evolve over time, and F is indexed by manufacturer type, allowing for different prior beliefs regarding the quality distribution for different manufacturer types.[27]

3. For each product, prior to the initial period of EU entry, we assume the firm receives a noisy but unbiased i.i.d. signal via product testing, $A_j^{EU} = Q_j^* + \nu_j^{EU}$ where $\nu_j^{EU} \sim N(0, \sigma^2_{EU})$. We assume this testing is costless (at the margin) to the firm given its infrastructure in place for ongoing research and development, and the results provide sufficient information to satisfy EU regulatory requirements. We also assume the resulting posterior information set of the firm $\mathcal{I}_j^1 := (Q_j^1, \sigma_j^1)$ is revealed to the EU regulators via the approval process and to physician consumers upon EU launch.

4. A *UStrials* firm will begin US trials for a stent if expected discounted lifetime profits (with a monthly discount premium of ζ) given the required US trial length T_{US}^c:[28]

$$E\left[\Pi_j(T_{US}^c)|\,\mathcal{I}^1\right] = E\left[\sum_{t=\underline{t}_j+T_{US}^c}^{\bar{t}_j+T_{US}^c}\left(\sum_h q_{jht}(p_{jht}-mc_j)\right)^{1-\zeta(t-\underline{t}_j-T_{US}^c)}\middle|\,\mathcal{I}^1\right] \quad (1)$$

an appropriate estimate of the true distribution of product qualities $F(Q)$. Second, the set of firms who enter in the data (under additional assumptions in Section 4.3) should be a superset of the firms who would enter in a counterfactual equilibrium with more restrictive testing requirements. Finally, it implies that no firm will decide not to enter after learning "bad news" from further testing. Although this is a strong assumption, it seems to be supported by the fact that 10 percent of EU products in the data generate lifetime profits of $1.3M or less, and most entries are from existing companies. In addition, our analyses show that this tail of lower profit products is marginal in its welfare effects, so to the extent that there are some marginal firms that might for some reason enter under more restrictive entry policies, it seems unlikely that they would meaningfully affect our analysis.

[27] Our GMM estimation approach will recover each Q_j^* without parametric restrictions on the distribution, and the results are reasonably close to normally distributed. We are limited in our ability to allow σ_Q to vary as estimation of this parameter relies on pooling the estimated Q_j^*.

[28] The expected profits for the entry/testing decision are computed using the estimated quality after EU testing, $\mathcal{I}_j^1 = (Q_j^1, \sigma_j^1)$, and taking the expectation over any subsequent learning that will occur. We further assume that products will exit after the fixed number of periods for which we observe them purchased in the data. Expected profits are a sum (discounted at one percent per month in our primary specification) of the profits accrued during this time $[\underline{t}_j + T_{US}^c, \bar{t}_j + T_{US}^c]$ in the market. Quantities q_{jht} and prices p_{jht} are equilibrium outcomes and a function of all firms' beliefs and actions as well as consumer beliefs. We assume perfect foresight among all players regarding the set of products completing EU testing, their timing of completing this testing, and their quality estimates at that point, so that $\mathcal{I}^1 = \{\mathcal{I}_j^1\}_{j\in\mathcal{J}}$. Expectations are taken over subsequent learning for own and competitor products from this point.

exceed the trial costs.

5. In subsequent periods $t = 1, 2, ...,$ prices p_{jht} are set, quantities q_{jht} are realized via consumption decisions, and surplus is accrued. Then signals are observed and beliefs are updated before actions are taken the following period. Letting age $a := t - \underline{t}_j$ denote the time in months since product j was introduced into the EU, signal A_{jha} received by hospital h is given by:

$$A_{jha} = Q_j^* + \nu_{ja} + \tilde{\nu}_{jha} \text{ where } \begin{cases} \nu_{ja} \sim N(0, (1-\gamma)\sigma^2_{A^c}), & \tilde{\nu}_{jha} \sim N(0, \gamma\sigma^2_{A^c}) \quad \text{if in clinical trials} \\ \nu_{ja} \sim N(0, (1-\gamma)\sigma^2_{A}), & \tilde{\nu}_{jha} \sim N(0, \gamma\sigma^2_{A}) \quad \text{if not} \end{cases} \tag{2}$$

where σ_A and σ_{A^c} measure the noise of signals generated by market usage and clinical trials, respectively.[29] The parameter $\gamma \in [0, 1]$ allows hospital learning to occur as weighted combination of market and hospital specific information ($\gamma = 0$ corresponding to perfect correlation in signals across hospitals; $\gamma = 1$ corresponding to completely independent signals).

Given these signals, beliefs about product quality are updated via Bayes' rule, resulting in posterior beliefs distributed $N(Q_{jha+1}, \sigma^2_{jha+1})$ where

$$Q_{jha+1} = \frac{\sigma^2_{jha}}{\sigma^2_{jha} + \sigma^2_{A_{jha+1}}} A_{jha+1} + \frac{\sigma^2_{A_{jha+1}}}{\sigma^2_{jha} + \sigma^2_{A_{jha+1}}} Q_{jha} \quad ; \quad \sigma^2_{jha+1} = \frac{\sigma^2_{A_{jha+1}}}{\sigma^2_{jha} + \sigma^2_{A_{jha+1}}} \sigma^2_{jha}. \tag{3}$$

Next, we describe how the model translates these beliefs into stent demand.

4.2 Demand and Surplus

We now turn to characterizing physician preferences over stents given their information set. Let $h(v_{iht}, x_j, Q_j^*)$ be the perfect information, ex post health state for a given individual i from an implanted stent j at hospital h in period t. Patient/physician characteristics are denoted by the vector v_{iht}, x_j is a vector of observable stent characteristics (e.g. bare metal, drug eluting) that affect its suitability for patient i, and Q_j^* the stent's true mean performance. Physicians have constant absolute risk aversion (CARA) preferences that incorporate the patient's health as well the cost of the device to the hospital, p_{jht}: $u_{ijht} = -\frac{1}{\rho} \exp\left(-\rho(h(v_{iht}, x_j, Q_j^*) - \theta^p p_{jht})\right)$, where $\rho = -u''(\cdot)/u'(\cdot)$ is the coefficient of absolute

[29] We assume that information release from a clinical trial accrues to the market with a consistent signal each month. This fits with the regular release of interim results at major meetings and in journal articles and subsequent further diffusion via word of mouth. We have examined and found no evidence of a discrete demand response in the EU upon US trial completion or FDA approval.

risk aversion.[30]

Physicians choose from the set of available stents at a point in time, $\mathcal{J}_t$, including the option of not implanting a stent, which has utility normalized to zero.[31] The true stent clinical performance is unobserved at time of implantation, so physicians must make their decisions based on their current information set, $\mathcal{I}_{ht} := (\{Q_{jht}\}_{\mathcal{J}_t}, \{\sigma_{jt}\}_{\mathcal{J}_t})$, which summarizes the expected performance, and uncertainty about that performance, for all available stents.[32] In this framework, "ex post regret" occurs any time a patient receives a stent that results in lower utility than the stent she would receive under perfect information. "Ex post harm" occurs when a patient receives a stent that results in lower utility than the outside good, $h(v_{iht}, x_j, Q_j^*) < 0$. Thus, a regulatory approach that allows entry of a larger set of products $\mathcal{J}_t$ can make consumers worse off by increasing the likelihood of ex post regret and harm if that set includes products that perform below average and/or have high uncertainty in their expected performance.

For each patient, physicians choose the stent that maximizes ex ante expected utility, given their information set, $E[u_{ijht}|\mathcal{I}_{ht}] = \int u_{ijht} dN(Q_{jht}, \sigma_{jt}^2)$. The normality of the distribution of beliefs over Q implies this problem is equivalent to maximization of the mean-variance representation $U_{ijht} = E[h(v_{iht}, x_j, Q_j^*)|\mathcal{I}_{ht}] - \frac{\rho}{2}\sigma_{jt}^2 - \theta^p p_{jht}$, and we follow the consumer learning literature (e.g. see the review in Ching et al. (2013)) in working with this representation directly.[33]

In order to take the model to the data, we parameterize $h(v_{iht}, x_j, Q_j^*) := Q_j^* + \xi_{jh} + \epsilon_{ijht}^g + (1 - \lambda^g)\epsilon_{ijht}$ so that:

$$U_{ijht} = Q_{jht} - \frac{\rho}{2}\sigma_{jt}^2 - \theta^p p_{jht} + \xi_{jh} + \epsilon_{ijht}^g + (1 - \lambda^g)\epsilon_{ijht} \tag{4}$$

where ξ_{jh} captures preference deviations of the physicians at a given hospital over product features that are known with certainty but unobserved to the econometrician. The deviations

[30] This closely follows the modeling of utility over health outcomes in the health insurance choice literature (Cardon and Hendel 2001; Handel 2013). The conceptual difference is that instead of choosing from insurance plans that affect ex post consumption over a pre-specified distribution of potential health states, our agents choose among products that each represent different distributions of potential ex post health states.

[31] Because our data consists of product usage, we do not directly observe the set of stents available at a given hospital. We proceed with the assumption that any hospital could potentially purchase any stent available in the market at that time.

[32] For a stent j and calendar time t, age a is implicit, so we suppress it for ease of notation, e.g. $E[Q_j^*|\mathcal{I}_{ht}] = Q_{jht} = Q_{jha}$ for the appropriate a.

[33] Although it is less frequently discussed in the consumer learning and health care contexts, a large literature in portfolio choice has documented that the outcome achieved by maximizing the mean-variance representation often provides an excellent approximation to the optimal outcome for a consumer who discounts uncertainty, even in cases where the underlying distributions are not normal. In Appendix D.2 we explore less parametric specifications and find the Normal-Normal learning model provides a parsimonious approximation that fits the data well.

are distributed according to the type of the device, $\xi_{jh} \sim N(0, \sigma_H^g)$ with $g \in \{bms, des\}$. Larger values of the standard deviations imply greater variation in tastes across hospitals. The iid error term, $\epsilon_{ijht}^g + (1 - \lambda^g)\epsilon_{ijht}$, captures the preference deviation relative to the population average of physician/patient i for device j with characteristic g. This is a random coefficients utility model where the random coefficients are on indicators for whether the stent is drug-eluting or bare-metal, which is equivalent to a nested logit specification under the assumptions in Cardell (1997) where $\epsilon_{ijht}^g + (1 - \lambda^g)\epsilon_{ijht}$ is distributed generalized extreme value with mean zero, scale parameter 1, and $0 \leq \lambda^g < 1$.[34]

We further assume that physicians maximize myopically, treating each patient as she arrives and ignoring the impact of the current stent choice on future stent choices. Integrating over the distribution of patient/physician i heterogeneity then yields the familiar nested logit closed forms for product-hospital-month specific: choice probabilities, $cp_{jht} := Pr[U_{ijht} > U_{ikht}, \forall k \in \mathcal{J}_t]$; elasticities with respect to price $\eta_{jkht} := \frac{\partial cp_{jht}}{\partial p_{kht}} \frac{p_{kht}}{cp_{jht}}$; and ex-ante expected consumer surplus (relative to the outside option) $CS_{ht}(\mathcal{J}_t, \mathcal{I}_{ht})$.[35] Combined with the number of patients receiving diagnostic procedures, M_{ht}, these map directly into quantities, substitution patterns, and welfare that enter supplier and regulator decisions.

4.3 Supply

A key objective of our paper is to better understand the equilibrium impact of changing regulations on product performance information generation. Therefore, we need to model the impact of changing information requirements on the supply-side behavior of stent manufacturers. Increasing information requirements for device approval is costly as it requires larger clinical trials and trial costs are increasing in the number of trial subjects. A larger trial also means a longer trial as it takes time to recruit patients. Therefore, increasing information requirements for device manufacturers can decrease and delay the entry of new products. The supply-side model should allow for the possibility that increased information requirements impact the firms' entry decisions and, conditional on deciding to enter, the timing of product introduction.

A natural supply-side approach to employ is to estimate the parameters of a dynamic entry/exit game and use the model to solve for new counterfactual equilibria. However, a full model of dynamic entry and exit posses conceptual and computational challenges as we

[34] When the nesting parameter, $\lambda^g = 0 \ \forall g$, this is the standard multinomial logit model. As $\lambda^g \to 1$, products within the category become closer substitutes to each other than to goods outside the category. We have experimented with allowing for finer nest classifications for some of the specialty stents present in the EU such as inert metal stents and stents designed specifically for bifurcated lesions, but these categories are too sparsely used (for context, the total market share of bifurcated stents is an order of magnitude lower than the average BMS) to identify their nesting parameters with any reasonable amount of precision.

[35] See Appendix C.1 for the explicit formulas.

have a large and continuous state space, requiring approximations of the type explored in recent papers such as Ifrach and Weintraub (2017). Conditional on entry, the buyer-supplier network formation problem itself is a complex problem at the frontier of recent research (e.g. Lee and Fong 2013; Grennan and Swanson 2018; Ho and Lee 2019). Because of these challenges, we use our demand and learning models to analyze outcomes under two extreme cases on supply-side behavior. These cases are simple to compute yet informative about the impact of different policies.

Specifically, we seek to understand the impact of policies requiring (weakly) greater clinical evidence than under the current EU requirements. We consider such policies as multiples of the current US trials we observe, and denote them by additional months of trials required T^c (again, we focus on clinical trial length which maps into sample size), and the fixed cost of running those trials by $FC = \chi T^c$.[36] In our counterfactuals in 6.3, we discuss how this thought experiment might apply to considering both EU and US policy. For each T^c, the two cases we examine are:

More Entry (M) case, $\mathcal{J}^M$: Assume firms enter as if there is no direct cost of longer trials, $\chi = 0$, iff: $E\left[\Pi_j(T^c)|\mathcal{I}^1\right] > 0$

Less Entry (L) case, $\mathcal{J}^L$: Assume firms enter as if trials cost $\chi_j = \$1.6M$ per month,[37] but also under the belief that other firms enter as if there is no cost to trials, $\chi_{-j} = 0$, iff: $E\left[\Pi_j(T^c)|\mathcal{I}^1, \chi_{-j} = 0\right] > \chi_j T^c$

Under the assumptions discussed in Section 4.1, these two cases provide reasonable approximations to a superset (M) and subset (L) of products expected to be present in equilibrium in the market at any point in time $\mathcal{J}_t$, given a regulatory policy T^c.[38] These cases also help decompose the welfare consequences of more pre-market clinical testing into two components: 1) delayed entry due to increased trial length requirements; and 2) decreased entry due to higher trial costs.

[36] In reality, there are a mix of trial costs, some up front, some ongoing, and some at the end of a trial. We do not have detailed estimates of this break-down, and so we make the simplifying assumption that all are accrued up front. To the extent that trials can be abandoned at significant cost savings, this may slightly overstate the expected cost of trials. But we have learned of very few trials being abandoned. Also, this would primarily affect lower expected quality products, which contribute very little to our welfare analysis.

[37] \$1.6M per month from the survey by Makower et al. (2010) of the costs of US trials.

[38] In a prior working paper version of this manuscript (Grennan and Town 2018) we referred to these cases as "bounds" on firm entry and used them to partially identify counterfactual surplus measures and optimal policies. The challenge with that approach is that one cannot (to our knowledge) prove these bounds theoretically for general demand and supply, even under the further assumptions we make regarding timing and information. Deviations from our timing and information assumptions could cause further violations of these bounds. For example, anything that would lead to entry of firms who do not enter under current policy could violate M as an upper bound, and M being an upper bound is essential to L being a lower bound. Thus we prefer to refer to these as helpful extreme "cases", moving away from the "bounds" terminology to avoid confusion regarding the degree of theoretical precision we claim.

The M case intuitively approximates a superset of products that might enter because, by assumption, the same products enter under any $T^c \geq 0$ as under current EU policy $T^c = 0$. Thus, the impact of increasing trial length on market structure under M weighs the benefit of increased potential learning through generating performance information via trials versus the cost of delayed access to the newest technologies.

The L case approximates a subset of products that might enter. The intuition for this claim is that while a focal product responds to its own entry costs, under L it acts under the belief that other products have zero additional costs, i.e. it calculates its expected profits as if *all other firms enter*. Because expected profits to a focal firm facing all other firms will tend to be lower than it would face in the full equilibrium where some firms do not enter due to the fixed cost of testing, computing L should result in less firms entering than under the actual equilibrium.

4.4 Modeling the Regulator

We treat the regulator as an agent that determines the device approval policy by choosing a mean performance threshold treatment effect that increases expected health to $\underline{h}$ and an associated significance level α over that treatment effect. After the clinical trial has been completed, and if the data indicate that $Pr[h(Q_j^*|trial) > \underline{h}] > 1-\alpha$, the regulator will then approve the product.[39] The regulator also determines the power of the test which, combined with the choice $(\alpha, \underline{h})$ and the underlying quality distribution F, dictates an optimal trial size $N^*(\underline{h}, \alpha)$, which (given a constant arrival rate of suitable subjects ϕ per month) implies a clinical trial length $T^{c*}(\underline{h}, \alpha) = N^*/\phi$ in months. This is why policy discussions often simply refer to the "length of trials" to capture the regulatory policy threshold and its temporal and monetary cost. To correspond with policy discussions and to simplify the analysis without losing much generality, we treat the regulator as choosing T^c with the understanding that trial length maps into sample size and, in turn, the statistical properties of the trial data.

Regulatory policy affects social surplus through two distinct channels: uncertainty and access to new products. Uncertainty is affected in that every ϕ subjects generate a signal, A_j, so a longer time in clinical trials provides information, which decreases uncertainty and brings market participants' estimates of a product's quality closer to its true quality. Access is affected directly because an additional month in trials delays consumer access to new stents by a month. Access is also affected indirectly because trials are costly – an additional month in clinical trials raises fixed costs of entry by χ, with the total costs $FC := \chi T^c$. In

[39] As noted above, given that trials are costly, the model implies that actual rejections will be rare because if the information from the trials indicates that the likelihood of device approval is low, the manufacturer will terminate the trial before its completion. The CoStar case discussed above and in Appendix B.1 provides such an example.

our counterfactual policy analyses in Section 6, we consider potential regulatory objectives based on consumer or total surplus.[40]

5 Model Estimation, Identification, and Results

5.1 Estimation and Identification

We estimate the parameters of the demand model using detailed data on prices paid and quantities of stents implanted at product-hospital-month level. We use only the EU sample in the estimation. This approach leverages the fact that the data contains: (1) variation in the information regimes across products, and (2) within the subset of products undergoing US trials, variation in the amount of information generated over time. The variation in generated information spans the range of the information gap between the EU and US policies, which is the primary range of interest in current policy debates. We implement the estimation via a generalized method of moments algorithm as detailed in Appendix D and summarized here.

A significant challenge faced in taking the model to the data is the EU choice set is large relative to the number of choice instances in a hospital-month. As a consequence, there are a large number of zero market shares at the product-hospital-month level. This issue is relatively common in fine-grained data and has been a topic of recent concern in the industrial organization and marketing literatures (Gandhi et al. 2013). Quan and Williams (2017) (henceforth QW) develop a novel solution that involves matching a combination of micro (in our case product-hospital-month) and aggregate (product-month) moments to estimate the distribution of preference heterogeneity across markets (in our case hospital-months) while explicitly allowing for zeros due to sampling variation. Our estimation strategy combines the intuition and approach of QW with our learning model. Their insight is aggregation across markets can generate enough purchase instances that the negligible sampling variation assumption can be restored to estimate product-specific utility parameters, while moments at the disaggregate level can still be included to estimate the distribution of heterogeneity in these parameters across hospitals.

[40] Appendix C.2 provides an explicit closed form solution for a simplified case that helps to clarify the regulator's tradeoff between access and uncertainty in requiring longer trials (more information):

$$TS_t(T^c+1) - TS_t(T^c) = \underbrace{\frac{\rho}{2}\left(\sigma^2(T^c) - \sigma^2(T^c+1)\right)}_{\text{gain from decreased risk}} - \underbrace{\ln\left(\frac{\sum_{j\in\mathcal{J}_{t+1}(T^c+1)} e^{Q_{jt}}}{\sum_{j\in\mathcal{J}_t(T^c+1)} e^{Q_{jt}}}\right)}_{\text{gain from tech change/entry}}$$

Following this logic, we rewrite utility to the mean consumer δ_{jht} in terms of aggregate and hospital-specific portions:

$$\delta_{jht} = \underbrace{Q_{jt} - \frac{\rho}{2}\sigma_{jt}^2}_{\delta_{jt}} - \theta^p p_{jht} + \xi_{jh} + \widetilde{Q}_{jht} \tag{5}$$

where $Q_{jt} := E_h[Q_{jht}]$ is the expected product quality estimate across hospitals, $\widetilde{Q}_{jht} := Q_{jht} - Q_{jt}$ is the product-hospital-month specific deviation from that aggregate expectation, and ξ_{jh} is already defined as a deviation with mean zero across hospitals. Following QW, we appeal to the law of large numbers in the number of hospitals H and (letting M_h denote the number of patients treated at h) set observed aggregate market shares equal to aggregated choice probabilities $s_{jt} := \sum_h \frac{M_h}{\sum_h M_h} s_{jht} = \sum_h \frac{M_h}{\sum_h M_h} c p_{jht}$, inverting the system to obtain:

$$\delta_{jt}(\mathbf{s}_t; \lambda, \sigma) = \ln(s_{jt}/s_{0t}) - \lambda^g \ln(s_{jgt}) - (1-\lambda^g)\ln(R(\sigma_g)) + \theta^p \sum_h \frac{M_h}{\sum_h M_h} p_{jht} \tag{6}$$

where $R(\sigma_g)$ is an adjustment to the mean utility accounting for aggregating over hospital heterogeneity

$$R(\sigma_g) := E_{j|g}\left[\exp\left\{\frac{\tilde{\delta}_{jht}}{1-\lambda^g}\right\}\right] = \exp\left\{\frac{\sigma_g^2 + \gamma \frac{(a_j - t_j^c)/\sigma_A^2 + t_j^c/\sigma_{Ac}^2}{1/\sigma_Q^2 + 1/\sigma_{EU}^2 + (a_j - t_j^c)/\sigma_A^2 + t_j^c/\sigma_{Ac}^2}\sigma_{jt}^2}{2(1-\lambda^g)^2}\right\} \tag{7}$$

where the expectation attains from the moment generating function of the normal distribution, and t_j^c denotes the cumulative time spent in clinical trials by device j by time t. $R(\cdot)$ follows directly from QW, which requires applying a law of large numbers in the number of products per category J_g. The only difference is that in our model, heterogeneity across hospitals at any point in time reflects both fixed preference heterogeneity (represented by σ_g^2, as in QW) and learning heterogeneity (represented by $\gamma \frac{(a_j - t_j^c)/\sigma_A^2 + t_j^c/\sigma_{Ac}^2}{1/\sigma_Q^2 + 1/\sigma_{EU}^2 + (a_j - t_j^c)/\sigma_A^2 + t_j^c/\sigma_{Ac}^2}\sigma_{jt}^2$, the fraction of uncertainty that is due to hospital-specific signals).

Aggregate moments – means: From (6) and (5), we form the standard linear moments:

$$\xi_{jt} = \ln(s_{jt}/s_{0t}) - \lambda^{g_j}\ln(s_{jg_jt}) - (1-\lambda^{g_j})\ln(R(\sigma_{g_j})) + \theta^p \sum_h \frac{M_h}{\sum_h M_h} p_{jht} - Q_j^* - \frac{\rho}{2}\sigma_{jt}^2 \tag{8}$$

where the econometric residual is the difference between the aggregate estimated product quality and the true product quality $\xi_{jt} := Q_{jt} - Q_j^*$, where $E_j[\xi_{jt}] = 0$ by the unbiased learning. We interact these residuals with a set of instruments Z^d which includes: product

fixed effects to identify product qualities; lagged mean prices $\sum_h \frac{M_h}{\sum_h M_h} p_{jht-1}$ to identify the price coefficient (following Grennan (2013) in exploiting the fact that changes in "stale" long-term contracts help identify demand);[41] a polynomial in the size of the within-group choice set $[J_{gt}, J_{gt}^2]$ (following Berry and Waldfogel (1999) with a growing choice set over time directly affecting within- vs. out-of-group substitution) to identify the nested logit substitution parameters λ; and a set of age dummy variables interacted with whether the product is currently undergoing clinical trials to jointly identify $-\frac{\rho}{2}\sigma_{jt}^2$. Further information is required to separately estimate learning $\sigma_{jt}^2(\sigma_A)$ and heterogeneity across hospitals (γ, σ^H).[42]

Aggregate moments – variances: The learning and demand model additionally implies that the variance of the prediction errors is tightly related to the aggregate uncertainty about product quality

$$E_j[\xi_{jt}^2 | (a_{jt}, t_{jt}^c) = (a, t^c)] = \sigma_{jt}^2(a, t^c) \ . \tag{9}$$

Recall that $\sigma_{jt}^2(a, t^c) = \left(1/\sigma_Q^2 + 1/\sigma_{EU}^2 + (a - t^c)/\sigma_A^2 + t^c/\sigma_{A^c}^2\right)^{-1}$, and in particular, note that this second moment is independent of the risk aversion parameter, ρ. Thus, variation in usage identifies the learning signal parameters σ_A as age a and time in trials t^c vary.

We further impose a consistency assumption that the variance of the estimated product quality parameters equal the prior belief about the distribution of product qualities that enter the EU market $Var_j(Q_j^*) = \sigma_Q^2$. Combined with the variance moments (9) in the first period a product is introduced (when $a_{jt} = 0, t_{jt}^c = 0$), this also identifies the information provided by EU trials, σ_{EU} because $\sigma_{jt}^2(a = 0, t^c = 0) = \left(1/\sigma_Q^2 + 1/\sigma_{EU}^2\right)^{-1}$.

These two sets of aggregate moments clarify how learning is identified by the degree to which the variance in product-specific quality estimates decreases over time. Risk aversion is then identified by how choice probabilities increase (or don't) as learning decreases uncertainty.[43] This relates directly back to the reduced form evidence in Figure 2. For products in

[41] Grennan (2013) estimates the model using quasi-differences $\xi_{jt} - \rho\xi_{jt-1}$, appealing to changes in information over time. We account for that in part by controlling for the evolution of uncertainty directly, but we could use quasi-differences in addition. Our attempts to do so resulted in difficulty converging to estimates that fit the data well, presumably due to extracting too much of the signal from the data.

[42] A simple and semi-parametric way to estimate Equation (6) would be to regress the inclusive shares $\ln(s_{jt}/s_{0t})$ on product and age fixed effects interacted with whether a product is in clinical trials or not to allow for differential learning rates. In this research design, the age fixed effects – paired with the exogenous variation in learning – would then capture the combined treatment effect of risk aversion and learning on utility. However, because we are interested in questions that involve market reactions to different learning rates and levels of uncertainty, we need to add structure via the learning model to disentangle these forces. Comparison to the fixed-effect model in Appendix D.2 provides a useful benchmark for assessing the fit of the more parsimonious and parametric learning model, which we consider quite good.

[43] The typical discussion of identification of learning versus risk aversion in the related literature estimating similar models from aggregate market share data (see Ching et al. (2013) for an overview) notes, correctly, that in the context of the Normal-Normal model, the two are in principle separately identified by the shape and level of the first moment over time. As those models are almost always estimated via maximum likelihood

trials, the variance decreases with age, identifying learning. As this variation decreases, the mean inclusive share increases, identifying risk-aversion. Observational learning is identified by the dynamic behavior of share volatility for products not in a US trial. These parameters are identified using the within-product variation, conditional on the product fixed effects (whose parameters provide estimates of the product qualities Q_j^*).

Micro-moments: The parameters left to be identified are those measuring the dispersion in hospital preferences σ^H and the extent to which learning signals differ across hospitals, γ. We follow the strategy developed in QW, adding hospital level micro-moments based on the probability of observing a zero market share for each product-hospital-month,

$$Pr\left[s_{jht}=0\right] = \left(1 - cp_{jht}(\sigma^H, \gamma)\right)^{M_{ht}}. \tag{10}$$

We match this probability to the data by simulating (over the distribution of ξ_{jh}) moments equating the empirical proportion of zeros to the model's predictions $\sum_h \mathbf{1}_{\{s_{jht}=0\}} = E_h\left[\left(1 - cp_{jht}(\sigma^H, \gamma)\right)^{M_{ht}}\right]$. The distribution of preference heterogeneity across hospitals σ_g^H is then identified by the extent to which large variance in ξ_{jh} is needed to match the zeros in the data for each product category, on average over time. The extent of hospital-specific learning, γ, is identified by how that proportion of zeros changes with learning as age and time in trials change.

5.2 Demand Parameter Estimates

The parameter estimates from the model are presented in Table 1. We focus on interpretation and validation of the estimates from our full, preferred model described in the previous section. Appendix D.2 presents further results with a less parametric learning model, simpler utility models nested within our preferred model, and alternative models of observational and hospital-specific learning / diffusion of information.

The demand estimates are sensible. Turning first to the utility parameters that capture physician preferences and substitution patterns. The parameter on price, θ^p, is negative and statistically significant indicating that demand is downward sloping but relatively insensitive to stent price. Both nesting parameters ($\lambda^{des}, \lambda^{bms}$) are also statistically significant and imply that products within the same nest are much closer substitutes than products in different nests. The estimated standard deviations ($\sigma_H^{des}, \sigma_H^{bms}$) of preferences across hospitals ξ_{jh} are both significant and economically meaningful. At nearly 0.2 logit utils, they are an order of magnitude larger in effect than the 0.03 util effect from a \$316 change in price (one standard

or bayesian methods, they implicitly use information from the second moment as well in estimation. Our use of GMM makes explicit the value of the second moment in identifying learning.

Table 1: Estimates of demand/learning model parameters

Preference/substitution parameters:					
θ^p (utils/\$)	λ^{des}	λ^{bms}	σ_H^{des}	σ_H^{bms}	$\rho \cdot \theta^p$ (1/\$)
0.10E-3	0.81	0.82	0.19	0.18	3.26E-3
(0.04E-3)	(0.02)	(0.01)	(0.04)	(0.02)	(1.47 E-3)
Learning process parameters:					
$\sigma_Q^{UStrials}$	σ_Q^{not}	$1/\sigma_{EU}^2$	$1/\sigma_{A^c}^2$	$1/\sigma_A^2$	γ_H
0.26	0.34	18.79	1.61	0.00	0.00
(0.01)	(0.02)	(2.75)	(0.67)	(0.23)	(0.10)

Estimates for demand model $\ln(s_{jt}/s_{0t}) = \lambda^{g_j} \ln(s_{j|_g ht}) + R(\sigma_H^{g_j}, \gamma_H) - \theta^p p_{jt} + Q_j^* - \frac{\rho}{2}\sigma_{jt}^2 + \xi_{jt}$ with separate nests for DES and BMS, and additional $E[\xi_{jt}^2]$ moments to identify learning, and $Pr[s_{jht} = 0]$ moments to identify heterogeneity in preferences and learning across hospitals. $N_{JHT} = 407,191$ product-hospital-months and $N_{JT} = 4,888$ product-months. Standard errors in parentheses, estimated via delete-10 jacknife, clustered by month ($N_T = 114$).

deviation and 26 percent of mean DES price). These results are all consistent with qualitative reports of strong physician brand preferences, the importance of DES/BMS-patient match, and estimates of coronary stent demand in other studies (Grennan 2013, 2014; Grennan and Swanson 2018).

We also estimate that physicians are risk averse in their selection of stents with a coefficient of absolute risk aversion of $\rho \cdot \theta^p = (3.26 \times 10^{-3}\$)^{-1}$. This estimate is within the range of estimates of risk aversion in well identified studies such as Cohen and Einav (2007).[44]

In addition to the uncertainty and risk aversion measures, the utility and learning models are linked through a rational expectations assumption on the distribution of product qualities $F(Q_j^*)$. Our demand model includes product fixed effects, and under our assumption of unbiased learning, the coefficients on these dummy variables provide consistent estimates for the true product quality for each product introduced to the EU market $\{Q_j^*\}$. These product quality estimates then provide a nonparametric estimate for $F(Q)$ (plotted in Appendix Figure A11), and rational expectations requires that consumers' priors about $F(Q)$ are consistent with this distribution.

In Figure 3, panel (a) we show that our revealed preference estimates of Q_j^* are correlated with the clinical quality measure target vessel revascularization (TVR) rate in the sample of products for which we were able to collect clinical trial data. This result provides support for the validity of our approach as the revealed preference estimates from our demand model

[44] As noted in Train (2015) (and more recently in Brown (2017) using a model very similar to ours), an ex post utility maximizing agent will also discount uncertainty when forced to make decisions over multiple uncertain options because of a "winners curse" phenomenon, even with risk neutral demand. Intuitively, in hospital-time markets where quality is overestimated ($Q_{jht} > Q_j^*$, the cases with ex post bad news) product j will also be used more that it should be (exactly because quality is overestimated), and conversely when quality is underestimated. This interaction means that increasing the second moment of beliefs decreases ex post welfare for risk neutral consumers. We report such an ex post loss number in Table 2, but we prefer the model with risk aversion and ex ante welfare measure for their link to the broader literatures on consumer learning and preferences over health states.

align with the clinical trial results.[45] We return to the product quality estimates themselves when we consider the role of technological change in generating gains from access to new products in Section 6, but we first discuss their role in the learning model.

The variation in product performance estimates are $\sigma_Q^{UStrials} = 0.26$ and $\sigma_Q^{not} = 0.34$. This uncertainty exceeds the magnitude of heterogeneity in preferences across hospitals suggesting that, without additional information, consumers selecting a new product for insertion face a non-trivial probability that it will perform worse than expected. These estimates imply greater device performance variation for EU-only devices. Also, consistent with the reduced form evidence (now controlling for a variety of other factors that influence demand), the estimates imply that there is significant learning in the EU from FDA clinical trials as $1/\sigma_{A^c}^2$ much greater than zero. The estimates also imply no experiential learning as $1/\sigma_A^2$ is very small and not significantly different from zero. The estimated precision of EU trials $1/\sigma_{EU}^2$ implies that the learning from trials pre-EU introduction is equivalent to almost 12 months of US trials. Finally, the estimates imply no hospital specific learning. The parameter γ_H is a precise zero, suggesting information accruing to the market is highly correlated across hospitals (ruling out correlation less than 0.8 with 95 percent confidence).

6 Technological Change, Uncertainty, and Optimal Trial Length in the Coronary Stent Market

With the parameters of the model estimated, we now turn to answering several policy relevant questions. Specifically, we use the model and the estimated parameters to: (1) calculate the size and source of technological change as new stents are introduced during our sample; (2) assess the role of information in affecting risk and resulting consumer usage patterns; and (3) estimate the optimal regulatory policy to balance the risk-access tradeoff under existing and alternative market and information environments.

6.1 Technological Change in the EU Coronary Stent Market 2004-2013

Not only does the rate of technological change in medical care have an important impact on aggregate welfare (Murphy and Topel 2006), it is also a key determinant of the optimal regulatory policy towards product information provision. The rate of technological change affects the value of access to the newest devices relative to those already available in the market. Typically, estimates of the value of a medical technology focus solely on measuring clinical outcomes and do not assess preferences and substitution patterns. Here we apply

[45] Here we focus on TVR, but the same pattern holds for major adverse cardiac events (MACE).

the tools that are standard in the industrial organization literature to assess welfare improvements associated with coronary stents over time. We compute the rate of technological change by calculating the ex post average treatment effect (ATE) for each stent, i.e. the mean surplus (relative to the outside option) of having stent j implanted on the average angiography patient: $ATE(Q_j^*) := \theta^{scale} \ln(1 + e^{Q_j^*})$.[46]

Figure 3: Clinical Quality, Revealed Quality, and Technological Change.

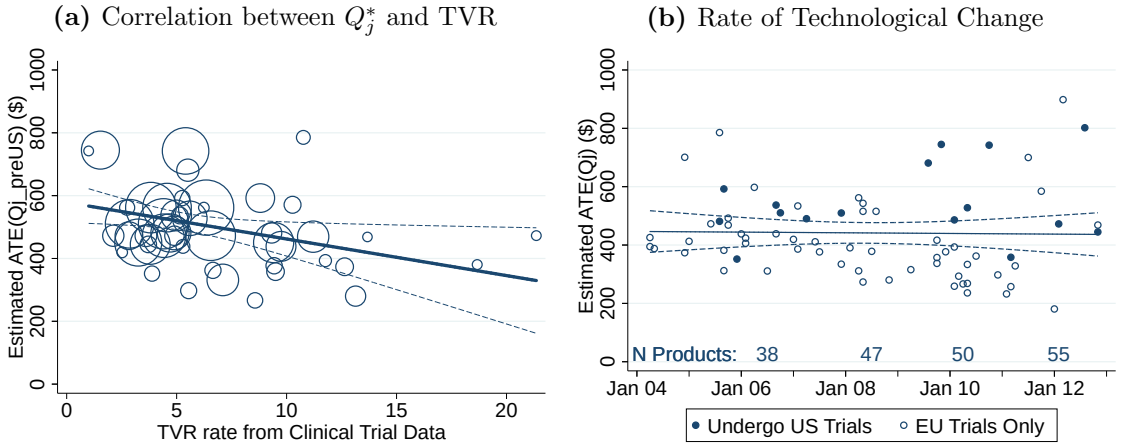


Figure 3, panel (b) presents these results, plotting the ATE for each product introduced against calendar time of the product introduction. During our sample period, the trend of mean product quality over time is flat.[47] However, the set of devices available grows from 38 to 55 over this same time period, which translates into a meaningful increase of 9.6 percent in the utility consumers receive from access to coronary stents.

This finding is salient for analysis of the optimal regulatory policy. If technological change is driven by increases in average product quality, the impact of changing costs of entry with changing regulatory standards of evidence will likely have a smaller impact on welfare than if the change is, as we find here, driven by increases in product variety. Niche products will by their nature have smaller market opportunities and thus find it more difficult to incur the

[46]Because we are concerned that they lack of price sensitivity we estimate may not accurately scale utils into dollars if physicians are imperfect agents for patients/hospitals, we choose θ^{scale} to normalize the total surplus per stenting procedure to $5,000, which is the approximate median of the estimated dollars in quality adjusted life years from the procedure relative to a coronary artery bypass graph surgery (a more invasive alternative to receiving angioplasty and a stent) among studies reported in the Cost Effectiveness Analysis Registry (https://research.tufts-nemc.org). Scaling into dollars using the standard approach of the inverse of the price coefficient $\frac{1}{\theta^p} = 10,482$, would approximately double all related consumer welfare estimates. This alternative scaling is only for translating welfare measures into dollars – we continue to use the estimated θ^p in quantity and elasticity calculations, as revealed preference indicates this is the level of price sensitivity that best fits the demand patterns in the data.

[47]This is likely in part due to increasing quality of alternative treatments, such as less-invasive and beating-heart CABG (Kalyanasundaram and Karlheinz 2014).

fixed cost of greater testing.

6.2 Uncertain Quality and Market Outcomes

Optimal regulatory policy should also take into account the potential welfare loss due to the risk that new products may not improve health much as expected. The magnitude of this uncertainty effect depends upon the mean and variance of expected quality levels across products as well as the amount of information consumers possess.

Table 2 explores the role of uncertainty in the market by using the demand model to calculate the percent of patients undergoing a diagnostic angiography who receive a stent relative the outside good $(1 - s_0)$, total surplus per stent $(\frac{TS}{1-s_0})$, and the expected ex post difference between the realized and expected utility from the chosen stent $(\sum_j E_i[Q_j^* - Q_{jt}|j = \arg\max_k U_{ikht}] = \sum_j \frac{s_{jt}(Q_j^* - Q_{jt})}{1-s_{0t}})$. Here we posit hypothetical markets where all products have uncertainty in their quality, varying from the unconditional uncertainty of the quality distribution σ_Q (if there were no testing/learning at all – the first column), to the estimated uncertainty upon first entering the EU $\sigma^1 = \sigma_{T^c=0}$ (after undergoing EU requirements – the second column), to varying lengths of US trials σ_{T^c}. In order to focus purely on the role of uncertainty, this is a partial equilibrium analysis as we do not consider firms' strategic responses to these different parameters via pricing or entry.

Table 2: The effect of uncertainty on number of stenting procedures, surplus per stent implanted, and expected ex post loss (all reported numbers take the average across all months in our period of study).

	σ_Q = 0.31	$\sigma_{T^c=0}$ = 0.19	$\sigma_{T^c=6}$ = 0.16	$\sigma_{T^c=12}$ = 0.14	$\sigma_{T^c=18}$ = 0.13	$\sigma_{T^c=24}$ = 0.12	$\sigma_{T^c=30}$ = 0.11
$1 - s_0$ (%)	12.5	24.0	26.4	27.9	29.0	29.7	30.3
	(2.5)	(1.4)	(1.3)	(1.3)	(1.3)	(1.4)	(1.4)
$\frac{TS}{1-s_0}$ ($)	5776	6103	6184	6238	6276	6304	6327
	(176)	(167)	(167)	(168)	(169)	(170)	(171)
$E[Q_j^* - Q_{jt}\|j^*]$ ($)	-1096	-560	-429	-348	-292	-252	-221
	(127)	(23)	(37)	(41)	(41)	(39)	(37)

Table 2 makes several important points. First, without any learning (and holding the strategies of the firms constant), the stent market would shrink significantly due to the large amount of performance uncertainty. This can be seen in the first column of the table in which the percentage of consumers having a stent implanted is about half that of the cases with testing. This implies that clinical testing and information gathering of the type done currently in the EU provides the necessary information to make this market operate.[48]

[48]We see this point as illustrative and potentially an underestimate of the value of EU testing. In addition to the partial equilibrium caveat applying to this entire table, the EU process may solve asymmetric

Second, modest increases in the information available to consumers generates significant improvements in welfare. Moving from a world in which there are no clinical trials to one in which there is EU testing plus an FDA clinical trial of 6 months leads to meaningful increases in the number of procedures performed and the surplus created, and decreases in expected ex post loss due to choice "mistakes" per procedure.[49] Increasing additional required FDA trial length beyond 6 months generates smaller increases, with the difference between 18 and 24 or 30 months of trials not statistically different from zero at typical thresholds.

We can also use this framework to calculate the impact of the beneficial information spillover from US testing to the EU. Comparing the estimated total surplus from the observed data and our model to the hypothetical $\sigma_{T^c=0}$ case, we calculate that if the US were to stop requiring additional testing beyond the EU levels, and the EU were to hold its current policy fixed, that EU total surplus would decrease by 6.4 percent.

Finally, these effects are driven by symmetric yet imperfect information, rather than the informational asymmetries which have been the central concern in much of the quality information literature. This suggests that in the case of regulating testing/disclosure, taking into account uncertainty and amount of information provided can be just as important as solving asymmetric information problems.

6.3 Optimal Clinical Testing Regulation

Next, we turn to the fundamental question that motivates this paper: In an industry where new products are developed with uncertain quality, what is the optimal amount of pre-market testing to require? To answer this question, we use our demand and supply model to calculate counterfactual equilibrium outcomes under different regulatory policies. We start with a baseline of current EU requirements, and we consider the effects of requiring incrementally longer trials by T^c months (where recruitment timing and thus information for these trials is assumed to mimic our estimates from current US trials). This is the primary region in which the policy debate has been focused, with the EU considering increasing testing requirements and the US considering loosening them.

There are a few important caveats to note in using our estimated model to examine EU and US policies. Considering EU policy is relatively straightforward, as the model is estimated directly on EU data. Beyond the standard cautions regarding extrapolations

information problems in addition to providing testing signals, making this result further out of sample both in data and conceptual terms.

[49] Appendix Figure A2 provides additional results on how these effects vary with quality of new products relative to the outside option – the key insight from that analysis is that the value of reducing uncertainty increases as mean product quality increases because higher quality products are used more frequently, so in this sense quality and information are complements.

in any counterfactual modeling exercise, a key decision is how to account for the potential information spillovers the EU receives from ongoing US trials. We address this by calculating optimal policies in our estimated model with and without this spillover.

Using our model to consider US policy requires a bit more care. First, one needs to assume the estimated EU preferences and learning patterns reasonably proxy US preferences and learning patterns over stents. The reduced form evidence in Figure 2 suggests this is a reasonable assumption, but it nevertheless remains an untestable one. Second, we need to determine how large are the fixed costs of US entry beyond regulatory testing might affect entry decisions of manufacturers that do not enter the US in the data. The precise nature and size of these costs are outside of our modeling exercise, and we set them to zero in these counterfactuals. This makes our counterfactual modeling more akin to the thought experiment of a US where more firms have incurred these costs, which might be what we would expect in the longer-run if the US were to loosen its testing policy, making it more attractive for firms to incur the costs of setting up sales and distribution.

In our baseline analysis, we model any policy change as taking place at the beginning of our data period, January 2004, so products having entered before then are not directly affected. For products entering after January 2004, there are several effects: First, entry is delayed by T^c months. Second, fixed costs of entry increase by \$1.6M$\times T^c$. This may cause some firms to decide not to run these additional trials and not enter the market. Third, for products that do enter the market, uncertainty faced by consumers upon entry decreases according to $\sigma^2(T^c) = \left(\frac{1}{\sigma_Q^2} + \frac{1}{\sigma_{EU}^2} + \frac{T^c}{\sigma_{Ac}^2}\right)^{-1}$. Finally, quantities and surplus generation adjust in equilibrium to the set of products in the market and information about product quality. In our primary estimates, we hold prices fixed at the observed prices in the data and calibrate marginal costs for each product to be half the minimum price observed for that product in the data. As discussed in Section 4.3, we impose several simplifying assumptions on supplier behavior to develop cases that are relatively easy to compute, yet still informative regarding policy in this market.

Figure 4 plots expected surplus measures versus T^c for the more entry (M) and less entry (L) cases developed in 4.3. Recall both of these cases incorporate the benefit of more clinical testing decreasing uncertainty, but in M the only cost is delayed access by T^c months, whereas L incorporates the further cost of products not entering if their expected profits do not exceed the fixed costs of testing. Thus, the cases are identical by construction at no trials beyond the current EU requirements $T^c = 0$, but as T^c increases they are driven apart by fewer products entering in L. The surplus values are calculated by using the learning and demand models to simulate the 10 year period we study, and computing a sum of payoffs, discounted at a one percent per month nominal rate after January 2004. In order to focus

on the economic tradeoffs of pre-market clinical testing, the graphs and the first row of the table below them are computed *without* any learning after products enter the market (such as spillovers of US trial information to EU consumers). The second row of the table makes the same computations, but allows for the fact that the EU learns from US trials for products that undertake them.[50] We discuss the results without US spillovers first, and turn to the implications of such spillovers at the end of this subsection.

Appendix E reports further details on this exercise including information on the computation, additional figures for the number of new products entering, fixed costs, and producer surplus. We also report multiple robustness checks to assumptions regarding costs, pricing, and discounting. We find that variations in marginal cost assumptions or modeling price changes using a bargaining equilibrium make very small and statistically insignificant differences in our results. The implications of discounting are more substantive – because products having entered prior to January 2004 are not directly affected in our counterfactuals, it takes time for more of the products in the market to be ones that have benefited from more testing, so that less discounting of the future favors more testing. In the extreme where surplus is computed as an undiscounted sum over our 10 year period, surplus gains are 40-80 percent larger than those reported below, and optimal testing times tend to be slightly longer (though often not statistically different).[51]

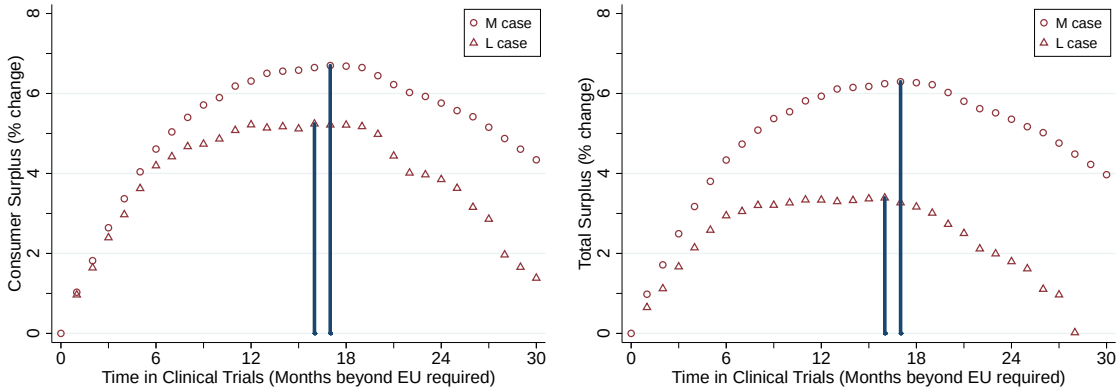
The left subfigure (a) of Figure 4 (and the left panel of the table below) shows the results for consumer surplus. Under both M and L cases, the benefits of risk reduction documented in Table 2 dominate the cost consquences and result in surplus increasing as trial length grows from the current EU baseline, $T^c = 0$. Despite the fact that the sets of products in the two cases quickly diverge (by $T^c = 3$ the M and L cases involve 78 and 46 new products entering over our 10 year sample, respectively), this has a relatively small effect on consumer surplus for low values of T^c because the first products not entering are on average lower quality (beliefs after initial EU testing Q_j^1 are sufficiently correlated with true quality Q_j^* that this is the case). As trial lengths increase, however, the benefits of learning begin to

[50]Here we assume the set of products that undergo US testing and the amount of testing required are held fixed as they are observed in the data. Because we estimate that there is no learning outside of trials in the EU market, EU policy does not affect US policy through that mechanism. It is possible that learning from increased EU trials could provide further certainty about product quality that would change US testing/entry decisions for some products (or that "harmonization" efforts across the regions might allow some trial evidence to count in both), which would result in EU policy affecting US policy and thus require consideration of equilibrium between the two regulatory regimes.

[51]We chose 1 percent nominal per month as it roughly the upper envelope of cost of capital estimates for medical device firms (Harrington 2012). Intermediate discount rates give intermediate results, as one might expect. Using OMB suggested real rates of 7 and 3 percent for regulatory analysis (and average inflation 2004-13 of 2.36 percent) yields surplus gains 10-20 percent and 20-50 percent larger than those at 1 percent nominal per month. These are differences for regulator discounting only. Firms discount expected profit calculations at 1 percent nominal per month in all of our L case entry calculations.

Figure 4: Optimal Regulation: Red markers represent estimates for M (more entry) and L (less entry) cases. Blue lines demarcate optimal trial lengths T^{c*}.

(a) Consumer Surplus, $\Delta CS(T^c)$ **(b)** Total Surplus, $\Delta TS(T^c)$



State of Market at Policy Change	$\Delta CS(T^{c*})$ (%)		T^{c*}_{CS} (months)		$\Delta TS(T^{c*})$ (%)		T^{c*}_{TS} (months)	
	(L)	(M)	(L)	(M)	(L)	(M)	(L)	(M)
Jan 2004	5.2	6.7	16	17	3.9	6.3	16	17
	(1.8)	(1.9)	(4)	(2)	(1.6)	(1.9)	(4)	(2)
Jan 2004, EU given US spillovers	2.0	3.1	11	13	1.0	2.7	6	13
	(1.1)	(1.3)	(4)	(3)	(0.8)	(1.2)	(4)	(3)

$N_{JHT} = 407,191$ product-hospital-months and $N_{JT} = 4,888$ product-months. Standard errors, clustered by month ($N_T = 114$) using a delete-10 block jacknife, in parentheses.

taper off. The products exiting in L are of increasingly high quality, causing M and L to diverge in the consumer surplus implications of longer trials. Our estimates suggest that the optimal policy implications for the two cases are, however, quite similar at $T^{c*}_{CS} = 16$ months for L and 17 for M. The additional consumer surplus generated at the optimum under these cases differs more at $\Delta CS(T^{c*}_{CS}) = 5.2$ percent for L and 6.7 for M (though this is not a statistically significant difference at standard confidence levels).

The total surplus results in the right panel of Figure 4 are roughly similar to the consumer surplus findings, with a few notable differences. First, the gap between the M and L cases is larger. This is due primarily to the widening gap in fixed costs of testing incurred by producers, which drives the lower bound on total surplus to decrease more rapidly.[52] The spread in additional total surplus generated at the optimum is wider at $\Delta TS(T^{c*}_{TS}) = 3.9$ percent for the L case and 6.3 percent for the M case (but again we cannot reject these are

[52]For example, by $T^c = 12$ the bounds on firm entry have widened and span 76 to 32 new products entering over our 10 year sample. See Appendix Figure A13 for entering products, fixed costs, and producer surplus plots. Note that the increase in producer surplus with testing is partially driven by the fact that greater testing benefits Jan 2004 incumbents by delaying competition.

the same at typical significance levels). In the consumer surplus analysis the optimal trial length is the same for both cases. Note that the difference between the L cases for consumer and total surplus generated by testing suggests a rationale for why private incentives may not induce firms to test optimally – producer surplus gains are often outweighed by the fixed costs.

These results speak to the policy debates in the EU and US over the medical device approval pathway. They support those who advocate for stronger clinical requirements in the EU. The results also support the FDA argument that reductions in their standards for device approval will reduce consumer welfare. Our results stand in contrast to Peltzman's (1973) influential analysis of the 1962 Kefauver-Harris Amendments to the Food, Drug, and Cosmetic Act which required proof of efficacy and made the testing procedures required to prove that efficacy subject to FDA oversight. He concludes that the Amendments led to a significant decrease in welfare. Of course, we are comparing a different time and product market than Peltzman considered. And Peltzman's analysis does not speak to the optimal informational requirements pharmaceutical manufacturers should face when introducing a new molecular agent. To the best of our knowledge, our analysis is the first that provides an estimate of the optimal policy on the amount of information creation in regulating product testing.

With the caveats mentioned above, we find that the current FDA policy for stents (the mean lag between US and EU entry is 10 months for all products and 17 months for DES) falls near our confidence interval for the optimal policy in terms of both consumer and total surplus maximization. These results also suggest that surplus could be increased in the EU (5-7 percent consumer and 3-6 percent total) by increasing the pre-market clinical trial requirements, but these numbers do not take into account information spillovers from US trials to the EU. Incorporating the spillovers from products undergoing US testing (shown in the second row of the table at the bottom of Figure 4), the potential surplus gains are cut by more than half, and in the L case the total surplus gains from further testing are no longer statistically significantly different from zero. This provides one partial justification for the low amount of testing that has been required in the EU.

6.4 Sensitivity of Optimal Regulation to Key Parameters

Table 3 explores comparative statics on several important parameters to help better understand the factors that determine the optimal regulation. In order to more clearly focus on the trade-offs between risk and access, we do not include any post-market learning (such as spillovers from the US to EU) in these computations. Thus, the first row here is identical to the analysis without spillovers in the last Section, providing a baseline for the other analyses.

Appendix E reports outcomes for more extensive sets of parameter values than shown here.

Table 3: Sensitivity of Optimal Regulation to Key Parameters: .

State of Market at Policy Change	$\Delta CS(T^{c*})$ (%)		T^{c*}_{CS} (months)		$\Delta TS(T^{c*})$ (%)		T^{c*}_{TS} (months)	
	(L)	(M)	(L)	(M)	(L)	(M)	(L)	(M)
Jan 2004	5.2	6.7	16	17	3.9	6.3	16	17
	(1.8)	(1.9)	(4)	(2)	(1.6)	(1.9)	(4)	(2)
Jan 2004, FC * 2	3.9	6.7	9	17	2.4	6.3	7	17
	(1.6)	(1.9)	(4)	(2)	(1.3)	(1.9)	(6)	(2)
Jan 2004, FC * 5	1.7	6.7	7	17	0.0	6.3	0	17
	(1.1)	(1.9)	(3)	(2)	(0.4)	(1.9)	(2)	(2)
Jan 2004, no DES	3.6	4.6	7	17	2.4	3.9	6	17
	(2.0)	(2.3)	(4)	(5)	(1.6)	(2.3)	(3)	(5)
Jan 2004, no stents	5.4	6.1	7	8	3.7	5.4	7	8
	(2.5)	(2.8)	(3)	(4)	(2.1)	(2.7)	(3)	(4)
Jan 2004, σ_Q * .5	0.1	0.2	1	5	0.0	0.1	0	1
	(0.3)	(0.5)	(2)	(3)	(0.1)	(0.5)	(1)	(4)
Jan 2004, σ_Q * .75	2.2	3.1	8	13	1.3	2.9	6	13
	(1.3)	(1.5)	(4)	(3)	(1.0)	(1.4)	(4)	(3)
Jan 2004, σ_Q * 1.33	8.6	10.2	19	19	6.9	9.6	17	19
	(2.1)	(2.2)	(3)	(3)	(2.0)	(2.1)	(4)	(3)
Jan 2004, σ_Q * 2	12.1	13.7	19	19	10.2	13.1	18	19
	(2.3)	(2.4)	(3)	(3)	(2.2)	(2.3)	(4)	(3)
Jan 2004, $1/\sigma^2_{A^c}$ * .2	0.0	0.1	0	1	0.0	0.1	0	1
	(0.1)	(0.5)	(2)	(4)	(0.0)	(0.4)	(0)	(4)
Jan 2004, $1/\sigma^2_{A^c}$ * .5	1.5	2.6	7	13	0.7	2.4	6	13
	(1.2)	(1.5)	(4)	(4)	(0.9)	(1.4)	(5)	(4)
Jan 2004, $1/\sigma^2_{A^c}$ * .75	3.5	4.9	16	17	2.2	4.6	8	17
	(1.6)	(1.8)	(4)	(3)	(1.4)	(1.7)	(5)	(3)
Jan 2004, $1/\sigma^2_{A^c}$ * 1.33	6.9	8.4	13	17	5.5	8.0	16	17
	(1.9)	(2.1)	(3)	(2)	(1.8)	(2.0)	(3)	(2)
Jan 2004, $1/\sigma^2_{A^c}$ * 2	9.6	10.7	13	17	7.9	10.2	13	17
	(2.1)	(2.2)	(3)	(3)	(1.9)	(2.1)	(3)	(3)
Jan 2004, $1/\sigma^2_{A^c}$ * 5	14.2	15.3	10	12	12.6	14.5	9	11
	(2.3)	(2.3)	(2)	(2)	(2.1)	(2.2)	(2)	(2)

$N_{JHT} = 407,191$ product-hospital-months and $N_{JT} = 4,888$ product-months. Standard errors, clustered by month ($N_T = 114$) using a delete-10 block jacknife, in parentheses.

Fixed Costs of Trials: The second panel of Table 3 reports the relationship between optimal trial length and the costs of trials. By construction, this only affects the L case, and as trial costs move toward zero, the L case converges to the M case. As trial costs increase from our assumed \$1.6M/month, optimal trial length and the additional surplus generated from testing both decrease. For costs double the base case, the L case estimates are $T^{c*}_{CS} =$ 9 months and $\Delta CS(T^{c*}_{CS}) = 3.9$ percent, which are statistically significantly less than the M case. Once costs exceed five times the base case, we can no longer reject that the consumer surplus generated at the optimal under the L case is equivalent to the EU status quo of $T^c = 0$. Point estimates reach zero at fixed costs 10 times those we assume. As expected,

the amount of optimal testing under a total surplus criterion declines even more quickly as fixed costs of testing increase.

Less Existing Substitutes / Larger Quality Increase Due to Technological Change: The third panel of Table 3 reports the estimates from two different scenarios that demonstrate how the impact of regulatory policy changes as the quality of existing technology decreases (and thus new potential entering technologies represent a larger increase in the average quality and variety available in the market). We calculate the optimal trial length as described above but remove (1) all DES; and (2) all stents that were introduced prior to 2004 from the analysis – thus any change in trial length impacts the availability of DES (a significant technological improvement) or of any stent. There are two opposing forces here relative to our baseline. The complementarity between quality and information discussed at the end of Section 6.2 applies here – lower quality incumbent technology means new products will be used more, increasing the value of uncertainty reductions. On the other hand, lower quality incumbents also means the benefit to accessing new technologies sooner is higher. We find that optimal trial lengths tend to decrease as the quality of existing technology decreases, indicating that the (relative change in the) value of access dominates the value of the complementarity between quality and uncertainty reduction for the scenarios we consider. However, we find that welfare is still improved by increasing trial length relative to current EU policy (though by less than in the baseline, due to the tradeoff just discussed, and L cases under total surplus are no longer statistically distinguishable from zero).

Uncertainty of Innovation Quality: The fourth panel documents that the benefit of product testing is tightly linked to the prior uncertainty surrounding the quality of innovations, σ_Q in our model. As prior uncertainty decreases, optimal policy involves less testing. The point estimates become small and statistically indistinguishable from current EU policy when σ_Q is half of our estimate. As prior uncertainty increases, the surplus gains from that testing also increase dramatically. For example, for σ_Q double of our estimate, the L case estimates are $T^{c*}_{CS} = 19$ months and $\Delta CS(T^{c*}_{CS}) = 12.1$ percent, a surplus more than double that of the optimal in our baseline estimates at similar trial lengths.[53]

Precision of Learning in Clinical Trials: The fifth and final panel documents how changes in the precision of learning from clinical trials, $1/\sigma^2_{A^c}$, influences optimal policy. As one might expect, the faster the learning from testing, the more surplus that can be generated by testing. However, the *length* of optimal testing follows an inverted-u shape as the precision of testing increases. For clinical trial learning precision at one fifth of our

[53] Appendix Table A5 shows results for changing risk aversion, which has similar effects to changes in prior risk, with even larger changes in optimal trial length and surplus at similar multiples.

parameter estimate, the optimal policy is short and are statistically indistinguishable from no additional testing beyond the current EU level. As precision increases, optimal trial length and surplus generated increases until it reaches is apex near our baseline parameter estimate. As precision increases beyond our point estimate, the optimal trial length decreases in precision (for consumer surplus this happens around 0.75, and for total surplus about 1.3, times our baseline optimal value). For precision five times our estimate, the optimal L case estimates are $T_{CS}^{c*} = 10$ months and $\Delta CS(T_{CS}^{c*}) = 14.2$ percent, a surplus almost triple at a trial length less than two thirds that of our baseline.

The intuition for this result is that increasing trial precision causes testing to approach "complete learning" more rapidly. Thus, for any amount of testing, the learning benefit is greater in level, but the slope of the benefit from additional learning also flattens out at a shorter trial length. We find this result both interesting and encouraging in that the learning rate from trials is something that policy might hope to influence (e.g., through encouraging surrogate endpoint validation as discussed in Budish et al. (2015)).

6.5 The Value of Post-market Learning

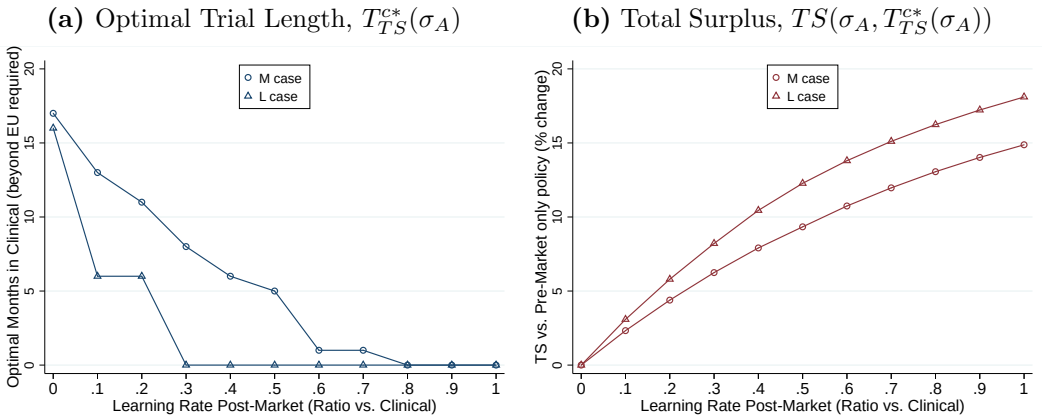
One frequently proposed change to FDA regulatory policy is to relax pre-market clinical standards and increase post-market surveillance. This policy proposal has a direct connection to our model as intends to increase the rate of post-market approval learning. In the language of our model, this implies increasing the precision of the signals that arrive outside of FDA required clinical trials, $1/\sigma_A^2$. We estimate the post market learning rate is effectively zero for the set of products in our data. There are several potential reasons for this finding. For reasons that are familiar to economists, observational learning from real world use make it difficult to infer the causal treatment effect of the device as there is no randomization into treatment and control groups. More fundamentally perhaps, there is currently no infrastructure in place to systematically collect data, perform analysis, and disseminate performance results.

We analyze this policy by varying σ_A (assuming no additional costs), and calculating the corresponding optimal trial length $T_{TS}^{c*}(\sigma_A)$ under a total surplus maximization metric and total surplus generated $TS(\sigma_A, T_{TS}^{c*}(\sigma_A))$ at the optimal. Figure 5 displays the results.

When observational learning approaches clinical trial learning in precision, there is no reason to run additional pre-market trials at all (again assuming required EU testing establishes symmetric information). There is no longer a tradeoff between access and learning because learning can happen after access is granted.[54] Total surplus is increased up to 18

[54]If the welfare of pre-market clinical trial participants is for some reason treated differently than that of post-market users, then there is still a consumer surplus gain to removing uncertainty prior to market

Figure 5: The Value of Post-Market Surveillance: Plots of optimal trial length (left panel (a)) and total surplus (right panel (b)) as observational learning precision $1/\sigma_A^2$ varies from zero to the clinical trial precision $1/\sigma_{A^c}^2$.



percent relative to no observational learning. To calibrate the value of this increased welfare due to increased post-market learning, baseline estimates of utilization of coronary stents in the US and a value of $5,000 per treatment (from the clinical literature cited earlier) yields a $576 million per year increase in welfare.[55]

Before reaching this extreme, as the precision of observational learning decreases (relative to clinical trial learning), it is optimal to require longer pre-market clinical trials.[56] The lesson from this policy experiment is clear. The argument that requiring shorter trials with post-approval testing can improve consumer welfare has merit. However, the gains from this policy critically depend on the rate and cost of learning via post-market surveillance. The viable rate of post-market learning will, in turn, depend on the investments made in collecting, generating, and disseminating performance information.

7 Conclusion

The tradeoff between access to new products and consumer risk in regulating the information required for market entry is important in a variety of industries, and, in particular, in medical devices. Informed by qualitative and quantitative evidence that the US regulatory environment requires more information than the EU via pre-market testing, we develop and

access. See Appendix Figure A14.

[55] In 2009, over 640,000 stent procedures were performed in the US (Auerbach 2012).

[56] Part of the tradeoff with this TS metric is driven by our assumption that post-market learning is costless on the margin and pre-market trials are costly. The decrease in optimal pre-market trial length is slightly less dramatic under the CS metric considered in Appendix E.7.

estimate a structural model with products introduced when quality is still uncertain, learning over time, and regulator and manufacturer decisions regarding product testing and market entry and pricing. We then conduct welfare analyses of counterfactual policies affecting: (1) the length of clinical trials required before market entry; and (2) observational learning after market entry.

For coronary stents 2004-13, we estimate that clinical testing is critical to market function. Without any testing, quality uncertainty plus risk aversion combine to keep many consumers from choosing a stent over alternative treatments. We estimate that the US is close to the optimal policy in terms of trading off testing versus access to innovation, but the EU is too lax (despite free-riding off of information generated by US trials). We also estimate that if it is possible to achieve post-market learning rates close enough to those we observe from clinical trials, then embracing recent calls for more active "post-market surveillance" could further increase total surplus by as much as 18 percent.

We additionally conduct a number of comparative static exercises to examine how optimal policy changes with the parameters of our model, and one takeaway is that results vary enough within reasonable parameter ranges that extrapolating to policy for other products should be done with care. The model we develop provides guidance for how this extrapolation should depend on the uncertainty in quality of new product introductions, the rate of technological improvement, the learning rate in (and cost of) clinical trials, and the observational learning rate for any type of device being considered. But it is difficult to give precise guidance without clear estimates or assumptions regarding these parameters.

At a more phenomenological level, the coronary stent case we analyze here will tend to be most similar to other Class III medical devices, where EU controls involving materials, manufacturing, and smaller clinical trials will typically satisfy FDA safety requirements, and the primary policy debate regarding pre-market clinical trial size focuses on the amount of information generated regarding product efficacy. These devices are also typically used by a relatively small community of specialist surgeons whose expertise and attention to new technology make the assumption of symmetric information among players a reasonable first-order approximation. Also similar in this regard would be considering clinical trial requirements and off-label prescribing of pharmaceuticals by specialists, though these may tend to diverge to the extent that they have meaningful side effects, and thus properly modeling learning and consumer surplus would involve allowing for multidimensional heterogeneity in information and preferences over treatment efficacy and side effects.[57] Further from our context would

[57]In particular, oncology comes to mind, where many cancers are treated with various "cocktails" that combine drugs and have not been tested the same way in which they are often used in practice. In general, off-label prescribing refers to when a drug that is FDA approved for treating a particular disease state is prescribed to treat a different disease state. Thus the drug still has a baseline level of safety, given its

be cases where a drug (or device) is used by generalists or other providers less expert and informed in the specific technology area, where asymmetric information might play a larger role. Finally, furthest from our context would be areas where considering any regulation at all, even basic safety testing, is the policy margin of interest.

Because the model is quite general and the type of data we use is available for many markets, we hope that we have provided a starting point for analysis of regulation and market structure in other industries where new product development and testing are important. As discussed above, other product areas may also suffer from asymmetric information problems or allow more learning via usage. Extending the model to allow for these features and to further explore the extent to which certification solves asymmetry (in addition to amount) of information problems offers another promising (and challenging) area for future research.

We also hope to have provided a building block in the push toward a more complete picture of how regulation affects market structure, innovation, and ultimately welfare. While our exercise here, estimating the welfare effects of the access/uncertainty tradeoff for an exogenously given set of "mid-stage" innovations, is an important step towards better understanding this phenomenon, a more complete understanding would allow for the regulatory regime to effect research and development at even earlier stages. Analysis of this type would require a significant extension to the theory and additional data on innovative activities of the firms. Developing this type of early-stage innovation data, in a way that links to product markets, is a challenge shared with the innovation literature more broadly (Sampat 2018).

References

Ackerberg, D. A. (2003). Advertising, Learning, and Consumer Choice in Experience Good Markets: An Empirical Examination. *International Economic Review*, 44(3):1007–1040.

Aguiar, L. and Waldfogel, J. (2018). Quality predictability and the welfare benefits from new products: Evidence from the digitization of recorded music. *Journal of Political Economy*, 126(2):492–524.

Berry, S. and Waldfogel, J. (1999). Free entry and social inefficiency in radio broadcasting. *RAND Journal of Economics*, 30(3):397–420.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *RAND Journal of Economics*, 25(2):242–262.

approval, but the body of evidence regarding its efficacy for the off-label disease state tends to be less than that required to obtain FDA approval.

Brown, Z. Y. (2017). An empirical model of price transparency and markups in health care. Working Paper, University of Michigan.

Budish, E., Roin, B., and Williams, H. (2015). Do firms underinvest in long-term research? evidence from cancer clinical trials. *American Economic Review*, 105(7):2044–2085.

Cardon, J. H. and Hendel, I. (2001). Asymmetric information in health insurance: Evidence from the national medical expenditure survey. *The RAND Journal of Economics*, 32(3):408–427.

Ching, A. T. (2010). A dynamic oligopoly structural model for the prescription drug market after patent expiration. *International Economic Review*, 51(4):1175–1207.

Ching, A. T., Erdem, T., and Keane, M. P. (2013). Learning models: An assessment of progress, challenges, and new developments. *Marketing Science*, 32(6):913–938.

Cohen, A. and Einav, L. (2007). Estimating risk preferences from deductible choice. *The American Economic Review*, pages 745–788.

Collard-Wexler, A., Gowrisankaran, G., and Lee, R. S. (2014). Bargaining in bilateral oligopoly: An alternating offers representation of the" nash-in-nash" solution. NBER Working Paper #20641.

Crawford, G. and Yurukoglu, A. (2012). The welfare effects of bundling in multichannel television. *American Economic Review*, 102(2).

Crawford, G. S. and Shum, M. (2005). Uncertainty and learning in pharmaceutical demand. *Econometrica*, 73(4):1137–1173.

Dranove, D. and Jin, G. (2010). Quality disclosure and certification: Theory and practice. *Journal of Economic Literature*, 48(4):935–63.

Erdem, T. and Keane, M. P. (1996). Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets. *Marketing Science*, 15(1):1–20.

Filson, D. (2012). A markov-perfect equilibrium model of the impacts of price controls on the performance of the pharmaceutical industry. *RAND Journal of Economics*, 43(1):110–138.

Gandhi, A., Lu, Z., and Shi, X. (2013). Estimating demand for differentiated products with error in market shares.

Gowrisankaran, G., Nevo, A., and Town, R. J. (2014). Mergers when prices are negotiated: Evidence from the hospital industry. *The American Economic Review*, 105(1):172–203.

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

Grennan, M. (2013). Price discrimination and bargaining: Empirical evidence from medical devices. *American Economic Review*, 103(1):145–177.

Grennan, M. (2014). Bargaining ability and competitive advantage: Empirical evidence from medical devices. *Management Science*, 60(12):3011–3025.

Grennan, M. and Swanson, A. (2018). Diagnosing price dispersion. Working paper.

Grennan, M. and Town, R. J. (2018). Regulating innovation with uncertain quality: Information, risk, and access in medical devices. NBER working paper 20981.

Hamilton, B. H., Jungheim, E., McManus, B., and Pantano, J. (2018). Healthcare access, costs, and treatment dynamics: Evidence from in vitro fertilization. forthcoming in the American Economic Review.

Handel, B. R. (2013). Adverse selection and inertia in health insurance markets: When nudging hurts. *American Economic Review*, 103(7):2643–82.

Harrington, S. E. (2012). Cost of capital for pharmaceutical, biotechnology, and medical device firms. In *The Oxford Handbook of the Economics of the Biopharmaceutical Industry*. Oxford.

Ho, K. and Lee, R. S. (2019). Equilibrium provider networks: bargaining and exclusion in health care markets. *American Economic Review*. forthcoming.

Horn, H. and Wolinsky, A. (1988). Bilateral monopolies and incentives for merger. *RAND Journal of Economics*, 19:408–419.

Ifrach, B. and Weintraub, G. (2017). A framework for dynamic oligopoly in concentrated industries. *The Review of Economic Studies*, 84(3):1106–1150.

Kalyanasundaram, A. and Karlheinz, P. (2014). Comparison of revascularization procedures in coronary artery disease. *Medscape*. http://emedicine.medscape.com/article/164682.

Kamenica, E. and Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, 101(6):2590–2615.

Kaplan, A. V. and Stern, A. D. (2018). The central and unacknowledged role of the us food and drug administration in the design and execution of medical device pivotal trials. *JAMA Cardiology*, 3(1):5–6.

Kyle, M. (2007). Pharmaceutical price controls and entry strategies. *Review of Economics and Statistics*, 89(1):88–99.

Lee, R. S. and Fong, K. (2013). Markov-perfect network formation an applied framework for bilateral oligopoly and bargaining in buyer-seller networks. *working paper*.

Makower, J., Meer, A., and Denend, L. (2010). FDA impact of us medical technology innovation: A survey of over 200 medical technology companies. Stanford University Report.

Miravete, E. J., Seim, K., and Thurk, J. (2014). Complexity, Efficiency, and Fairness in Multi-Product Monopoly Pricing. Technical report, Working paper.

Murphy, K. and Topel, R. (2006). The value of health and longevity. *Journal of Political Economy*, 114(3):871–904.

Nistor, C. and Tucker, C. E. (2015). Certification intermediaries: Evidence from the medical device industry. working paper.

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*, 27:175–200.

Peltzman, S. (1973). An evaluation of consumer protection legislation: The 1962 drug amendments. *Journal of Political Economy*, 81(5):1049–1091.

Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy*, 110(4):705–729.

Philipson, T., Berndt, E. R., Gottschalk, A. H., and Sun, E. (2008). Cost-benefit analysis of the FDA: The case of the prescription drug user fee acts. *Journal of Public Economics*, 92(5-6):1306–1325.

Quan, T. and Williams, K. (2017). Product variety, across market demand heterogeneity, and the value of online retail. Cowles Foundation Discussion Paper No. 2054.

Roberts, J. H. and Urban, G. L. (1988). Modeling multiattribute utility, risk, and belief dynamics for new consumer durable brand choice. *Management Science*, 34(2):167–185.

Sampat, B. (2018). A survey of empirical evidence on patents and innovation. NBER working paper 25383.

Schreyögg, J., Stargardt, T., Tiemann, O., and Busse, R. (2006). Methods to determine reimbursement rates for diagnosis related groups (DRG): a comparison of nine european countries. *Health Care Management Science*, 9(3):215–223.

Scott Morton, F. and Kyle, M. (2012). Markets for pharmaceutical products. In Pauly, M., McGuire, T., and Barros, P., editors, *Handbook of Health Economics*, volume 2, chapter 12. Elsevier.

Seim, K. and Waldfogel, J. (2013). Public monopoly and economic efficiency: Evidence from the pennsylvania liquor control board's entry decisions. *American Economic Review*, 103(2):831–62.

Sorenson, C. and Kanavos, P. (2011). Medical technology procurement in europe: a cross-country comparison of current practice and policy. *Health Policy*, 100(1):43–50.

Spence, M. A. (1975). Monopoly, quality, and regulation. *Bell Journal of Economics*, 6(2):417–429.

Stern, A. D. (2017). Innovation under regulatory uncertainty: evidence from medical technology. *Journal of Public Economics*, 145:181–200.

Stern, A. D., Kramer, D. B., Ouellet, M., and Kesselheim, A. S. (2017). Review times and adverse events for cardiovascular devices. *Nature*, 1(0013):0013.

Timmins, C. (2002). Measuring the Dynamic Efficiency Costs of Regulators' Preferences: Municipal Water Utilities in the Arid West. *Econometrica*, 70(2):603–629.

Train, K. (2015). Welfare calculations in discrete choice models when anticipated and experienced attributes differ: A guide with examples. *Journal of Choice Modelling*, 16:15–22.

Van Norman, G. A. (2016). Drugs and devices: comparison of european and us approval processes. *JACC: Basic to Translational Science*, 1(5):399–412.

ELECTRONIC APPENDICES—NOT FOR PRINT PUBLICATION

A Data Appendix

A.1 Dataset construction

The dataset used in this paper is from Millennium Research Group's Marketrack survey of catheter labs, the source that major device manufacturers subscribe to for detailed market research. The goal of the survey is to provide an accurate picture of market shares and prices of medical devices. For our purposes, the key variables in the data are the price paid and quantity used for each stent in each hospital in each month. In addition, the hospitals report monthly totals for different procedures performed, such as diagnostic angiographies. The data span January 2004 through June 2013 and cover the US and EU markets.

There are three main challenges in constructing a usable dataset from the raw survey data. First, the survey was not as concerned with collecting price data as it was with collecting quantity data. Second, the survey measures stent usage rather than availability, and our data go back only to 2004, so it is not always possible to infer regulatory approval dates from the data (and while our independent research found most introduction dates, we were not able to find all). Finally, there is some apparent misreporting in the survey. The following tables illustrate how key sample summary statistics compare across our cleaning steps for the EU and US datasets. These steps are summarized below; full detail can be found in the Stata code used to execute them, `cleaning-eu-data-3-sample.do` and `cleaning-us-data-3-sample.do`.

EU dataset modifications

	Diagnostic procedures	No. of stents implanted	No. of BMS products	No. of DES products	Average stent age	Stent-hospital-months	Hospital-months	Hospitals
Raw data	151	108	3.8	3.3	54.3	88,144	15,064	542
Rm. suspect q	161	98	3.3	2.8	54.5	61,098	13,477	540
Rm. if q¿2*diagnostics	152	107	3.8	3.3	54.3	86,672	14,812	537
Rm. suspect diagnostics	151	108	3.8	3.3	54.4	87,349	14,933	542
Rm. outlier p	148	106	3.8	3.3	54.4	81,646	14,149	532
Rm. unknown entry	150	108	3.8	3.3	54.0	87,516	14,995	541
Final sample	160	95	3.2	2.8	54.6	54,771	12,313	524

US dataset modifications

	Diagnostic procedures	No. of stents implanted	No. of BMS products	No. of DES products	Average stent age	Stent-hospital-months	Hospital-months	Hospitals
Raw data	137	76	2.2	2.5	36.8	68,603	17,183	526
Rm. suspect q	147	68	1.9	2.1	37.8	44,218	14,631	509
Rm. if q¿2*diagnostics	138	76	2.2	2.5	36.7	67,783	16,982	517
Rm. suspect diagnostics	138	76	2.2	2.5	36.8	67,857	16,997	526
Rm. outlier p	136	75	2.2	2.5	37.1	66,293	16,720	525
Final sample	147	67	1.8	2.1	38.0	41,779	13,900	478

The table rows record the sample means for key summary statistics across various cleaning steps. The summary statistics are means of quantities calculated at the hospital-month level. The means reported are of the total number of stents implanted; the total number of diagnostic angiographies; the number of different bare-metal stents (BMS) used; the number of different drug-eluting stents (DES) used; and the weighted average age, in months, of the stents used. The table also shows the total number of stent-hospital-month observations, number of hospital-month observations, and number of hospitals in each sample.

The table rows correspond to different samples. The first row of each table summarizes the raw EU and US survey data. The second row drops hospital-months with suspect total quantities. The criteria for dropping are threefold. First, we drop hospital-months for which the total quantity of stents changes by more than 50% relative to the previous month in which the hospital appears in the data. Second, for "low-quantity" hospitals with mean monthly stent quantities below 15, we drop hospital-months with usage strictly greater than 1.5 standard deviations from the hospital's mean. For "high-quantity" hospitals with mean monthly stent quantities (weakly) greater than 15, we drop hospital-months with usage strictly greater than 3.0 standard deviations from the hospital's mean. Third, for hospital-months with flagged quantity changes that were accompanied by a 30% or greater change in diagnostic angiography procedures, the hospital-months were undropped. Diagnostic angiography procedures are performed prior to coronary stent implantation, so large changes in monthly stent quantities should be accompanied be similarly large changes in angiographies.

The third and fourth rows of the table drop hospital-months with suspect diagnostic angiography counts. Diagnostic angiographies should be bounded below by some multiple of the number of stents used; in our data and anecdotally according to clinicians, there are on average about two stents implanted per procedure. The third row drops hospital-months if the number of diagnostic angiographies is less than two times the number of stents implanted in that hospital-month. The fourth row drops hospital-months if the number of diagnostic angiographies is more than 2 standard deviations away from the hospital's mean and if the ratio of angiographies to stents was 2 standard deviations from the hospital's mean.

The fifth row of the table drops hospital-months with problematic prices. We drop hospital-months with outlier prices based on a regression of log-price on the hospital's number of BMS products and number of DES products used that month, in addition to a hospital fixed-effect. Hospital-months with products whose regression residuals were more than 2 standard deviations from the mean of all residuals were dropped.

The sixth and penultimate row of the E.U. table drops hospital-months with positive quantities for stents for which E.U. regulatory approval dates are not known. Since the age of the product is an important component of our analysis, the products for which an entry

date could be pinned down with reasonable certainty must be removed from the analysis. This drop affects only a few products, none of which were frequently used. There are no products for which the US approval dates could not be ascertained.

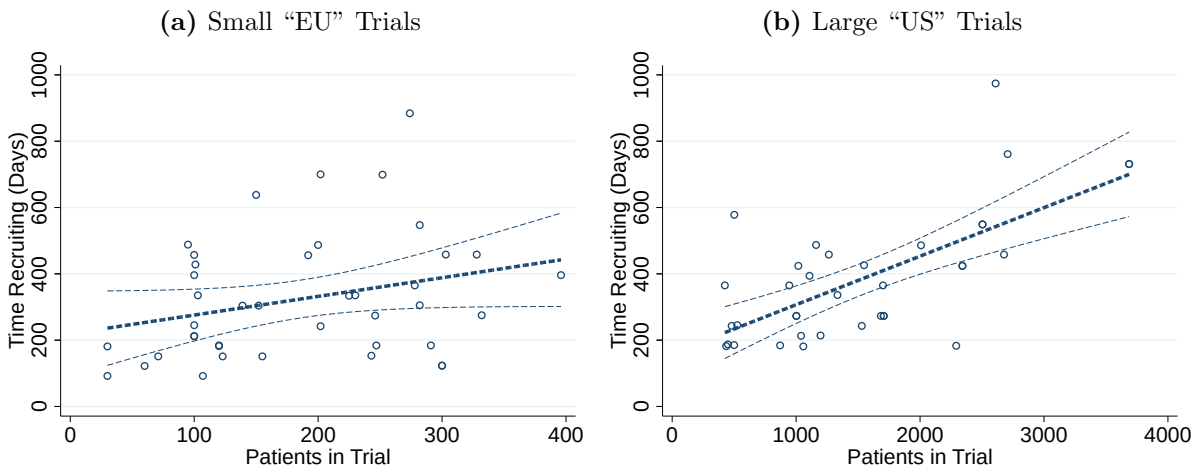
The final row in each table reports summary statistics for the final sample, which drops all observations that meet one or more of the dropping criteria described above.

A.2 Clinical trial data

Our collected clinical trial data, and a detailed document on the sources, are available in the online archive and upon request from the authors.

In addition to clarifying the differences between EU and US trial policy and validating our product quality estimates, the trial data make clear the strong relationship between the size of clinical trial in terms of patients and the time spent on the trial via the time it takes to recruit patients. Figure A1 plots the data on patients and length of recruitment in days for smaller and larger trials (broken down to roughly correspond to the scale of trials required for "EU" and "US" approval). One can see from the fitted lines that larger trials take longer. The fit is not perfectly linear, as there are of course idiosyncracies to particular trials, but especially for the larger "US" trials, which tend to be run by professional units within large firms or third party research organizations that do this as their core business, the fit is reasonably tight, implying an average arrival rate of 186 patients per month.

Figure A1: Relationship between trial size and time.

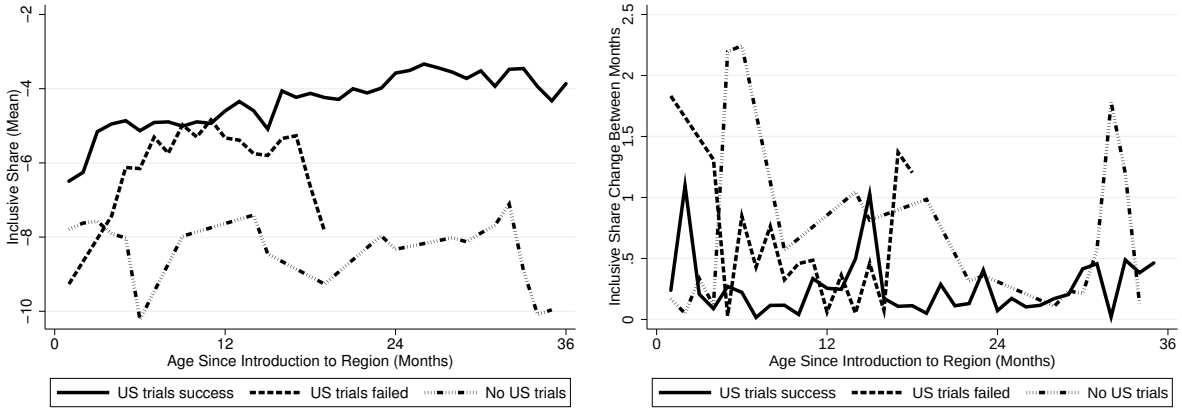


B Robustness and Alternative Explanations: Supplemental Figures and Discussion

B.1 Evidence of learning from individual products

Figure A2: Learning patterns for selected individual products. Three representative products that receive good and bad news from trials or not much (useful) news at all. Left panel (a) plots mean utility estimate for each product $\ln(s_{ja}/s_{0a})$ by age since introduction into the EU. Right panel (b) plots absolute differences $|\ln(s_{ja}/s_{0a}) - \ln(s_{ja+1}/s_{0a+1})|$ by age, which should be larger with more uncertainty, and converge toward zero with learning.

(a) $\ln(s_{jt}/s_{0t})$ **(b)** $|\ln(s_{jt}/s_{0t}) - \ln(s_{jt+1}/s_{0t+1})|$



	mean	s.d.	p25	p50	p75	N
$\Delta_t \ln(s_{jt}/s_{0t})\|_{a=1}$	0.24	1.14	-0.16	0.12	0.60	27
$\Delta_t \ln(s_{jt}/s_{0t})\|_{a=12}$	0.17	0.50	-0.04	0.08	0.27	29
$\Delta_t \ln(s_{jt}/s_{0t})\|_{a=24}$	-0.11	0.30	-0.31	-0.06	0.11	32

Averaging across products conditional on age provides patterns in the data that have direct relation to expected patterns in our model. However, these averages cloud heterogeneity across products. Figure A2 provides two types of evidence of this variation. First, the figures in the panels provide patterns for a few individual products, illustrating how learning does not always bring good news, and lack of learning brings a volatile mix of good and bad over time. Second, the table below the panels provides summary statistics on the raw changes in usage patterns with age $\ln(s_{jt}/s_{0t}) - \ln(s_{jt+1}/s_{0t+1})$ for products in the EU, undergoing US trials.

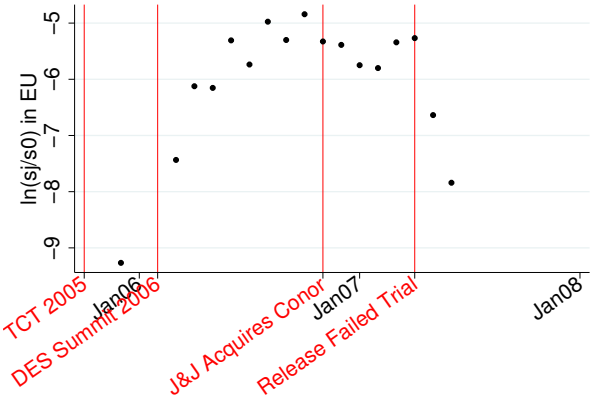
The patterns documented previously regarding decreases in volatility and increasing mean usage with age might be worrisome if they were driven by increasing usage for all product with age that then asymptotes as in a diffusion process. The table on the raw usage changes

show this is not the case—there is a large fraction of changes that are "bad news" for products.

B.1.1 Case Study: CoStar and the Role of Bad News

Here we focus on a clear example of the impact of bad news. A small firm named Conor Medsystems developed a drug-eluting stent with an intuitively appealing new design for drug release that performed well in small early trials (CoStar I (87 patients) and EuroStar I (149 patients)), which were received enthusiastically at conferences in late 2005 through 2006. During this period, pivotal US trials were begun. The stent saw growing market share after receiving a CE mark and being released in the EU in February 2006.[58] In November 2006, Johnson & Johnson was sufficiently optimistic about CoStar to buy Conor for $1.4B. J&J took over CoStar's pre-market notification submission to the FDA. In May 2007 J&J announced the results of a large US trial (CoStar II (1675 patients)), where safety evidence was good but efficacy was disappointing with TLR rates 8% for CoStar versus 4% for its competitor and the control stent, Taxus. Shortly after, J&J announced that it was terminating its FDA mandated clinical trials as the stent was failing to meet its primary endpoints.[59]

Figure A3: Evolution of $\ln(s_{jt}/s_{0t})$ for CoStar.



The CoStar story demonstrates many of the themes of our analysis. CoStar's usage rose as early trial results were communicated at physician conferences and it underwent US trials. As more information was generated via the clinical trial, that information is reflected in the inclusive share. Presumably J&J shared this optimism and did not possess differential information, even after due diligence that would have made it privy to the same information

[58] See http://www.ptca.org/pr_conor/20060217.html

[59] See http://www.investor.jnj.com/releasedetail.cfm?releaseid=241182.

as Conor. And when trial results on efficacy were unfavorable, market share dipped and the product was pulled from the market.

B.2 Robustness and Mechanism Tests: Supplemental Figures and Discussion

B.2.1 Placebo Test: PTCA Balloons

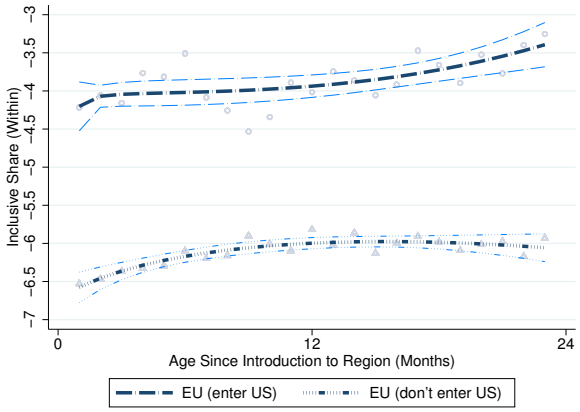
One alternative explanation for the findings in Figure 2 would be that the set of manufacturers/products that undergo US trials promote their products differently than other products in the EU, and also differently than for the same products upon US introduction. While we believe the evidence on decreasing variance and on the same products upon US launch make this unlikely, it is not impossible. To further explore this possibility, we perform a placebo test using percutaneous transcatheter coronary artery (PTCA) balloons, which are FDA Class II devices and thus face similar regulatory requirements in both the EU and US. Thus PTCA should not display the differential signs of learning we document for stents if our proposed mechanism is true. The results here show that we do see more total entry in the EU (presumably due to pre-existing complementary sales and distribution assets in the US for some manufacturers); but the differences in amount of entry are smaller than in stents, there is no gap in time of entry on average, and usage patterns with age show no evidence of learning.

As another check that our results are indeed capturing learning in the EU from US clinical trials, we perform a "placebo" type analysis by looking at a device where we know such trials are not required. We perform the analysis on PTCA balloons catheters, which are often used to clear a blockage in the artery before the stent is placed. Standard balloons (ones that do not have drug coatings or special cutting capabilities) typically have little, if any, gap between US and EU approval requirements. This is evident in the lag between US and EU introduction of on average two months (here we calculated entry from first observation in the data instead of looking up press releases, and so the confidence interval includes zero when sampling error is taken into account). Despite this lack of lag for those products introduced in both the US and EU, we still observe many balloons introduced only in the EU because they are sold by the same sales force as stents, but are much lower revenue products, so that only a few companies enter the US market for the purpose of selling balloons only. During our ten year sample, 40 manufacturers sell 113 different balloons in the EU and 6 manufacturers sell 40 different balloons in the US. Thus we can execute our same research design on balloons, with the expectation of no differential learning between products that are EU only versus those that enter the US as well.

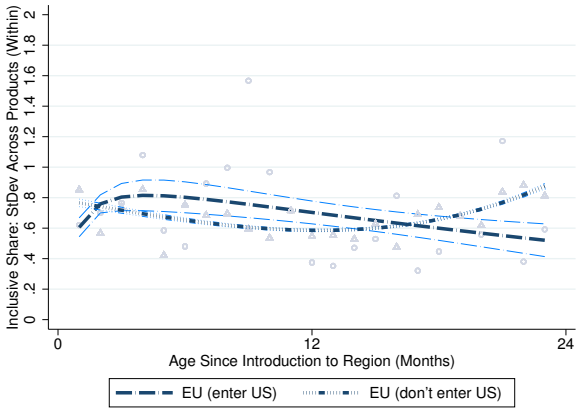
Figure A4 shows the results of this placebo test, comparing EU data for products that do

Figure A4: PTCA Balloons—EU only, products that enter US vs. not.

(a) $\text{Mean}_{j|a}\ \ln(s_{jt}/s_{0t})$ (within product)



(b) $\text{SD}_{j|a}\ \ln(s_{jt}/s_{0t})$



	$x_{a=1}$	$x_{a=24}$	$x_{24} - x_1$	$(x_{24}^{trials} - x_1^{trials}) - (x_{24}^{not} - x_1^{not})$
$\text{Mean}_{j\|a}^{EU\|_{trials}}\ \ln(s_{jt}/s_{0t})$	-4.22	-3.25	0.96	0.37
	(0.17)	(0.27)	(0.32)	(0.33)
$\text{Mean}_{j\|a}^{EU\|_{not}}\ \ln(s_{jt}/s_{0t})$	-6.52	-5.93	0.60	
	(0.16)	(0.18)	(0.21)	
$\text{SD}_{j\|a}^{EU\|_{trials}}\ \ln(s_{jt}/s_{0t})$	0.62	0.59	-0.03	0.01
$\text{SD}_{j\|a}^{EU\|_{not}}\ \ln(s_{jt}/s_{0t})$	0.85	0.81	-0.04	

$N = 789$ product-month observations (all in EU). Standard errors clustered by month $N_t = 114$ in parentheses.

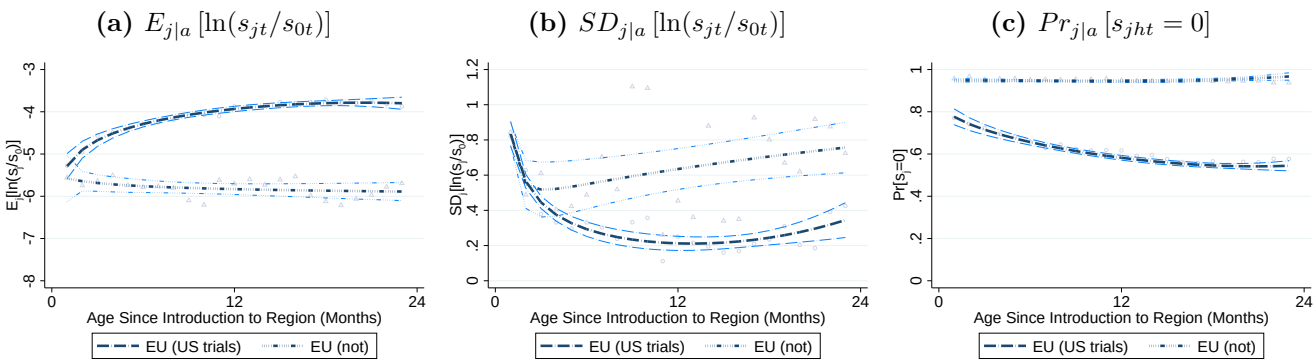
and do not enter the US as well. The results illustrate the importance of looking at learning evidence in the volatility along with trends in means as well as the importance of having comparison groups to be able to look at differences-in-differences. There is no evidence of learning in the volatility figure. Mean usage of products in both groups trend up slightly with age, but these trends are statistically identical, suggesting a slight diffusion process that affects all balloons in the EU that is not driven by learning about product quality.

B.2.2 Alternative Explanation: Observational Learning with Different Initial Sample Size

Another potential explanation for the results in Figure 2 is that there is learning in the EU sample undergoing US trials, but this learning is observational (all or in part). The difference between the patterns in the two samples is then plausibly driven by the fact that those stents undergoing US trials enter with higher usage levels, which generate sufficient sample sizes for observational learning to occur, whereas the EU sample not undergoing trials contains too many products that do not gain enough early traction to enable learning.

We examine this hypothesis by reformulating the same figures and tests for a set of products with overlapping support on initial values of $\frac{1}{J_a}\sum_j \ln(s_{ja}/s_{0a})$ at $a_j = 1$, so they all have similar chances to generate early observational learning. The pattern in Appendix Figure A5 is essentially identical to that in Figure 2, suggesting that our results are not driven by selection on initial quality/usage levels.[60]

Figure A5: Stent usage patterns after product entry, by region and trial status (subsample matched on $age = 1$ usage)



	$\theta_{a=1}$	$\theta_{a=24}$	$\Delta\theta_a$	$\Delta\theta_a^{EU\|UStrials} - \Delta\theta_a^{row}$
$E_{j\|a}^{EU\|UStrials}\ \ln(s_{jt}/s_{0t})$	-5.27	-3.87	1.40***	
	(0.26)	(0.15)	(0.30)	
$E_{j\|a}^{EU\|not}\ \ln(s_{jt}/s_{0t})$	-5.56	-5.70	-0.13	1.54***
	(0.33)	(0.31)	(0.46)	(0.56)
$SD_{j\|a}^{EU\|UStrials}\ \ln(s_{jt}/s_{0t})$	0.82	0.42	-0.40***	
	(0.18)	(0.05)	(0.19)	
$SD_{j\|a}^{EU\|not}\ \ln(s_{jt}/s_{0t})$	0.84	0.72	-0.12	-0.28
	(0.08)	(0.12)	(0.15)	(0.25)
$Pr_{j\|a}^{EU\|UStrials}(s_{jht} = 0)$	0.78	0.54	-0.24***	
	(0.04)	(0.02)	(0.05)	
$Pr_{j\|a}^{EU\|not}(s_{jht} = 0)$	0.95	0.96	0.01**	-0.25***
	(0.01)	(0.02)	(0.02)	(0.05)

$N^{EU|UStrials} = 197$ (8 products), and $N^{EU|not} = 159$ (10 products) product-month observations. Standard errors clustered by month $N_t = 114$ in parentheses. $\Delta\theta_a := \theta_{a=24} - \theta_{a=1}$.

B.2.3 Alternative Explanation: Asymmetric Information and Signaling

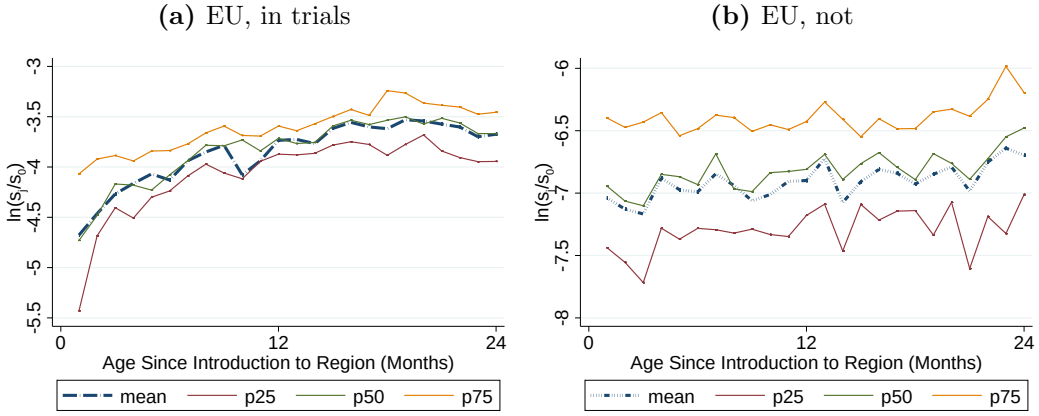
Another potential explanation that could rationalize Figure 2 is manufacturer signaling. Under this hypothesis, after the release of EU trial data, manufacturers retain a sufficiently

[60] For this matched sample, selection into US trials must be based on level shifts in expected US profit due to the fact that those products that enter the US all have pre-existing complementary assets for sales and distribution (while those that don't enter do not). This is consistent with the challenges firms such as Biotronik have faced in develop US sales forces. See, "Tipping the Odds for a Maker of Heart Implants," *New York Times*, April 2, 2011.

large degree of private information about expected product quality, and so undertaking costly US trials signals expected product quality to physicians. To produce the observed data patterns, such a signaling model also needs to include some combination of slow signal diffusion across hospitals and/or increasing signal strength as a trial continues. We explore this hypothesis by looking more closely at the shapes of the distribution of $\ln(s_{jt}/s_{0t})$ with age.

Appendix Figure A6 shows the evolution with age of different quantiles of the $\ln(s_{jt}/s_{0t})|_a$ distribution. Under a model where manufacturers and physicians are similarly informed about quality after the release of trials for EU entry, and then learn similarly as data from US trials is released, the distribution of product quality estimates should converge symmetrically to the true product quality distribution. In an asymmetric information setting, consumers do not receive direct information about quality, but instead infer quality must be above some threshold if a manufacturer is willing to continue with costly testing (see Appendix Figure A7 below for more on this intuition). Learning in this way would cause the lower tail of the distribution for product in US trials to become truncated. In the Figure, the 25th and 75th percentiles appear to move symmetrically towards the median as information arrives. Below the figure, we present relevant test statistics. The change in the skewness of the distribution and the change in the ratio of the 75th-50th percentile to the 50th-25th are both insignificant.

Figure A6: Symmetry of changes in quality distributions

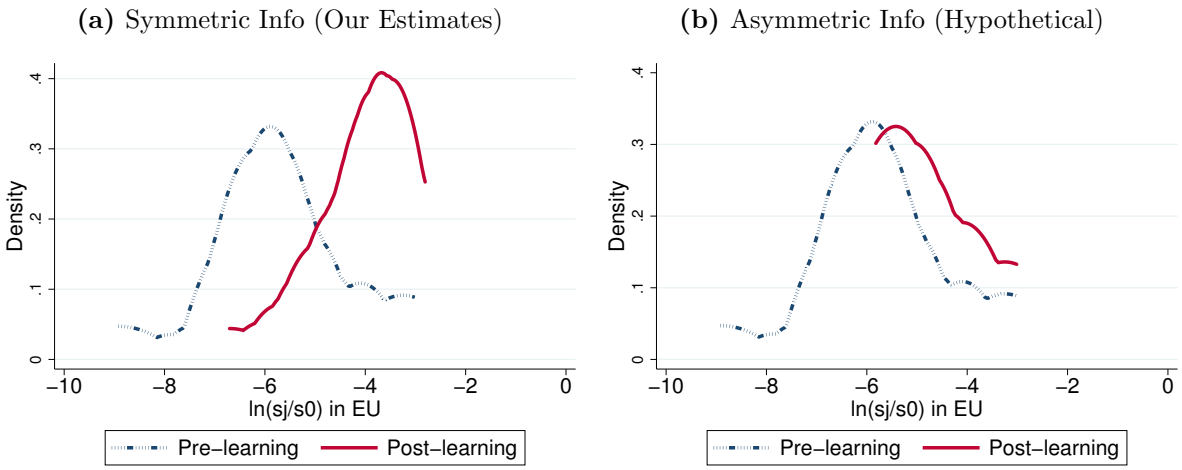


	EU, in trials			EU, not		
	$\theta_{a=1}$	$\theta_{a=24}$	$\Delta\theta_a$	$\theta_{a=1}$	$\theta_{a=24}$	$\Delta\theta_a$
$\left(\frac{\mu-p50}{\sigma}\right)_{j\|a}$ $\ln(s_{jt}/s_{0t})$	0.06	-0.08	-0.14	-0.09	-0.11	-0.01
$\left(\frac{p75-p50}{p75-p25}\right)_{j\|a}$ $\ln(s_{jt}/s_{0t})$	0.49	0.41	-0.08	0.53	0.42	-0.10

$N = 383$ product-months (in EU; US trials). Standard errors clustered by month $N_t = 114$ in parentheses.

Our test of information symmetry in Figure A6 relies upon the intuition that symmetric learning (as we assume in our model) suggests that the inferred distribution of product qualities should tighten from both ends of the distribution as learning occurs (and also shift up if consumers are risk averse). This contrasts with a model where suppliers have private information about their product qualities, where consumer learning should take the form of realizing that manufacturers who engage in costly testing must have product quality exceeding some threshold, which suggests that the inferred distribution of product qualities should tighten from the bottom as learning occurs. Figure A7 illustrates these ideas graphically.

Figure A7: Learning effects on inferred product quality distributions under symmetric and asymmetric information mechanisms.



The left panel (a) plots two distributions directly from our EU data for stents undergoing US trials: (Pre-learning) plots the density of $\ln(s_{jt}/s_{0t})|_{a=1}$; and (Post-learning) plots the density of $\ln(s_{jt}/s_{0t})|_{a=12}$. As one would expect from Figure A6, the distribution shifts up and tightens symmetrically after 12 months in US clinical trials.

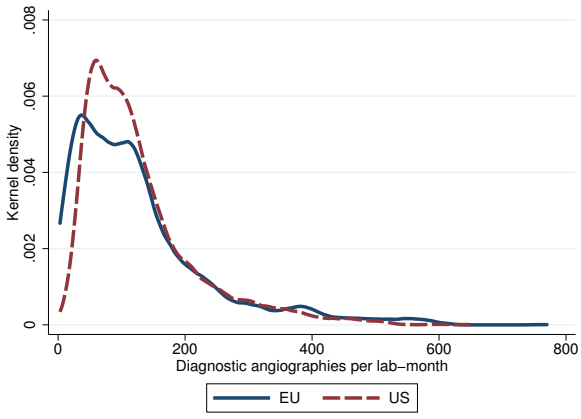
The right panel (b) plots the same pre-learning distribution, and displays the expected post-learning distribution from applying a truncated learning rule $\ln(s_{jt}/s_{0t})|_{a=1,\ln(s_{jt}/s_{0t})>-6}$. The plot illustrates the type of distribution we might expect if there were learning with asymmetric information. This is clearly different from the symmetric model and from our data, which is why our test in Figure A6 fails to reject the null hypothesis of symmetric learning.

B.3 EU vs. US: Other Differences Driving Entry and Diffusion Patterns?

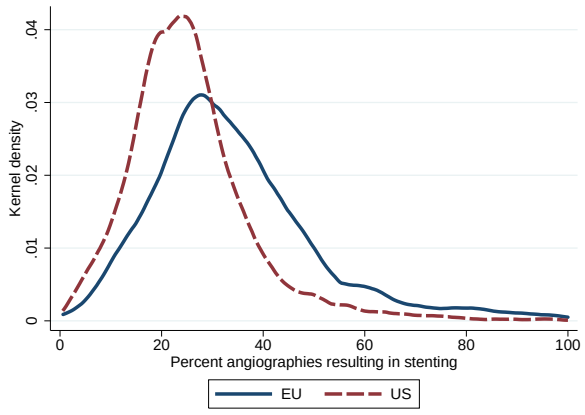
In theory it could be that the differences in usage patterns between the US and EU are driven by differences in disease incidence, preferences for angioplasty and stents, or variation in price setting regimes between the US and EU. However, all the evidence that we have been able to gather indicates that these explanations do not plausibly explain the patterns in the data described above. For example, the average ischemic heart disease mortality rate is very similar between the US and the EU, suggesting that the disease incidence is also similar. The 2010 mortality rate in the US for ischemic heart disease was 126.5 deaths per 100,000; and the corresponding figure for the EU is 130.0 per 100,000.[61] This modest differential seems unlikely to account for the stark differences of entry rates between the two regions.

Figure A8: Comparison of diagnostic procedure patterns, EU vs. US. Left panel (a) plots the distribution of number of diagnostic procedures across hospitals—the US and EU are nearly identical. Right panel (b) plots the distribution across hospitals of the probability that a diagnostic procedure results in stenting—the EU is shifted slightly to the right of the US, with a mean of 32 versus 28 percent.

(a) Diagnostic Angiographies



(b) Percent Resulting in Stenting



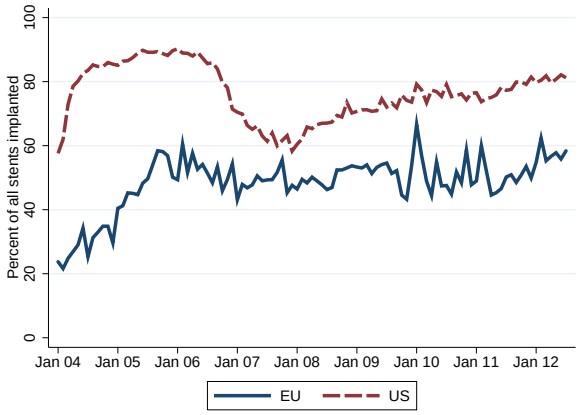
Prior to performing an angioplasty in which a stent may be inserted, the patient must undergo a diagnostic angiography. In this procedure, the blood flow through the coronary artery is visualized and this information is used to determine whether the patient should receive a stent or some other medical intervention. If the difference in the number of stents available between the EU and the US was driven by higher demand for stents, then it should show up in the data with the EU performing a larger number of angiographies or having a higher rate of stenting conditional on the angiography rate. Figure A8 documents the

[61] OECD *Health at a Glance, 2013.*

distributions of the number of diagnostic angiographies performed across the hospitals in our data and percent of those diagnostic procedures resulting in a stenting procedure across hospitals in the US and EU samples. The distributions are close to identical statistically, with the EU having a few more small volume hospitals and hospitals that are more likely to place a stent conditional upon a diagnostic procedure. In the EU, 32 percent of patients received a stent conditional on an angiography while in the US that figure was 28 percent. Like the evidence on heart disease prevalence, this small difference seems unlikely to explain the large disparity in entry rates between the two regions.

Figure A9: Comparison of usage and price patterns EU vs. US.

(a) DES Share of All Stents



(b) Average Price Paid

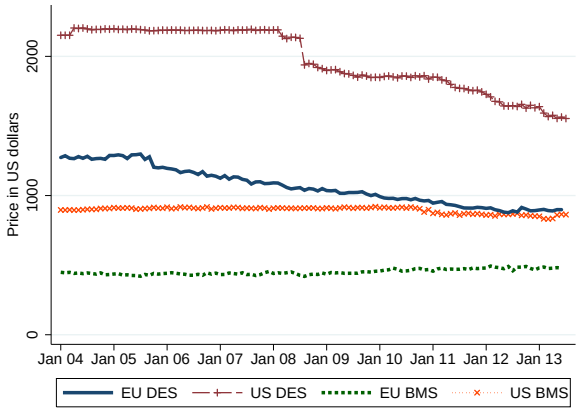


Figure A9 documents that DES usage as a percentage of all stents used is lower in the EU but follows similar patterns to the US over time. If the increased DES entry in the EU was driven by higher demand, we would expect the opposite pattern. Figure A9 also shows that the prices and hence profits per stent sold are lower in the EU. This is true for both BMS and DES and is true over our entire sample period. Both of these patterns are likely the result of lower reimbursement levels for stent procedures overall, lower DES reimbursement levels in particular, and more competing devices in the EU market. These findings suggest that conditional upon FDA approval, average variable profit in the US is higher making it a more attractive entry target than the EU. This, in turn, suggests that the differential entry rates are driven by differences in regulation and not underlying demand.

C Theory Appendix

This appendix provides formulas and proofs to supplement the results provided in the body of the paper.

C.1 Nested Logit Demand Formulas

Choice probabilities are given by:

$$\begin{aligned} cp_{jht} &= Pr[U_{ijht} > U_{ikht}, \forall k \in \mathcal{J}_t] \\ &= \frac{\exp\left(\frac{\delta_{jht}}{1-\lambda^{g_j}}\right)}{\sum_{k\in\mathcal{J}_t^{g_j}} \exp\left(\frac{\delta_{jht}}{1-\lambda^{g_j}}\right)} \frac{\left(\sum_{k\in\mathcal{J}_t^{g_j}} \exp\left(\frac{\delta_{jht}}{1-\lambda^{g_j}}\right)\right)^{1-\lambda^{g_j}}}{1+\sum_{\mathcal{J}_t^{g_j}\subset\mathcal{J}_t}\left(\sum_{k\in\mathcal{J}_t^{g_j}} \exp\left(\frac{\delta_{jht}}{1-\lambda^{g_j}}\right)\right)^{1-\lambda^{g_j}}} \end{aligned} \tag{11}$$

where $\delta_{jht} := Q_{jht} - \frac{\rho}{2}\sigma_{jt}^2 - \theta^p p_{jht} + \xi_{jh}$ is the mean ex-ante expected utility across patients given beliefs regarding the mean stent performance characteristics and the variance of those beliefs. The corresponding elasticity of choice probabilities with respect to own price is given by

$$\eta_{jht} := \frac{\partial q_{jht}}{\partial p_{jht}} \frac{p_{jht}}{q_{jht}} = -\theta^p \left(\frac{1 - \lambda^{g_j} cp_{jht|g} - (1-\lambda^{g_j}) cp_{jht}}{1-\lambda^{g_j}} \right) p_{jht}. \tag{12}$$

The ex-ante expected consumer surplus (relative to the outside option) as a function of information and choice set is

$$CS_{ht}(\mathcal{J}_t, \mathcal{I}_{ht}) = \theta^{scale} \ln \left(1 + \sum_{\mathcal{J}_t^{g_j}\subset\mathcal{J}_t} \left(\sum_{j\in\mathcal{J}_t^{g_j}} \exp\left(\frac{\delta_{jht}}{1-\lambda^{g_j}}\right) \right)^{1-\lambda^{g_j}} \right) . \tag{13}$$

where θ^{scale} is set to make fully informed the average treatment on the treated vs. non-stent alternatives for DES introduced to the US equal to \$5000, as motivated by the clinical literature discussed in the body of the paper.

$$5000 = \theta^{scale} \frac{1}{|\mathcal{J}_{DES,US}|} \sum_{j\in\mathcal{J}_{DES,US}} \frac{\ln\left(1+\exp\left(\delta_{jht}\right)\right)}{\left(\frac{1+\exp\left(\delta_{jht}\right)}{\exp\left(\delta_{jht}\right)}\right)} \tag{14}$$

C.2 Regulator's Total Surplus Tradeoff: Illustrative Case

The general total surplus function is complicated by the entry policies of firms, tracking observational learning for firms that entered the market at different times, and potential

distortions due to heterogeneity in marginal costs and price markups. To clearly see the core tradeoff between uncertainty and access in the model, it is helpful to consider a simple case of a simple logit demand model with testing and entry costless, no observational learning, homogenous marginal costs (normalized to zero for convenience), and no distortions in usage due to price. In this case, the regulator's tradeoff simplifies as follows:

$$\begin{aligned} TS_t(T^c+1) - TS_t(T^c) &= \ln\left(\frac{\sum_{j\in\mathcal{J}_t(T^c+1)} e^{Q_{jt}-\frac{\rho}{2}\sigma^2_{jt}(T^c+1)}}{\sum_{j\in\mathcal{J}_t(T^c)} e^{Q_{jt}-\frac{\rho}{2}\sigma^2_{jt}(T^c)}}\right) - \chi\left|\mathcal{J}^e_t(T^c+1)\setminus\mathcal{J}^e_t(T^c)\right| \\ &= \ln\left(\frac{\sum_{j\in\mathcal{J}_t(T^c+1)} e^{Q_{jt}-\frac{\rho}{2}\sigma^2_{jt}(T^c+1)}}{\sum_{j\in\mathcal{J}_t(T^c)} e^{Q_{jt}-\frac{\rho}{2}\sigma^2_{jt}(T^c)}}\right) \qquad (15) \\ &= \underbrace{\frac{\rho}{2}\left(\sigma^2(T^c)-\sigma^2(T^c+1)\right)}_{\text{gain from decreased risk}} \underbrace{-\ln\left(\frac{\sum_{j\in\mathcal{J}_{t+1}(T^c+1)} e^{Q_{jt}}}{\sum_{j\in\mathcal{J}_t(T^c+1)} e^{Q_{jt}}}\right)}_{\text{gain from tech change/entry}} \qquad (16) \end{aligned}$$

where (15) follows from no fixed costs, and (16) follows from no observational learning and recognizing $\chi = 0 \Rightarrow \mathcal{J}_t(T^c) = \mathcal{J}_{t+1}(T^c+1)$.

C.3 Learning and Risk Aversion Predictions for Shares

The data exploration section explores several patterns of distribution of market shares, conditional on age. This Section provides further justification for the relationship between these patterns and learning / risk aversion.

Prediction 1 (Learning): If initial product quality is uncertain $\sigma_Q > 0$, then learning $1/\sigma^2_A > 0$ implies that expected volatility in product-specific quality estimates (demeaned by true product quality) converge by decreasing with age to zero $Var_{j|a}(\delta^a_j - \bar{\delta}_j) \searrow^{a\to\infty} 0$.

Proof: It is clear from the model setup and Bayes' rule that nonzero precision of the learning process $1/\sigma^2_A > 0$ (in and/or out of trials, so here we suppress that subscript) implies convergence of quality estimates to the true quality $Q^a_j \longrightarrow^{a\to\infty} Q^*_j$ and the convergence of uncertainty about that estimate to zero $\sigma^a_j \searrow^{a\to\infty} 0$ for any product j. Our further claim is that evidence for this learning will be found by looking at

measures of volatility of the mean utilities across products:

$$\begin{aligned}
\lim_{a\to\infty} Var_{j|a}(\delta_j^a - \bar{\delta}_j) &= \lim_{a\to\infty} Var_{j|a}(Q_j^a - Q_j^*) - 0 \\
&= \lim_{a\to\infty} Var_{j|a}\left(\frac{\sigma_{EU}^2}{a\sigma_{EU}^2+\sigma_A^2}\sum_{\tau=1}^{a}\left(Q_j^*+\epsilon_j^\tau\right) + \frac{\sigma_A^2}{a\sigma_{EU}^2+\sigma_A^2}Q_j^1 - Q_j^*\right) \\
&= \lim_{a\to\infty} E_{j|a}\left[\frac{\sigma_{EU}^2}{a\sigma_{EU}^2+\sigma_A^2}\sum_{\tau=1}^{a}\epsilon_j^\tau + \frac{\sigma_A^2}{a\sigma_{EU}^2+\sigma_A^2}(Q_j^1 - Q_j^*)\right]^2 \\
&= \lim_{a\to\infty} \left(\frac{\sigma_{EU}^2}{a\sigma_{EU}^2+\sigma_A^2}\right)^2 a\sigma_A^2 + \left(\frac{\sigma_A^2}{a\sigma_{EU}^2+\sigma_A^2}\right)^2 {\sigma_j^2}^1 \\
&\searrow 0 \qquad (17)
\end{aligned}$$

where the first equality follows from the mean utility specification used in the body of the paper (and relies on linear separability of the quality term). The second substitutes for Q_j^a using Bayes' rule. The third from distributing the Q_j^* term and the fact that the term inside brackets has expectation zero. The fourth from distributing the square and taking expectations. And monotonically decreasing convergence of that quantity to zero is clear. *Q.E.D.*

Prediction 2 (Risk Aversion): If consumers (doctors making decisions on behalf of their patients) are risk averse ($\rho > 0$), then expected product usage, conditional on age, will increase strictly as learning occurs ($E_j[\delta_j^a] \nearrow^{a\to\infty} \overline{Q}$).

Proof: This again follows from the basic structure of the learning model. Consider the quantity of interest:

$$\begin{aligned}
\lim_{a\to\infty} E_j[\delta_j^a] &= \lim_{a\to\infty} E_j[Q_j^a] - \frac{\rho}{2}{\sigma_j^2}^a \\
&= E_j[Q_j^*] - 0 \\
&= \bar{Q} \qquad (18)
\end{aligned}$$

where the first line follows from the mean utility specification. The second line follows from the convergence of $\{Q_j^a\}$ and $\{{\sigma_j^2}^a\}$. And ${\sigma_j^2}^a \searrow^{a\to\infty} 0$ and $\rho > 0$ guarantee the convergence is increasing and strictly so. *Q.E.D.*

D Demand/Learning Estimation: Supplementary Details

D.1 Demand/learning estimation algorithm

The estimation approach is to construct a generalized method of moments estimator that matches the observed market shares in the data (and knowledge of which products are in clinical trials when) to the demand and learning model. The Matlab code for this estimator is available in the electronic archive *code4RegulatingInnovation.zip*. This appendix outlines the main steps of the algorithm.

1. Construct an initial estimator for σ_Q using the empirical equivalent from the Q_j^* from the estimator using age by trial status fixed effects instead of the learning model.
2. Guess initial values for learning precisions $(\sigma_{EU}, \sigma_A, \sigma_{A^c})$ and hospital heterogeneity $(\sigma_H^{des}, \sigma_H^{bms}, \gamma_H)$.
3. Compute the full vector of σ_{jt}^2 implied by σ_Q^2, the learning precision parameters, and which products are in trials when.
4. Least squares then gives an estimator for the linear parameters $(\rho, Q_j^*, \theta^p, \lambda^g)$.
5. Repeat 2-4 until minimize the GMM objective function.
6. Recompute σ_Q using the empirical equivalent from the Q_j^* from this stage.
7. Repeat 2-6 until σ_Q converges.

D.2 Robustness and Alternative Structural Demand Models

Table A1 displays results for several robustness checks on our demand/learning model specification. The first two columns use age fixed effects, interacted with a dummy variable indicating whether the product is in US clinical trials, to provide a less parametric way to capture how demand changes over time with age and trial status. The first column (NL) estimates a simple nested logit model, shutting down any variation in preferences across hospitals. The second column (NLQW) estimates the Quan and Williams (2017) model. The results show how across hospital heterogeneity is important for fitting the data as the criterion function reaches a lower minimum with this added flexibility. As expected, this acts as a selection correction for the product-hospital-months with zero shares, which shifts the product fixed effect estimates.

The third column (NLQWNN) adds the structure of the Normal-Normal learning model in place of the age and trial status fixed effects. There are two primary differences: (1) the

Table A1: Estimates of physician preference and learning model parameters

	NL	NLQW	NLQWNN	$\sigma_A(q_{jt-1})$	H Lags
θ^p (utils/\$1000)	0.21	0.20	0.10	0.10	0.11
	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)
λ^{des}	0.88	0.84	0.81	0.81	0.81
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
λ^{bms}	0.91	0.88	0.82	0.82	0.81
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
σ_H^{des}		0.14	0.19	0.19	0.19
		(0.03)	(0.04)	(0.04)	(0.04)
σ_H^{bms}		0.07	0.18	0.18	0.18
		(0.01)	(0.02)	(0.02)	(0.03)
$\rho \cdot \theta^p$ (1/\$1000)			3.36	3.29	3.66
			(1.70)	(2.30)	(1.99)
$1/\sigma_{EU}^2$			18.82	18.75	18.52
			(2.16)	(2.09)	(2.74)
$1/\sigma_{A^c}^2$			1.73	1.70	1.88
			(0.51)	(0.60)	(0.70)
$1/\sigma_A^2$			0.00	0.00	0.00
			(0.00)	(0.15)	(0.13)
γ_H					
β^q ($\frac{1}{\sigma_A^2}$/100)				0.00	
				(0.00)	
μ^{lag} (months)					0.00
					(0.10)
$\overline{Q}_j$	-2.06	-2.58	-2.37	-2.37	-2.41
	(0.05)	(0.06)	(0.10)	(0.35)	(0.31)
$\sigma_Q\|_{UStrials}$	0.27	0.30	0.26	0.26	0.27
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)
$\sigma_Q\|_{not}$	0.30	0.36	0.34	0.34	0.34
	(0.02)	(0.03)	(0.01)	(0.01)	(0.02)
$age \times UStrials$ FE	Y	Y	N	N	N
$\min(GMM_{criterion})$	101.47	15.47	15.53	15.54	15.54
$RMSE(\xi_{jt})$	0.413	0.401	0.281	0.279	0.297

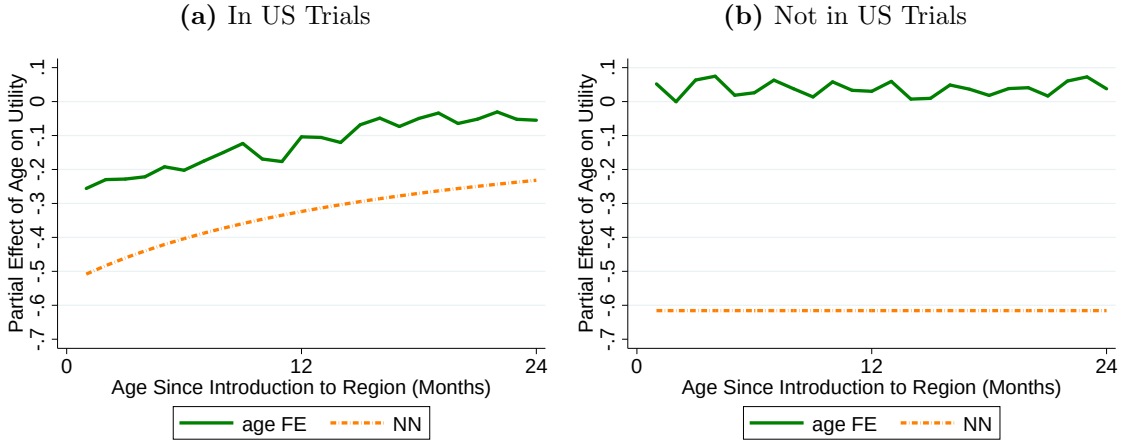
Estimates for demand model $\ln(s_{jt}/s_{0t}) = \lambda^{g_j} \ln(s_{j|_g t}) - \theta^p p_{jht} + Q_j^* - \frac{\rho}{2}\sigma_{jt}^2 + \xi_{jt}$ with separate nests for DES and BMS. $N_{JHT} = 407,191$ product-hospital-months and $N_{JT} = 4,888$ product-months. Standard errors in parentheses and are clustered by month ($N_T = 114$).

learning model parameterization forces learning to be smooth over time (vs. the nonparametric fixed effects); and (2) the learning model uses the rational expectations assumption to link the product fixed effect estimates Q_j^* to how demand evolves with age and trial status. Under rational expectations, the fixed effect estimates must be consistent with the prior distributions $F^{UStrials}(Q)$ and $F^{not}(Q)$, and the precision parameters in the learning model ($1/\sigma_{EU}^2$, $1/\sigma_{A^c}^2$, $1/\sigma_A^2$) link the prior to how the variance and levels moments of product usage evolve with age and trial status.

Figure A10 plots the age fixed effects in NLQW and uncertainty discounts $-\frac{\rho}{2}\sigma_{jt}^2$ in NLQWNN versus age. The left panel shows the products in US trials; the right panel

products not in US trials. The patterns show that: (1) With regards to the smooth parameterization of the learning model, the fit is still quite close to the pattern of the age fixed effects, so the parametric form imposes very little on the data (this can be seen in the figures and also in the $\min(GMM_{criterion})$ (fitting the aggregate usage, aggregate volatility, and hospital moments) and $RMSE(\xi_{jt})$ (of aggregate usage moments only) in the table being close for the two models). (2) The rational expectations assumption allows the model to extract much more information from the data – the prior is now linked to the fixed effects, and so we can infer the amount of learning from EU trials/approval from the gap between that and the variation in usage patterns at $age = 1$. Pinning uncertainty to these two points then allows us to infer the amount of uncertainty remaining as learning does or doesn't occur, and it then requires the product quality estimates Q_j^* to adjust for this. Finally, the NN learning model separates learning/uncertainty and risk aversion parameters – these structural parameters have a clear interpretation that allows for validation of the results, and they allow estimation of counterfactuals where the nature of uncertainty/learning might change due to regulatory changes.

Figure A10: Comparison of estimates from fixed effect and learning models. Plots the estimated discount due to uncertainty versus product age in Normal-Normal learning model $-\frac{\rho}{2}\sigma_{jt}^2$ vs. model with age and $age \times UStrials$ fixed effects.

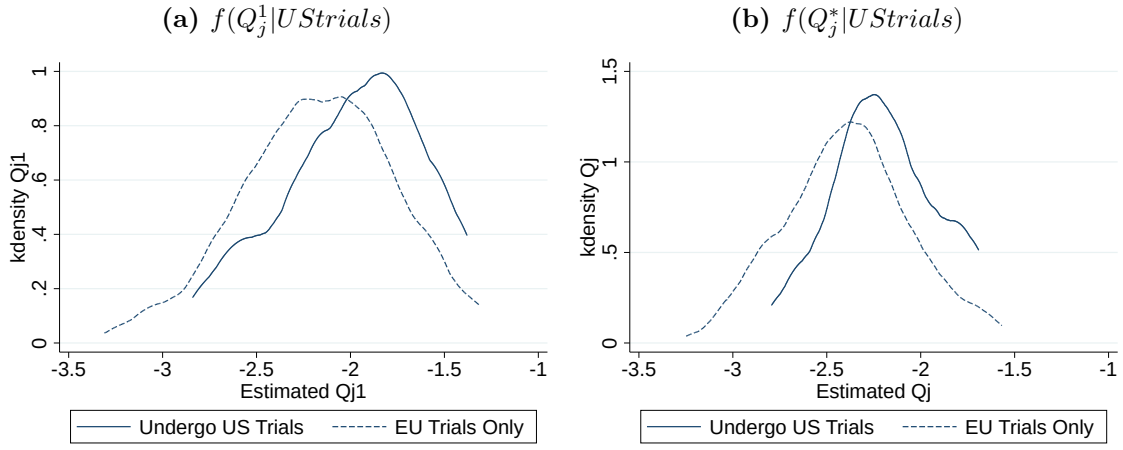


Returning to Table A1, the final two columns show the results for two extensions of our model. Column 4 allows observational learning in the market place to be a function of past demand, $\sigma_A(q_{jt-1})$, in order to check that the lack of market learning that we estimate is not being driven by low usage levels. This is similar in spirit to our "overlapping $Q_{j,age=1}$" test in the reduced form section, and like that test, we do not find any evidence of observational learning for products not in US trials, even for those that are used in large quantities.

The final column, $HLags$, reports the estimates for a model that allows different hospitals to learn with different random lags from each information shock. The goal is to allow for the patterns in the data to potentially be generated by an information diffusion process that is not intrinsically linked to information generation (in which case we might be conflating this diffusion process with the trial information generation process, which is the process the regulator controls most directly). Specifically, each hospital receives each signal with a delay of a number of months drawn from a Poisson(μ^{lag}) distribution, and we estimated μ^{lag} using simulated GMM. Similar to the hospital signal correlation parameter in our preferred model, this lag parameter is identified by the difference in the aggregate vs. hospital level patterns of usage. The estimate does not suggest that learning lag heterogeneity across hospitals explains the patterns in the data.

D.2.1 Estimated product quality distributions

Figure A11: Estimated Quality Distributions. Density plots of the product fixed effect parameters $\{Q_j^*\}$ and market expectations upon EU entry $\{Q_j^1\}$.



One advantage of the GMM algorithm vs. ML (besides the ability to use instruments, which is of course important) is that it allows a nonparametric distribution of product quality estimates. Figure A11 plots the distribution of Q_j^1 (left panel (a)) and Q_j^* (right panel (b)). The results help to validate several of the maintained assumptions. The distributions are not perfectly normal, but appear to be symmetric and reasonably approximated by normals (especially since the tails are inherently difficult to estimate). Also, it does seem that the $UStrials$ distribution may indeed be best thought of as a different distribution with a slightly higher mean and smaller variance. But the distribution does not appear to have a different

shape in a way that would make the two groups difficult to compare or that would suggest an asymmetric information signaling equilibrium.

E Counterfactuals: Supplementary Details

E.1 Partial equilibrium effect of risk: dependence on quality relative to outside option

Table A2 replicates Table 2 in the paper body, allowing the mean qualities to vary by shifting the entire quality distribution by plus or minus a standard deviation of the logit horizontal error term. As referred to in the paper, the effects of decreasing risk become more dramatic as mean quality increases relative to the outside option.

Table A2: The effect of uncertainty on number of stenting procedures, surplus per stent implanted, and expected ex post loss.

		$\sigma_Q =$ 0.312	$\sigma_{a_{EU}=0} =$ 0.185	$\sigma_{T^c=6} =$ 0.160	$\sigma_{T^c=12} =$ 0.143	$\sigma_{T^c=18} =$ 0.131	$\sigma_{T^c=24} =$ 0.121	$\sigma_{T^c=30} =$ 0.113
Baseline $\overline{Q_j^*}$	$1-s_0$ (%)	12.5	24.0	26.4	27.9	29.0	29.7	30.3
		(2.5)	(1.4)	(1.3)	(1.3)	(1.3)	(1.4)	(1.4)
	$\frac{TS}{1-s_0}$ ($)	5776	6103	6184	6238	6276	6304	6327
		(176)	(167)	(167)	(168)	(169)	(170)	(171)
	$E[Q_j^* - Q_{jt}\|j^*]$ ($)	-1096	-560	-429	-348	-292	-252	-221
		(127)	(23)	(37)	(41)	(41)	(39)	(37)
$\overline{Q_j^*} + \sqrt{\pi/6}$	$1-s_0$ (%)	33.8	52.9	56.1	58.0	59.2	60.1	60.8
		(5.0)	(1.8)	(1.6)	(1.6)	(1.6)	(1.6)	(1.6)
	$\frac{TS}{1-s_0}$ ($)	6525	7458	7663	7795	7887	7955	8007
		(301)	(233)	(230)	(232)	(234)	(236)	(238)
	$E[Q_j^* - Q_{jt}\|j^*]$ ($)	-1083	-554	-425	-344	-289	-249	-219
		(127)	(23)	(36)	(40)	(40)	(39)	(37)
$\overline{Q_j^*} - \sqrt{\pi/6}$	$1-s_0$ (%)	3.9	8.1	9.1	9.8	10.2	10.6	10.9
		(0.9)	(0.6)	(0.6)	(0.6)	(0.6)	(0.6)	(0.6)
	$\frac{TS}{1-s_0}$ ($)	5533	5611	5634	5651	5663	5672	5679
		(137)	(138)	(138)	(139)	(139)	(139)	(139)
	$E[Q_j^* - Q_{jt}\|j^*]$ ($)	-1102	-563	-432	-350	-294	-254	-223
		(127)	(23)	(37)	(41)	(41)	(40)	(38)

E.2 Algorithms for computing equilibrium counterfactuals

The Matlab code for the counterfactuals is available in the electronic archive *code4RegulatingInnovation.zip*. This appendix outlines the main steps of the algorithms.

For each potential $T^c = 0, 1, ..., 30$ (and given estimated or comparative static parameter values) we calculate the M and L cases as follows:

M case:

1. Simulate $r = 1, ..., 20$ draws from the posterior signal distribution for each observation, given $\sigma_{jt}(T^c)$ and qualities at EU entry Q_j^1.

2. Restrict sample to products that would be active in each month $\mathcal{J}^M(T^c) = \mathcal{J}_{t+T^c}$.

3. Given $\mathcal{J}^M(T^c)$, use demand/learning models to compute equilibrium quantities and surplus measures for each r.

4. Estimate expected surplus measures using mean of r realizations.

L case:

1. Given M case results and $\chi T^c = 1.6E6 * T^c$, restrict sample to products that would enter, under the naive assumption that firms assume other products enter as if $\chi = 0$ (i.e. single agent entry, assuming competing against $\mathcal{J}^M(T^c) \setminus j$), such that

$$E\left[\sum_{t=\underline{t}_j+T^c}^{\bar{t}_j+T^c}\left(\sum_h q_{jht}(p_{jht} - mc_j)\right)^{1-.01(t-\underline{t}_j-T^c)} |\mathcal{I}^1, \chi_{-j} = 0\right] > \chi_j T^c$$

   , with the expectation computed as the mean across the r simulations of the profits for product j competing against $\mathcal{J}^M(T^c) \setminus j$ (i.e. exactly the expected product profits from the M case). The set of products that would enter is $\mathcal{J}^L(T^c)$.

2. Given $\mathcal{J}^L(T^c)$, use demand/learning models to compute equilibrium quantities and surplus measures for each r.

3. Estimate expected surplus measures using mean of r realizations.

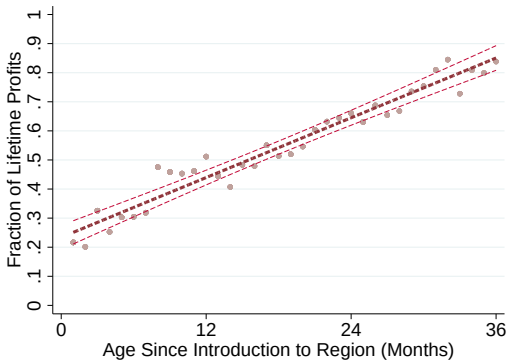
In the Appendix where we also model pricing, equilibrium prices are also computed simultaneously with quantities. Standard errors for counterfactuals are calculated using a nonparametric delete-10 jackknife, blocked at the month level.

E.3 Distribution of Profits Over Product Lifetime and Across Products

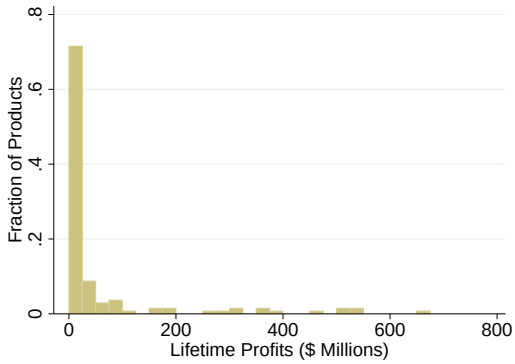
The counterfactual L case with fixed costs of entry require calculation of expected lifetime profits under the assumption that all firms who enter in the EU do enter in equilibrium. This number can be directly acquired from the EU data for the 41 of 109 products that both enter and exit the market during our sample period. However, for the other 68 products whose lifetimes are truncated at the beginning or end, we need to extrapolate.

Figure A12: Distribution of Profits Over Age and Across Products.

(a) Cumulative Profits Over the Product Lifecycle



(b) Distribution of Lifetime Profits over Products



	mean	s.d.	10ptile	median	90ptile	N
Products with full lifetime during sample period:						
Months in sample	21.5	19.8	5	15	47	41
Profit per month ($1000s)	179	612	18	70	211	41
Products with censored lifetime:						
Months in sample	53.1	34.9	10	46	100	68
Profit per month ($1000s)	1,347	2,119	41	262	4,067	68
Distribution of lifetime profits (extrapolated where necessary):						
Lifetime Profit ($M)	72.6	141.5	1.3	10.5	304.3	109

We perform this extrapolation by estimating the percent of cumulative lifetime profits the average product has earned at each age. We then use this percent to extrapolate the missing profits, for whatever age at which the truncation occurred. We do this unconditionally on any covariates besides age.

In our counterfactuals, we hold take time from entry until exit as exogenously given, and we use this same extrapolation to calculate expected lifetime profits for any product that has not exited by the end (or entered by the beginning) of the sample period in the counterfactual. Our counterfactual estimates are robust to a variety of approaches to this extrapolation. This is in part because the extrapolation is typically for the beginning or end of lifetime tail of product profits, so that lifetime profit projections are not very sensitive to the method we choose. Further, the products that are marginal in the sense that they exit as entry costs increase, are also marginal in their contribution to consumer surplus.

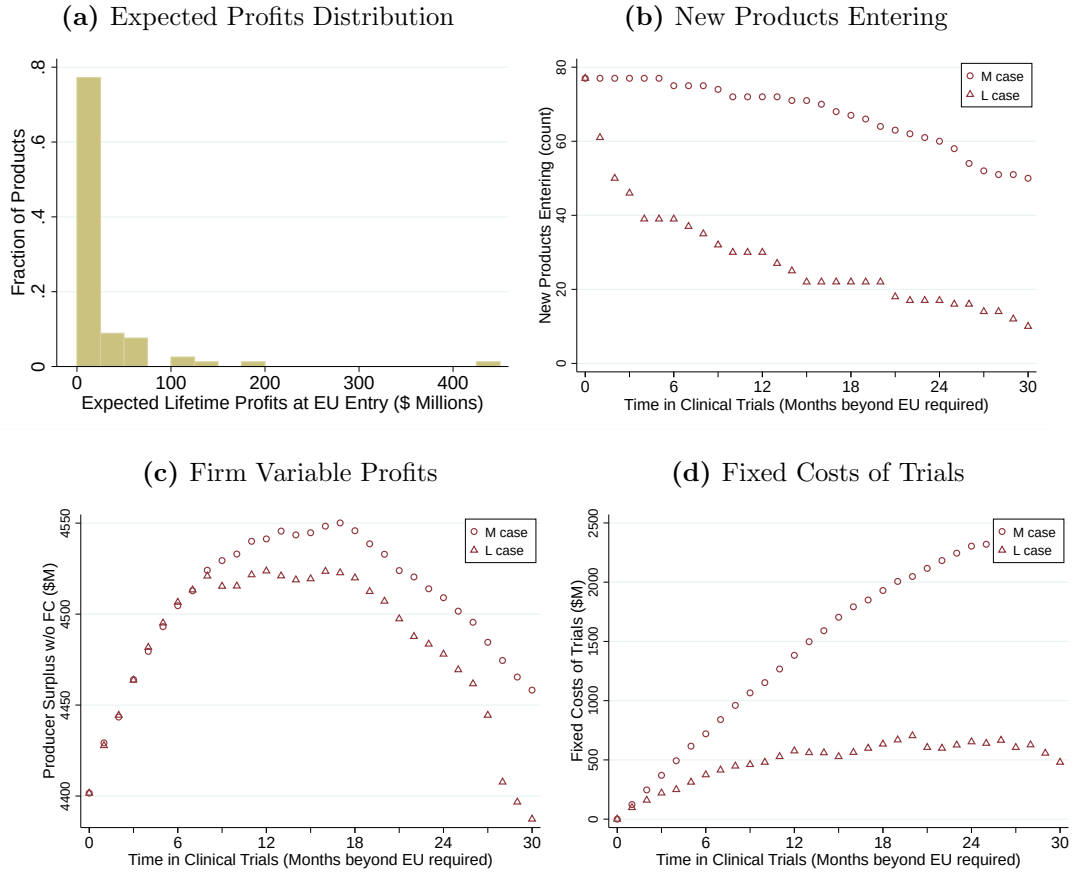
Also notable here is the distribution of estimated lifetime profits in the last row of the table at the bottom of Figure A12, which makes it clear that many products with quite low profitability enter the EU. This supports our assumption that the products in the EU market represent a reasonable approximation to the set of products developed that firms

might consider testing and bringing to market.

E.4 Additional Results for $E[\pi_j(T^c)]$,$\mathcal{J}(T^c)$, $PS(T^c)$, $FC(T^c)$

Each of our counterfactuals fully characterizes outcomes (under our M and L cases), but we only show results for CS and TS in the body of the paper. Figure A13 shows: (a) the distribution of expected profits across products given their information sets upon EU entry and beliefs about other product entry, (b) the number of new products entering, (c) variable producer surplus, and (d) fixed cost expenditures on trials.

Figure A13: Additional Results for $E[\pi_j(T^c)]$,$\mathcal{J}(T^c)$, $PS(T^c)$, $FC(T^c)$ Computed for our baseline scenario of a change in policy in January 2004 (no spillovers).



Several highlights from these results are discussed in the body of the paper. The full results here provide a more complete picture of how delay of entry from longer trials (subfigure (b), M case) affects the choice set of products available for our period of study. Additionally, they show how fixed costs of additional trials (subfigure (d)) plus the distribution across products of expected profits (subfigure (a)) can further combine to limit the set of products

available (subfigure (b), L case). The bulk of this decreased entry due to larger fixed costs of testing comes from the 76 percent of products whose expected lifetime profits are below $25M.

E.5 Robustness of Counterfactual Estimates to Modeling Assumptions

The paper makes a number of modeling assumptions to facilitate the counterfactual computations. Here we explore robustness of the results to the extent to which the government/regulator discounts future surplus, and to the assumptions on marginal costs and prices. Table A3 shows the results.

The assumed amount of discounting has rather large effects on the amount of surplus generated by optimal testing. The reason for this is that it takes time for many products used in the market to be products that have benefited from further testing, whereas the costs of lack of access to new products are more uniformly distributed in time. Thus more heavily discounting the future reduces the benefit from further testing. However, this effect is primarily borne out in the level of surplus gain from optimal testing, with the amount of increased testing under optimal policy changing little with the discounting applied.

Table A3: Robustness: Discounting, Marginal Costs, and Pricing. Top row repeats our baseline results (discounted at one percent nominal per month, $mc_j = .50 * \min(p|g_j)$, and prices held fixed). Subsequent rows modify these assumptions as indicated in the first column.

State of Market at Policy Change	$\Delta CS(T^{c*})$ (%)		T^{c*}_{CS} (months)		$\Delta TS(T^{c*})$ (%)		T^{c*}_{TS} (months)	
	(L)	(M)	(L)	(M)	(L)	(M)	(L)	(M)
Jan 2004	5.2	6.7	16	17	3.9	6.3	16	17
	(1.8)	(1.9)	(4)	(2)	(1.6)	(1.9)	(4)	(2)
Jan 2004, Govt. Discount 7% real	5.8	7.3	18	18	4.5	6.9	16	17
	(2.1)	(2.2)	(4)	(3)	(2.0)	(2.1)	(4)	(3)
Jan 2004, Govt. Discount 3% real	6.6	8.2	19	19	5.2	7.7	16	19
	(2.1)	(2.2)	(4)	(3)	(2.0)	(2.1)	(4)	(3)
Jan 2004, No Govt. Discount	7.7	9.5	19	19	6.2	8.9	16	19
	(2.1)	(2.2)	(4)	(3)	(2.0)	(2.1)	(4)	(3)
Jan 2004, $mc_j = .25 * \min(p\|g_j)$	5.4	6.7	13	17	3.9	6.3	13	17
	(1.8)	(1.9)	(3)	(2)	(1.6)	(1.9)	(4)	(2)
Jan 2004, $mc_j = .75 * \min(p\|g_j)$	5.2	6.7	16	17	3.8	6.3	13	17
	(1.8)	(1.9)	(4)	(2)	(1.6)	(1.9)	(4)	(2)
Jan 2004, Pricing Modeled	5.2	6.6	13	17	4.0	6.7	13	18
	(1.7)	(1.9)	(3)	(2)	(1.6)	(1.8)	(4)	(3)

$N_{JHT} = 407,191$ product-hospital-months and $N_{JT} = 4,888$ product-months. Standard errors, clustered by month ($N_T = 114$) using a delete-10 block jacknife, in parentheses.

Turning to marginal costs and prices, our primary specification holds prices fixed in the counterfactual and assumes marginal costs at half the minimum price observed within each

product group ($g \in \{bms, des\}$) for the purpose of computing firm expected and realized profits. The bottom rows of Table A3 show that results change very little, if at all, when we change marginal costs to be either one quarter or three quarters the group minimum, or when we instead model pricing and use that model to estimate marginal costs and compute new equilibrium prices in the counterfactuals.

E.5.1 Pricing: model, identification, estimation, and results

Here we describe the pricing model and estimation for the robustness check where we model pricing explicitly.

In the EU, device pricing practices vary somewhat across countries and hospitals, but are typically negotiated between manufacturers and either the hospital or some regional body responsible for procurement for a set of hospitals. For this Appendix, we model this process using a static Nash Equilibrium of Nash Bargaining models for each period, following the theory developed in Horn and Wolinsky (1988) and Collard-Wexler et al. (2014) and recent empirical work by Crawford and Yurukoglu (2012), Grennan (2013), and Gowrisankaran et al. (2014). These approaches assume that prices maximize the bilateral Nash product

$$\max_{p_{j\mathcal{H}t}} \left(\sum_{h \in \mathcal{H}} \pi_{j\mathcal{H}}(\mathcal{J}_t, I_{ht}, p_{j\mathcal{H}t}) \right)^{b_{jt}(\mathcal{H})} \left(\sum_{h \in \mathcal{H}} CS_{ht}(\mathcal{J}_t, I_{ht}, p_{j\mathcal{H}t}) - CS_{ht}(\mathcal{J}_t \setminus \{j\}, I_{ht}, p_{j\mathcal{H}t}) \right)^{b_{\mathcal{H}t}(j)} , \tag{19}$$

for each $j \in \mathcal{J}_t$ in each market (group of hospitals in bargaining unit $\mathcal{H}$ in each month t), taking other prices in the market $\{p_{k\mathcal{H}t}\}_{k \in \mathcal{J}_t}$ as given. Here $\pi_{j\mathcal{H}} := q_{jht}(p_{j\mathcal{H}t} - mc_j)$ are manufacturer variable profits at marginal cost, mc_j, for each device. CS_{ht} is the hospital level consumer surplus. The parameters $b_{jt}(\mathcal{H})$ and $b_{\mathcal{H}t}(j)$ are the Nash bargaining weights, determining the extent to which equilibrium prices weight manufacturer profit (minus its outside option of not producing the stent) versus hospital surplus (minus its outside option of optimal choice for each patient from a choice set that excludes j).[62]

With the demand and learning model parameter estimates in hand, we turn to estimating the parameters from the bargaining model between hospitals and device manufacturers. To take the bargaining model to the data, we follow Grennan (2013) and rewrite (19) as

$$p_{j\mathcal{H}t} = mc_j + \frac{b_{jt}(\mathcal{H})}{b_{jt}(\mathcal{H}) + b_{\mathcal{H}t}(j)} \sum_{h \in \mathcal{H}} \frac{M_{ht}}{\sum_{h \in \mathcal{H}} M_{ht}} AV_{jht}(\mathcal{J}_t, I_{ht}, p_{j\mathcal{H}t}, mc_j; \theta^D) , \tag{20}$$

[62] Assuming constant returns to scale in distribution and manufacturing on the margin at $\sum_{h \in \mathcal{H}} q_{jht}$. We also follow previous work in maintaining the Nash-like assumption that other prices remain the same in the case of disagreement, which is consistent with "passive beliefs" in the theory literature that provides noncooperative foundations for this concept.

where the added value of product j to hospital h is defined as

$$AV_{jht} := \left(1 + \frac{\partial q_{jht}}{\partial p_{jht}} \frac{p_{j\mathcal{H}t} - mc_j}{q_{jht}}\right) \frac{CS_{ht}(\mathcal{J}_t, I_{ht}, p_{j\mathcal{H}t}) - CS_{ht}(\mathcal{J}_t \setminus \{j\}, I_{ht}, p_{j\mathcal{H}t})}{q_{jht}} + p_{j\mathcal{H}t} - mc_j \tag{21}$$

To estimate the structural parameters, the bargaining weights and the marginal costs, we use our utility model estimates. We calculate the substitution patterns by simulating η_{jht}, the elasticities across hospitals as defined in Equation (12), over the distribution of hospital level unobservables $f_H(\xi_{jh}; \sigma_H^{g_j})$ (suppressing dependence on hospital-specific learning for simplicity since $1/\sigma_A^2$ and γ_H are both estimated to be zero). Similarly, we use the consumer surplus equation derived from the utility function in Equation (13) to compute the buyer surplus portion of the added value.

Because of physicians may be imperfect agents for patients and/or hospitals, estimated physician price sensitivity measures may not reflect the correct scaling for measuring hospital and/or consumer surplus. For this reason, we deviate slightly from the standard demand estimation approach in that we add a scaling coefficient to relate consumer surplus derived from consumer utility maximization to estimates of the dollar value of quality adjusted life years obtained in clinical studies. We normalize the total surplus per stenting procedure to \$5,000, which is the approximate median of the estimated dollars in quality adjusted life years from the procedure relative to a coronary artery bypass graph surgery, a more invasive alternative to receiving angioplasty and a stent.[63] This alternative scaling is only for translating welfare measures into dollars—we continue to use the estimated θ^p in quantity and elasticity calculations, as revealed preference indicates this is the level of price sensitivity that best fits the demand patterns in the data.

In addition to the standard set of issues that the bargaining literature has identified in estimating marginal costs and bargaining parameters, we face two additional challenges. First, the challenge in estimating demand at the hospital level means that our demand estimates only provide the distribution of added values across hospitals, not the hospital-specific added values. Second, because we only observe a sample of hospitals, we do not have added value measures for all hospitals in a group $\mathcal{H}$ in cases where our hospitals may negotiate as part of a group.

We address both of these supply estimation challenges by aggregating our estimation strategy across hospitals to the product-month level. Otherwise, we follow Grennan (2013,

[63] Among studies reported in the Cost Effectiveness Analysis Registry (https://research.tufts-nemc.org). We could also scale into dollars using the standard approach of the inverse of the price coefficient $\frac{1}{\theta^p} = 10,482$, which would approximately double all related consumer welfare estimates.

2014): We assume the econometric error enters relative bargaining weights multiplicatively:

$$mc_j := \mu_g^C \quad ; \quad \frac{b_{jt}(\mathcal{H})}{b_{\mathcal{H}t}(j)} := \frac{\beta_j}{\beta_{\mathcal{H}}} \nu_{j\mathcal{H}t} \tag{22}$$

where μ_g^C allows marginal cost to vary across BMS/DES, and β_j are product-specific bargaining parameters to be estimated. Substituting into Equation (20), rearranging and taking logs to obtain a linear equation in the unobservable, and aggregating over hospitals gives the equations at the product-month level that we take to the data:

$$\sum_h \frac{M_{ht}}{M_t} \ln\left(\frac{p_{jht} - \gamma_j}{AV_{jht}}\right) = \ln(\beta_j) + \underbrace{\sum_h \frac{M_{ht}}{M_t}(-\ln(\beta_h) + \ln(\nu_{jht}))}_{\tilde{\nu}_{jt}} \tag{23}$$

The parameters (β_j, γ_j) can then be estimated by a GMM algorithm, assuming $E_{jt}\left[\tilde{\nu}_{jt}|Z_{jt}\right] = 0$ for a set of instruments including product-specific constants and $\frac{\partial \tilde{\nu}_{jt}}{\partial \gamma_j}$.

Table A4 presents the parameter estimates from the bargaining model. As we estimate the parameters at the product level, we present the means and standard deviations of those estimates. The first two variables are the elasticity and average value parameters that come from the demand model that feed into the bargaining model. The price elasticity is somewhat higher from DES stents and the typic BMS stent adds $1155 of value while the average DES stent adds significantly more value at $1424.

Table A4: Structural parameter estimates for pricing model

	η^p_{jht}		AV_{jht} ($)		mc_j ($)		$\frac{b_{jt}(\mathcal{H})}{b_{jt}(\mathcal{H})+b_{\mathcal{H}t}(j)}$	
	mean	sd	mean	sd	mean	sd	mean	sd
BMS	-0.25	0.11	1155	118	87	-	0.41	0.12
	(0.06)	(0.06)	(172)	(41)	(124)		(0.06)	(0.04)
					(124)		(0.06)	(0.04)
DES	-0.42	0.14	1424	224	361	-	0.60	0.14
	(0.32)	(0.11)	(312)	(60)	(117)		(0.08)	(0.04)
					(117)		(0.08)	(0.04)

$N_{JHT} = 407,191$ product-hospital-months and $N_{JT} = 4,888$ product-months. Standard errors clustered by month ($N_T = 114$) using a delete-10 block jacknife in parentheses.

The next two sets of variables are parameter estimates from the bargaining model. The marginal cost estimates align with expectations and prior literature. BMS cost an average of $87 to produce while DES are more costly at $361. Finally, the last two columns present the estimates of the relative bargaining weights, $\frac{b_{jt}(\mathcal{H})}{b_{jt}(\mathcal{H})+b_{\mathcal{H}t}(j)}$. The results imply that for BMS, hospitals retain the majority of the surplus (manufacturers obtain 41 percent on average) from the implantation of the device, with a modest amount of variance across products.

However, for the newer DES technology, on average, the manufacturers receive the majority of the surplus (60 percent).[64]

E.6 Additional Comparative Static Counterfactual Results

Table A5 provides results for a more complete set of comparative static parameter values than those shown in the body of the paper in Table 3. The key takeaways were discussed in the body of the paper. We supply the full set of numerical values here for the interested reader. We find these useful in considering the robustness of our estimated results for coronary stents 2004-13, and also for thinking about how these results might extrapolate to other product categories with different primitives. Perhaps the most non-obvious and important qualitative result is the nonlinearity of optimal testing *length* with respect to the precision of information generated by testing. As mentioned in the body of the paper, this thought experiment is tightly linked to pushes for validating more intermediate/surrogate endpoints (e.g. for stents this would be measuring 6 month "loss" in the target vessel diameter instead of revascularization or mortality endpoints) that could allow trials to generate more information with smaller sample sizes.

[64]The DES bargaining parameter is nearly double that in Grennan (2013) in the US 2004-07 subsample, but this corresponds closely to the magnitude of our alternative scaling of consumer surplus, and may also be related to lower reimbursements to hospitals and the different competitive environment for DES in the EU relative to the US.

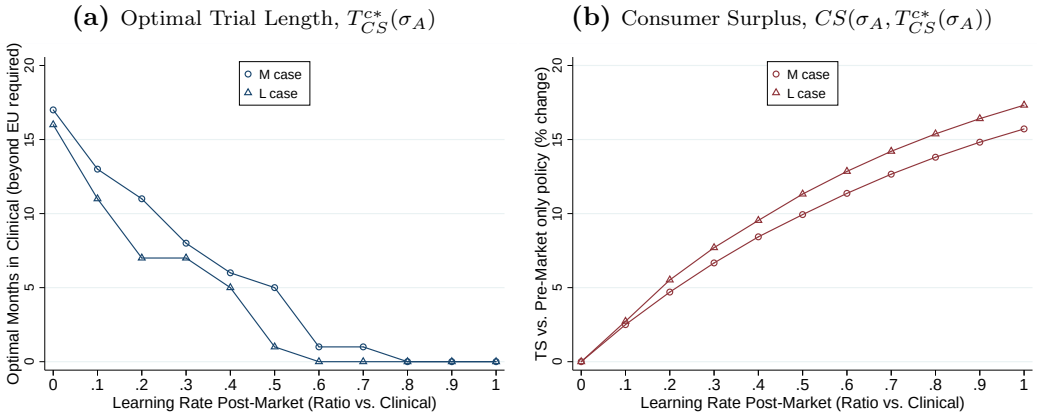
Table A5: Sensitivity of Optimal Regulation to Key Parameters: Full results for all parameter values we have explored.

State of Market at Policy Change	$\Delta CS(T^{c*})$ (%)		T^{c*}_{CS} (months)		$\Delta TS(T^{c*})$ (%)		T^{c*}_{TS} (months)	
	(L)	(M)	(L)	(M)	(L)	(M)	(L)	(M)
Jan 2004	5.2	6.7	16	17	3.9	6.3	16	17
	(1.8)	(1.9)	(4)	(2)	(1.6)	(1.9)	(4)	(2)
Jan 2004, FC * .1	6.6	6.7	17	17	5.9	6.3	17	17
	(1.9)	(1.9)	(3)	(2)	(1.8)	(1.9)	(3)	(2)
Jan 2004, FC * .2	6.5	6.7	17	17	5.7	6.3	17	17
	(1.9)	(1.9)	(3)	(2)	(1.8)	(1.9)	(3)	(2)
Jan 2004, FC * .5	6.2	6.7	14	17	5.0	6.3	13	17
	(1.9)	(1.9)	(3)	(2)	(1.8)	(1.9)	(3)	(2)
Jan 2004, FC * 2	3.9	6.7	9	17	2.4	6.3	7	17
	(1.6)	(1.9)	(4)	(2)	(1.3)	(1.9)	(6)	(2)
Jan 2004, FC * 5	1.7	6.7	7	17	0.0	6.3	0	17
	(1.1)	(1.9)	(3)	(2)	(0.4)	(1.9)	(2)	(2)
Jan 2004, FC * 10	0.0	6.7	0	17	0.0	6.3	0	17
	(0.3)	(1.9)	(1)	(2)	(0.0)	(1.9)	(0)	(2)
Jan 2004, σ_Q * .5	0.1	0.2	1	5	0.0	0.1	0	1
	(0.3)	(0.5)	(2)	(3)	(0.1)	(0.5)	(1)	(4)
Jan 2004, σ_Q * .75	2.2	3.1	8	13	1.3	2.9	6	13
	(1.3)	(1.5)	(4)	(3)	(1.0)	(1.4)	(4)	(3)
Jan 2004, σ_Q * .9	4.0	5.3	11	17	2.7	5.0	11	17
	(1.6)	(1.8)	(4)	(3)	(1.4)	(1.7)	(4)	(3)
Jan 2004, σ_Q * 1.1	6.4	7.9	18	18	5.0	7.5	16	17
	(1.9)	(2.1)	(3)	(3)	(1.8)	(2.0)	(4)	(2)
Jan 2004, σ_Q * 1.33	8.6	10.2	19	19	6.9	9.6	17	19
	(2.1)	(2.2)	(3)	(3)	(2.0)	(2.1)	(4)	(3)
Jan 2004, σ_Q * 2	12.1	13.7	19	19	10.2	13.1	18	19
	(2.3)	(2.4)	(3)	(3)	(2.2)	(2.3)	(4)	(3)
Jan 2004, ρ * .5	0.8	1.2	6	6	0.2	1.1	3	6
	(0.5)	(0.7)	(2)	(2)	(0.3)	(0.6)	(2)	(2)
Jan 2004, ρ * .75	2.7	3.6	9	13	1.8	3.3	6	13
	(1.2)	(1.3)	(3)	(3)	(1.0)	(1.3)	(3)	(3)
Jan 2004, ρ * 1.33	9.6	11.2	19	19	7.7	10.6	18	19
	(2.6)	(2.7)	(3)	(3)	(2.4)	(2.6)	(5)	(3)
Jan 2004, ρ * 2	16.3	20.0	20	26	14.1	19.1	20	26
	(4.1)	(4.3)	(3)	(2)	(3.9)	(4.1)	(3)	(3)
Jan 2004, $1/\sigma^2_{A^c}$ * .2	0.0	0.1	0	1	0.0	0.1	0	1
	(0.1)	(0.5)	(2)	(4)	(0.0)	(0.4)	(0)	(4)
Jan 2004, $1/\sigma^2_{A^c}$ * .5	1.5	2.6	7	13	0.7	2.4	6	13
	(1.2)	(1.5)	(4)	(4)	(0.9)	(1.4)	(5)	(4)
Jan 2004, $1/\sigma^2_{A^c}$ * .75	3.5	4.9	16	17	2.2	4.6	8	17
	(1.6)	(1.8)	(4)	(3)	(1.4)	(1.7)	(5)	(3)
Jan 2004, $1/\sigma^2_{A^c}$ * 1.33	6.9	8.4	13	17	5.5	8.0	16	17
	(1.9)	(2.1)	(3)	(2)	(1.8)	(2.0)	(3)	(2)
Jan 2004, $1/\sigma^2_{A^c}$ * 2	9.6	10.7	13	17	7.9	10.2	13	17
	(2.1)	(2.2)	(3)	(3)	(1.9)	(2.1)	(3)	(3)
Jan 2004, $1/\sigma^2_{A^c}$ * 5	14.2	15.3	10	12	12.6	14.5	9	11
	(2.3)	(2.3)	(2)	(2)	(2.1)	(2.2)	(2)	(2)

$N_{JHT} = 407,191$ product-hospital-months and $N_{JT} = 4,888$ product-months. Standard errors, clustered by month ($N_T = 114$) using a delete-10 block jacknife, in parentheses.

E.7 Post-Market Surveillance and Consumer Surplus

Figure A14: The Value of Post-Market Surveillance (Consumer Surplus): Plots of optimal trial length (left panel (a)) and total surplus (right panel (b)) as observational learning precision $1/\sigma_A^2$ varies from zero to the clinical trial precision $1/\sigma_{A^c}^2$.



As noted in the paper, the CS metric generates tighter bounds and greater returns to optimal pre-market policy. The CS metric is of special interest in the post-market surveillance case because it is derived from only the risk-access tradeoff, not the fixed costs savings from less trials. As a result, optimal pre-market trial length decreases less quickly with post-market learning under the CS criterion.