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

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 Horta csu, Robin Lee, Scott Stern, Mauricio Varela, and audiences at the BEA, Chicago, Clemson, EEI8 Conference, FDA, FTC, Georgia Tech, Harvard, HEC Montreal Summer IO Conference, iHEA, Johns Hop- kins, Kellogg Conference on Healthcare Markets, Minnesota, NBER IO and Health Care Summer Institutes, NUS, Ohio State, Penn State, Rice/Houston, Texas, UCLA, Wharton, and Yale for helpful comments. The University of Toronto, Wharton Dean's Research Fund and Mack Center for Innovation, and the Indian School of Business provided financial assistance. Ellie Prager, Mike Punzalan, Emi Terasawa, and Christine Yang 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 March 2016
JEL No. I11,L11,L51

ABSTRACT

We propose a model where new product quality is uncertain, but market participants learn over time. Regulation balances information's role in reducing consumer risk versus reducing access to innovation. Using new data and variation between EU and US medical device regulations, we document patterns consistent with the model and estimate its parameters. We find: (1) without information from testing, risk would severely inhibit usage; (2) US policy maximizes total surplus in our estimated model while the EU could gain 20 percent with more pre-market testing; and (3) more "post- market surveillance" could increase surplus 24 percent.

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

Robert Town
Health Care Management Department
The Wharton School
University of Pennsylvania
3641 Locust Walk
Philadelphia, PA 19104
and NBER
rtown@wharton.upenn.edu

1 Introduction

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

This paper uses new, detailed data and exploits exogenous regulatory differences between the US and EU to identify the impact of product quality information on market outcomes for medical devices. Among its many duties, 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 safety of the product. For the Class III medical devices we study, this difference is material.[2] 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 seek FDA approval, if they decide to enter the US at all.

The differences between the US and the EU in the medical device approval process have

[1] See, for example in electronics "European Environmental Rules Propel Change in U.S.", The New York Times, July 06, 2004; 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 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, world-wide sales of coronary stents exceeded $7.9 billion with the vast majority of those sales occurring in the US and the EU.

led to calls for reform in both regions. In the US, the FDA has faced attacks from both sides, with some claiming that a slower, tougher approval process is crippling innovation; and others claiming that the approval process is too lax, allowing too many dangerous devices into the market.[3] 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.[4]

Despite the broad importance, empirical research on 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 pre-market testing from manufacturers and as a consequence is both faster and less costly than the US process for any given device. We describe this difference in detail and argue it is due to historical political processes and not correlated with market demand for devices. As a result, we are able to observe outcomes for the same 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 subject to US trials. The key additional identifying assumption (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 data and a corresponding empirical framework that can then quantify the returns from increased information relative to the costs of decreased access. To address this challenge we acquire monthly data on product-level prices, quantities, and diagnostic procedures in the US and EU. Paired with the variation in participation in US clinical trials, we show how revealed preference arguments imply that such data capture the state of market knowledge and learning that changes that knowledge, and how these in turn affect consumer choice and thus welfare. This approach is similar in spirit to the pioneering work of Peltzman (1973) where he uses pre-/post- analysis to argue that the 1962 Kefauver-Harris Amendments to the Food, Drug, and Cosmetic Act (which require clinical trials for pharmaceuticals prior to their introduction to the market) harmed consumers by reducing access to drugs without increasing product information. Our ability to observe EU usage patterns for devices both in and out of US trials allows us to go further in directly estimating information generated and how consumer demand responds to information.[5]

[3]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.

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

[5]Our approach contrasts with studies of the FDA approval process which have used product introductions and withdrawals instead of market usage data (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

We begin the analysis by constructing a theoretical model to capture the risk/access tradeoff inherent in these policy debates, generate empirical predictions to take to the data, and provide a mapping from the data to welfare estimates. In the model, products are invented with uncertain quality, a regulator requires some level of premarket testing that a firm must invest in if it wishes to bring a product to market, and consumers can learn about product quality over time through these trials and/or through observational learning in the marketplace. Uncertainty and learning are symmetric across manufacturers, consumers, and the regulator.[6] We focus on the case where the rate of learning in premarket clinical trials is greater than the rate of learning after market entry. This introduces a tradeoff where more regulation leads to more information and less risk; but it also delays access and raises entry costs for new products.

Our empirical work begins by first documenting multiple reduced-form patterns in the data and linking these patterns to the predictions of our model. The EU enjoys greater access to the newest medical technologies, while also bearing greater risk by allowing entry of a wider range of device qualities, earlier in each device's lifecycle. On average, US physicians have 11 stents available to implant while their EU counterparts have 39 to from which to choose. Conditional on the product entering the US, EU physicians have access to the product 10 months earlier. The average share (relative to the outside good) of US entering products is much higher, and importantly, is flat over the product lifecycle. Average volatility in shares over time in the US is low and flat. Through the lens of our model, these patterns are consistent with no market learning regarding product quality after product launch in the US. In contrast, after product launch in the EU, the average volatility begins high and decreases over time, while average share increases over time, consistent with the predictions of our model if in fact there is learning.

We then establish that the patterns we observe in the EU market are driven entirely by information spillovers from US clinical trials. By focusing on within-product variation and comparing the same products launched in different regions (the US and EU) and also EU patterns for products that are and are not undergoing US trials (including an overlapping

more precise information on product quality, on which negative tail events such as recalls provide little information.

[6]The symmetric information assumption 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. We believe that many markets with published testing results and informed consumers 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). We also provide an empirical test that fails to reject the symmetric information assumption in our data.

support sample that have similar estimated qualities upon EU introduction), we are able to rule out alternative mechanisms such as selection on product quality, non-learning models of product diffusion, learning from observational use vs. learning from clinical trials, and selection on product uncertainty.

In order to develop welfare measures and address policy questions regarding optimal regulation, we estimate the structural parameters of the model. Using the EU market data and variation in product enrollment in US trials, we estimate the distribution of product qualities and risk as well as the speed of learning and preferences of consumers in the marketplace, and we validate our estimates with comparisons to a number of outside data sources. Consistent with the reduced form evidence, the demand parameter estimates indicate that FDA required clinical trials generate significant information while there is practically no observational learning via experience in the EU marketplace. Combined with product quality estimates that indicate significant variation in stent quality, this implies the returns to early product testing are large for stents during this time period. A partial equilibrium analysis suggests that without the EU safety testing the risk faced would be enough to deter a significant portion of consumers from choosing stenting versus the next best alternative, and the further US efficacy testing substantially decreases the risk of using an inferior product.

We then consider optimal regulatory policy by combining these estimates of the effects of information on risk with our estimates of product quality and the value of access to the newest products over time. We develop simple-to-compute cases that bound total surplus as a function of regulatory policy and firm behavior, and use these bounds to generate a partially identified set of optimal regulatory policies.[7] The results imply that total surplus is maximized when the average premarket clinical trial is at least seven months longer than the current EU requirements. Around the optimum, total surplus is relatively insensitive to the time spent in premarket testing, implying that, for stents 2004-13, US regulatory policy is statistically equivalent to the policy that maximizes total surplus in our estimated model. While the EU benefits from free-riding on the information generated by US trials, we estimate EU surplus could be increased by 20 percent by requiring more pre-market testing for stents.

Our final piece of analysis examines optimal policy under counterfactual worlds with greater "post-market surveillance". This idea, which is a centerpiece of the FDA reforms proposed in the 21st Century Cures legislation, has a straightforward logic: increased learning post-market approval could maintain risk-reduction while lowering pre-market requirements, thus decreasing entry costs and lags. Our estimate of no observational learning in the EU

[7] A full supply side model requires frontier work in bargaining and dynamic entry modeling. Our approach avoids the need for a full supply model by using simple assumptions on supply behavior and a partial identification approach (pioneered by Manski (2003) and more recently in Pakes et al. (2015) and others to estimate primitive parameters) on the set of counterfactual outcomes based on point identified primitives.

for coronary stents is perhaps not surprising, given that there is no systematic data collected that links stents used to clinical outcomes; but we find that if post-approval learning rates could approach those we observe from clinical trials at a comparable cost, the benefits from such a policy change are substantial.[8] In the extreme case where post-approval learning is fully informative and not too costly, the optimal policy is to require no pre-approval trials at all, which would yield a welfare increase of 24 percent in our estimated model.

Because our attention is on the issues of information generation, risk, and access, our analysis should be interpreted as holding the other roles of the regulator as fixed.[9] Though we do model entry decisions conditional on arrival of an innovation, we take the arrival process for new products as exogenous and abstract away from feedback effects from the device regulatory regime to the incentives to invest in the innovative process itself. In this sense our approach and results are related to the larger literature on quality information disclosure (e.g. Dranove and Jin 2010), but whereas the focus there is typically on the difference between no disclosure and some disclosure, our focus is on the *amount* of information required, given a basic disclosure regime. We also measure how this information affects market structure and welfare through entry and consumption decisions.

We believe this focus on market structure in our paper is complementary to recent empirical research on other regulatory tools that affect product entry incentives, such as patent breadth and length (Budish, Roin, and Williams 2015) and price regulations (Kyle 2007; Filson 2012). Whereas the focus of that literature has been on innovative activity with stylized monopoly market structures, we show that the welfare impact of regulation on market structure and buyer decisions can be large as well. New medical technologies with uncertain quality can only achieve their welfare potential if the necessary studies to document the product's clinical performance are performed.

More broadly, our work builds on recent empirical research on optimal regulation (Timmins 2002; Seim and Waldfogel 2013; Miravete, Seim, and Thurk 2014) and consumer learning (Roberts and Urban 1988; Erdem and Keane 1996; Ackerberg 2003; Crawford and Shum 2005; Ching 2010), and to our knowledge is the first to combine these two. This combination is essential in allowing us to build on the pioneering work of Peltzman (1973) in measuring the impact of regulatory information requirements. As we build on established models and use frequently available data, we provide an approach that others might find useful to the study of regulation and product approval/certification in other markets.

[8] In the US, the FDA has recently introduced a Unique Device Identifier system that could facilitate better post-market data collection. However, there is currently debate regarding whether UDIs will be added to patient claims databases.

[9] In addition to determining safety and efficacy requirements, the FDA also plays an important role in defining and certifying testing procedures themselves. The FDA also regulates post-market introduction activities: ensuring that only information established via these tests is allowed on product labeling and in marketing and sales communications; and monitoring good manufacturing processes.

Our analysis of the impact of different regulatory regimes not only speaks to the economics of information and product quality regulation, but also informs policies currently under debate with potentially large welfare consequences. The amount of economic activity regulated by the FDA and the Notified Bodies is significant. In the US the medical device market sales exceeded $150B in 2010 or 6 percent of total national health expenditures and approximately $130B (7.5 percent) in the EU.[10] 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 their direct impact on commerce. We conclude the paper by discussing how one might approach applying our analysis and results to products beyond coronary stents.

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.

Before detailing these regulatory differences, however, it is useful to keep in mind some basic facts about the structure of decision making and 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 importantly the financial benefits to the physician are unrelated to the specific brand of device used. Physicians typically have strong preferences about which brand of device 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 effect 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 or hospital systems.

For the purposes of this study, the most important features of the stent market to note are the constant innovations over time in terms of both vertical quality advances that make

[10] Donahoe and King, 2012; Medtech Europe, 2013

similar yet better products for the mass market, and also horizontally differentiated products designed to address less common niche markets such as small vessel and bifurcated lesions. 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 (TCT each October, ACC in March, and ESC 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 passage of the Medical Device Amendments Act of 1976. This law established the regulator pathway for medical devices in the US, placing oversight authority within the FDA. The criteria the FDA is mandated to use is "safe and effective." The Act established three classification of devices (I, II and III) which are assigned based on the perceived risks associated with using the device. Class III devices are defined as those used in in supporting or sustaining human life, of substantial importance in preventing impairment of human health, or presents 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.[11] For stents, the FDA generally requires the trial to demonstrate efficacy 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 approval process for Class III devices is very different than in the US. Medical devices are regulated by three EU Directives. The main directive is the Medical Devices Directive which passed in June, 1993 and 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" to implement regulatory control over devices. Notified bodies are independent, commercial organizations that are designated, monitored and audited by the relevant member states via "competent authorities." Currently, there are more than 70 active notified bodies within the

[11] There are two basic regulatory pathways within the FDA to bring a device to market: Pre-Market Approval (PMA) and the 510(k). The PMA process applies to Class III devices, while the 510(k) process generally applies to Class II and some Class I devices. Under the 510(k) process the manufacturer needs to demonstrate that the device is 'substantially equivalent' to a predicate device. Generally, bench testing data and perhaps a very small clinical study is all that is necessary for a device to demonstrate equivalency. While there is no standard timetable for 510(k) clearance, a straightforward clearance can typically be obtained within several months.

EU. A firm is free to choose any notified body designated to cover the particular type of device under review.[12] To obtain an CE mark a Class III medical device needs to only demonstrate safety and performance. Compliance with this standard usually can be demonstrated with much simpler and cheaper clinical trials than required by the FDA.[13]

The difference between the two regulatory regimes is largely a consequence of different histories that lead up to the passing of the primary medical device legislation in the two regions. 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 which garnered 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, lobbying, and government reports. For example, a series of Boston Consulting Group reports shows 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 10 devices that were approved in the EU, not approved by the FDA, and lead to significant adverse events in patients. The FDA study only focused on the negative consequences of the EU's relatively lax regulatory standards and does not acknowledge the benefits of greater access to devices in the EU. Perhaps most importantly, by focusing on extreme and 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 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 of Boston Scientific's Taxus Express II and 20 percent less than J&J's Cypher stent. Its target vessel failure (TVF) rate was 8 percent less than the Taxus stent.[14] The

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

[13]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, third, and fourth generation products, so coverage and payment determination has already been made prior to their introduction.

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

impact of TVF is significant as it requires additional interventions to restore vessel function.

One feature both the US and EU share is that outside of clinical trials and wide-spread, catastrophic device failure, useful information on device performance is scarce. There is virtually no ability in the post-market environment to systematically monitor the key clinical performance attributes (e.g. mortality, MACE, TVF) of medical devices.[15] This fact has received attention lately in the US, with the FDA introducing a new unique device identifier database and the 21st Century Cures act proposing decreased pre-market testing and increased post-market surveillance in some therapeutic areas.

While the consequences of the different regulatory regimes has generated significant policy debate, what is less controversial is that there are significant lags between the US and the EU in device introduction. Conditional on entry into both the US and the EU, BCG documents that medical devices are introduced into the US approximately four years after the EU.[16]

3 A Model of Learning, Regulation, and Choice

In order to guide our empirical analysis, we first develop a model that captures the trade-off between risk and access involved in regulating market entry of products with uncertain quality. This model also serves as the framework for the structural estimation and counterfactuals. In our model, products are developed with uncertain quality; this uncertainty is potentially resolved over time via exogenous signals (e.g. from reporting of clinical trials or other research, or from observational learning); a regulator requires costly premarket clinical trials to accelerate learning; firms enter when they expect entry to be profitable; and risk-averse consumers choose from the available products in the market at a point in time.

Our model captures many of the salient features of medical device markets and the role of the regulator. However, the medical device sector is complicated and there are notable institutional features that we purposefully de-emphasize in order to keep the model tractable and parsimonious.

As we have modeled, medical device quality is uncertain, but this uncertainty is symmetric among manufacturers, regulators, and consumers. If manufacturers are differentially

[15]The FDA maintains an adverse event reporting system called MAUDE to which certain entities are required to submit reports (e.g. manufacturers) and voluntary submissions are also allowed. Unfortunately, because the criteria for submission are not well defined, MAUDE receives hundreds of thousands of submissions per year, and reports often have nothing to do with the involved device per se. As the FDA states on their web site, the MAUDE system alone "cannot be used to establish rates of events, evaluate a change in event rates over time or compare event rates between devices." (https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfmaude/search.cfm) We have spent considerable time attempting to find post-market performance data in the EU, and to the best of our knowledge no attempt is made in the EU to systematically track device performance.

[16]BCG (2012) *Regulation and Access to Innovative Medical Technologies.*

informed about their devices' qualities, device regulation could solve a lemons problem (Leland 1979). At the extensive margin of whether to have any regulation at all, the lemons problem is surely relevant. However, our focus is on the appropriate standards of that regulation not on whether the regulation should exist at all. The variation that we exploit aligns with this focus. The EU is more lax in their standard relative to the US yet we are unaware of any significant evidence that the device market in the EU 'unravels' more than in the US. In fact, the presence of many more device offerings in the EU suggests that the variation in regulations between the US and EU is not a margin that would induce a lemons type market failure. This may be due to the fact that the relatively small number of interventional cardiologists who use the devices we study tend to stay well-informed about the most recent clinical updates for the products on the market (two days each at the three major conferences are devoted to reporting such data). While we believe that the institutional setting and the policy relevant margin suggest that the symmetric information framework is the appropriate lens through which to perform the analysis, we nevertheless test this assumption in Section 4.1.2 and we cannot reject symmetry of information.

We also do not model the possibility that the regulator will reject a device. We do, however, allow for manufacturers' optimal entry decisions in the face of clinical trial and entry costs conditional on developing the device. This amounts to an implicit assumption that no firm would enter with a product the regulator would want to reject, which is exactly the case under our symmetric information assumption. We have investigated the frequency of FDA rejections of coronary stents over the time frame of our data and have found no evidence of a single submitted PMA application for coronary stents that was ultimately rejected.[17] [18]

The next several subsections lay out the model. Section 3.1 describes how market participants learn about product quality over time, Section 3.2 describes consumer behavior and how it is affected by uncertainty about product quality, Section 3.3 turns to supplier pricing

[17] As the FDA does not report data on the number of PMA applications that are rejected, we performed searches of device manufacturers press releases and financial filings. It is important to note that there are frequent communications between the FDA staff and the device manufacturer over a given application and the evidence the FDA needs to see prior to granting approval. Thus, our impression is that there are few devices that submit a full PMA applications that are rejected as the process is costly for the manufacturer and they have a relative precise estimate of the likelihood of approval prior to submission. This differs from the pharmaceutical context where there is often disagreement between the manufacturer and regulator regarding the weighting of any positive treatment effect vs. potential negative side effects (and where an asymmetric information model might be more appropriate in some contexts such as drugs primarily marketed to non-expert general practitioners, and FDA acceptance sends an important quality signal to the market).

[18] Johnson and Johnson's CoStar stent was pulled off the market in the EU in 2007 after it performed poorly in its FDA required pivotal trials. Because of this, J&J did not submit a PMA application for this stent. See "Johnson and Johnson Won't Seek Approval for a Heart Stent," *New York Times*, May 8, 2007. As we discuss below and consistent with our framework, we observe CoStar's EU share erode significantly after the negative results are reported.

and entry, and finally Section 3.4 lays out the role for a regulator to affect total surplus via information requirements and their effect on risk and access.

3.1 Information and Learning

In our framework, consumers (and manufacturers and regulators) are uncertain about the quality of newly developed stents. The quality of the stent captures all features of the device that physicians and patients consider when selecting a particular device including ease of use and its efficacy/safety profile. Information on the quality of a given a stent accrues to the market over time from two sources, each with potentially different precision: First, products undergo clinical trials, and information from these trials diffuses to physicians through updates reported at major cardiology meetings throughout the year, published articles in medical journals, and discussions with peers and medical device representatives. Second, usage of the product may generate observational learning which is then shared and diffuses to the market.

Specifically, we assume innovative new devices j are each developed with quality Q_j according to a distribution $F_t(Q)$:[19]

$$Q_j \sim F_t(Q) := N(\overline{Q}_t, \sigma_Q^2). \tag{1}$$

where the subscript t allows for technological advancement over time.[20]

Over time, unbiased but noisy signals A arrive regarding the product's quality as new information from ongoing clinical trials and real world usage are generated, released, and diffused into the market. Letting age a denote the time since product j was introduced into the market (not calendar time), A_{ja} is given by:[21]

$$A_{ja} = Q_j + \nu_{ja} \qquad \text{where} \quad \nu_{ja} \sim \begin{cases} N(0, \sigma_{A^c}^2) & \text{if in clinical trials} \\ N(0, \sigma_A^2) & \text{if not} \end{cases} \tag{2}$$

Given these signals, beliefs about product quality are updated via Bayes' rule, resulting

[19] For simplicity, we assumed the prior and signal process to be normally distributed. In principle, we could relax this assumption. However, sample size limitations make taken a more non-parametric route undesirable, and we find that the simple normal model fits the data quite well with a small number of parameters.

[20] We abstract away from any feedback effects from the regulatory regime on the incentive of device manufacturers to invest in discovering new products and assume that the quality distribution and innovation arrival process are exogenously fixed. Depending on one's model of the nature of innovation, this could be thought of as a "shorter run" model of potential response to market changes.

[21] 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.

in posterior beliefs distributed normal with mean:

$$Q_{ja+1} = \frac{\sigma^2_{ja}}{\sigma^2_{ja} + \sigma^2_{A^{ja+1}}} A_{ja+1} + \frac{\sigma^2_{A^{ja+1}}}{\sigma^2_{ja} + \sigma^2_{A^{ja+1}}} Q_{ja} \tag{3}$$

and variance:

$$\sigma^2_{ja+1} = \frac{\sigma^2_{A^{ja+1}}}{\sigma^2_{ja} + \sigma^2_{A^{ja+1}}} \sigma^2_{ja}. \tag{4}$$

With this uncertainty and learning as a backdrop, the regulator must decide on the required length of clinical trials, trading off any costs of later access versus any benefits of more information and reduced risk. Once a product has been subjected to the required clinical trials, it is released to the market and consumers (doctors and patients) make decisions about which product to use given the current available choice set and information. We begin with the consumers' problem and work backwards.

3.2 Consumer Choice

Given beliefs regarding a stent's quality and the uncertainty over that quality, we assume consumers select the stent that yields the highest expected utility. For each patient/doctor combination i the indirect expected utility function from using device j at time t (where the t subscript refers to the calendar month, which will be associated with an age a for a product j) takes the form

$$u_{ijt} = Q_{jt} - \theta^p p_{jt} - \frac{\rho}{2}\sigma^2_{jt} - \gamma \ln(|\mathcal{J}_t|) + \epsilon_{ijt} =: \delta_{jt} + \epsilon_{ijt}, \tag{5}$$

where θ^p is the weight on price p_{jt} in the physician's choice, ρ is the coefficient of risk aversion, ϵ_{ijt} is an i.i.d. error term capturing the deviation of doctor preferences and/or patient appropriateness for device j relative to the population average, and $\gamma \ln(|\mathcal{J}_t|)$ is the adjustment suggested by Ackerberg and Rysman (2005) to model the fact that unobserved product space may become "congested" as the set of products available $\mathcal{J}_t$ grows. In our setting this helps to capture duplicative product innovations and/or the fact that not all products may be in the choice set for all patient/doctor combinations.

Assuming consumers choose the product j that maximizes expected utility, the set of patients for whom a doctor chooses product j (in month t) is then $\mathcal{A}_{jt} := \{i | j = \arg\max_{k \in \mathcal{J}_t} u_{ikt}\}$. Expected quantities are then given by the market size M_t and the choice probabilities:

$$q_{jt} = M_t s_{jt} = M_t Pr[j = \arg\max_{k \in \mathcal{J}_t} u_{ikt}] = M_t \int_{\mathcal{A}_{jt}} f_t(\epsilon) d\epsilon = M_t \frac{e^{\delta_{jt}}}{\sum_{k \in \mathcal{J}_t} e^{\delta_{kt}}}, \tag{6}$$

where the last equality obtains from the assumption that ϵ is distributed "logit" i.i.d. extreme value type I with unit variance. The choice set includes outside option $j = 0$, utility normalized to zero, representing the best non-stent treatment for that patient. Producer variable profits are then $\pi_{jt} := q_{jt}(p_{jt} - c_{jt})$.

Expected consumer surplus (ex-ante, relative to the best non-stent alternative) can then be obtained by summing over patient utility for the products chosen, scaling utils into dollars, and scaling for market size:

$$CS(\mathcal{J}_t) = M_t \frac{1}{\theta^p} \int_{\mathcal{A}_{jt}} u_{ijt} f_t(\epsilon) d\epsilon = M_t \frac{1}{\theta^p} \ln \left(\sum_{j \in \mathcal{J}_t} e^{\delta_{jt}} \right) , \tag{7}$$

where the final equality obtains from the logit distributional assumption on ϵ. Because the product quality distribution has support over the entire quality space (a stent can receive a large in absolute value negative draw from $N(\overline{Q}_t, \sigma_Q^2)$), the consumer surplus formulation explicitly accounts for the possibility that that a given stent could systematically underperform. In this way, our framework captures the health risks including significant harm from need for repeat procedures (TLR) or death (MACE) that consumers face when selecting a stent.

3.3 Bounds on Supply Effects on Total Surplus

We are interested in total surplus, $TS(\mathcal{J}_t) = CS(\mathcal{J}_t) + \sum_{j \in \mathcal{J}_t} \pi_{jt} - \phi^e \left| \mathcal{J}_t^e \right|$ (where $\mathcal{J}_t^e(T^c)$ is the set of firms who enter in period t), as the object that the regulator should seek to maximize in its choice of the optimal length of clinical trials T^c.[22] Total surplus is a function of the choice set $\mathcal{J}_t$, which in turn is a function of supplier entry behavior, given costs of trials required for entry $\phi^e(T^c)$ and expected profits after entry, conditional on the expected behavior of other firms. A fully specified supply model requires models of pricing and entry/exit dynamics. In the case of our analysis, both of these modeling efforts would entail a combination of approximating assumptions and work at the frontier of the business-to-business contracting and dynamic oligopoly literatures. For pricing, we would need to either construct an approximate expected price function at our region-time level of analysis, or build on the hospital-time level analysis in Grennan (2013, 2014) to allow for strategic choice in who contracts with whom (Lee and Fong 2013), an important feature of the EU market during the time frame we study. For dynamic entry and exit, we face a problem with

[22] Other regulator objective functions are also plausible. The goal of this paper is not to uncover the true regulator objective function but to examine the welfare implications of different regulatory approaches. We choose this objective function as it seems the most natural. We recognized that different regulator objective functions would have implications for the optimal policy but the regulator will still have to account for inherent trade-off between risk and access in the optimal policy.

a large and continuous state space, requiring approximations of the type explored in recent papers such as Ifrach and Weintraub (2014). Because constructing such models involves a substantial refocusing of the contributions of this paper to arrive at point estimates that would be caveat to a number of additional modeling and approximating assumptions, we instead take a bounds approach to partially identify the answers to our policy questions of interest.

We construct bounds on the total surplus generated under any regulatory policy $TS(T^c)$ that rely on weak assumptions on supply side behavior:[23]

Supply Assumption 1 (EU Entry Costs): EU entry costs are low enough such that all products developed with positive expected profits after EU testing enter.

Supply Assumption 2 (Entry Policy): The equilibrium entry policy of firms is increasing in own expected profits: $\pi_j > \pi'_j \Rightarrow Pr[Enter_{jt}|\pi_j] > Pr[Enter_{jt}|\pi'_j]$.

Supply Assumption 3 (Pricing): Prices are bounded by the marginal contribution of the product: $p_j - c_j \leq TS(\mathcal{J}) - TS(\mathcal{J} \setminus \{j\})$.

Under these assumptions, we can construct upper and lower bounds for total surplus. The upper bound is given by the case where there are no direct fixed costs of longer trials, so all firms enter in equilibrium, and the only impact of increasing trial length on market structure is to delay access to the newest technologies (in addition to increasing learning). The lower bound is given by assuming that the cost of trials is \$1.6M per month, but with firms' entry decisions made under the assumption that other firms enter as if entry costs are zero.[24] Fewer firms will enter than in the true equilibrium because this case doesn't allow expected market shares and prices to increase as fixed costs increase and the market becomes more concentrated. Full proofs are provided in Appendix A.1.

The advantage of these particular bounds is their simplicity of computation. Expected profits and thus entry decisions can be computed directly from the data, and then total surplus can be computed using the estimated demand model for the set of products that enter. How informative these bounds will be depends on the size of trial costs relative to the distribution of product qualities in the data. The bounds will be equal to each other and the full equilibrium model at $T^c = 0$ (the case of no US trials beyond EU requirements), then diverge as increasing entry costs drive a wedge between the entry assumed in the bounds and the entry that would obtain in a full equilibrium.

[23] Our bounds approach relies on necessary conditions from theory that are easy to compute from the data and demand model. Reguant (2016) provides a complementary approach to obtaining computational bounds in cases that are more challenging to compute.

[24] \$1.6M per month from the survey by Makower et al. (2010).

3.4 Modeling the Regulator's Tradeoffs

The total surplus equation above captures the primary tradeoff between access and risk: the longer time T^c that products spend in premarket clinical trials, the lower the risk from uncertainty about product quality in the market σ_{jt}, but the fewer new technologies available in the consumer choice set $\mathcal{J}_t$ at any point in time and greater costs of entry. This tradeoff can be formalized by writing total surplus as a function of time spent in premarket clinical trials and considering the marginal return to increasing the amount of time spent in premarket testing to $T^c + 1$ over any time period $t = 1, ..., \tau$:

$$\begin{aligned} TS(T^c+1) - TS(T^c) &= \sum_{t=1}^{\tau} \frac{M_t}{\theta^p} \ln\left(\frac{\sum_{j\in\mathcal{J}_t(T^c+1)} e^{\delta_{jt}(T^c+1)}}{\sum_{j\in\mathcal{J}_t(T^c)} e^{\delta_{jt}(T^c)}}\right) \\ &\quad + \sum_{j\in\mathcal{J}_t(T^c+1)} \pi_{jt}(T^c+1) - \sum_{j\in\mathcal{J}_t(T^c)} \pi_{jt}(T^c) - \phi^e \left|\mathcal{J}_t^e(T^c+1) \setminus \mathcal{J}_t^e(T^c)\right| \ . \end{aligned} \tag{8}$$

From the regulator's perspective, the optimal trial length sets (8) to zero.

One way to clearly see the tradeoff between access and risk as a function of trial requirements is to consider a simple scenario where there is no observational learning, no direct cost of trials, and no affect of price on total surplus. In this case, the per-period marginal return to increasing premarket testing simplifies (proof in Appendix A.2) to:

$$\frac{TS(T^c+1) - TS(T^c)}{\tau} = \frac{\rho}{2}(\sigma^2_{T^c} - \sigma^2_{T^c+1}) - \frac{1}{\tau}\ln\left(\frac{\sum_{j\in\mathcal{J}_\tau(T^c)} e^{Q_{jt} - \gamma\ln(|\mathcal{J}_t|)}}{\sum_{j\in\mathcal{J}_0(T^c)} e^{Q_{jt} - \gamma\ln(|\mathcal{J}_t|)}}\right) \ . \tag{9}$$

The first term captures the per period utility gain from decreased risk (and is determined by the unconditional uncertainty in product quality σ_Q and the rate of learning in trials σ_{A^c}). The second term captures the total surplus generated by the rate of technological improvement in product quality over time (which will be determined by the rate of drift in product quality over time $\overline{Q}_t$ and the expansion of variety in the choice set $\mathcal{J}_t$).

3.5 Model Predictions to Take to the Data

The model has several testable implications that we can take to the data. In order to map the model into the data, we match the choice probabilities implied by utility maximization to the market share data, and invert the system as in Berry (1994) to recover the mean utility estimates for each product in each month:

$$\ln(s_{jt}/s_{0t}) = \delta_{jt} := Q_{jt} - \theta_p p_{jt} - \frac{\rho}{2}\sigma^2_{jt} - \gamma\ln(|\mathcal{J}_t|) \ . \tag{10}$$

The model then implies the following (proofs in Appendix A.3):

Prediction 1 (Learning): If initial product quality is uncertain ($\sigma_Q^2 > 0$), then learning ($1/\sigma_A^2 > 0$) implies that volatility in product-specific quality estimates converge by decreasing with age ($E_j|\delta_{ja} - \delta_{ja+1}| \searrow^{a\to\infty} 0$ or $Var_j(\delta_{ja}) \searrow^{a\to\infty} \sigma_Q^2$).

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 as learning occurs ($E_j[\delta_{ja}] \nearrow^{a\to\infty} \overline{Q}$).

The summary statistics and reduced-form analysis in the next Section will use these predictions in exploring how much more stringent the US regulatory policy actually is relative to the EU, and the implications for learning, risk, and access in the market.

4 Data 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, providing the number of patients potentially eligible for a stent in each hospital-month. In our analysis, we aggregate the data to the region-month (US and EU) level to obtain accurate measures of market entry and usage of each device within a region, which is the relevant unit of observation for this study.[25]

As mentioned in Section 2, to our knowledge, high quality data on post-market device performance does not exist. The lack of such data is the topic of the policy debate regarding "post-market surveillance" that we analyze in Section 6. We believe usage data to be the next best thing because it captures, via revealed preference, the state of market knowledge of physicians across products and over time.

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 also collected clinical trial data (when available) from various journal articles, conference abstracts, press releases, and product catalogs. These provide further evidence

[25] As we are analyzing second generation stents, reimbursement policies have already be determined so that these region specific differences are not a consequence of difference in the timing of payment approvals. We also prefer the EU region vs. country analysis because it provides more precisely measured market shares, which are important for our demand model. Appendix D.3 shows variation in data and our estimates across countries within the EU, and Appendix D.4 shows the same within hospitals.

regarding the size and length of trials required for US versus EU entry. They also provide product TLR and MACE rates which we use to validate our revealed preference estimates of quality.

Table 1: US and EU differences in clinical trials and market structure.

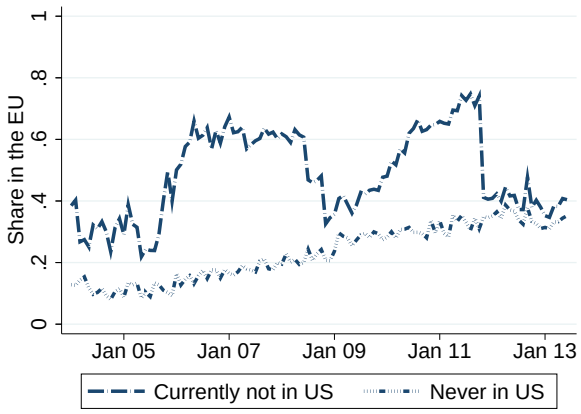
	US	EU
Clinical trial data:		
Percent of stents with publish clinical trials	45	16
Mean clinical trial size (patients)	1252	471
Mean clinical trial length (months)	32	11
Market structure data:		
Mean manufacturers in market	4 (3)	21 (5)
Mean products in market	11 (5)	39 (8)
Total products in market 2004-13	21 (11)	109 (22)
Mean months from EU to US entry	10	-
Mean months from EU to US entry (DES)	17	-
(Usage within hospital in parentheses.)		

The top third of Table 1 presents summary statistics for our clinical trial data. We were able to find such data for 45 percent of the products entering the US and 16 percent of the products that enter the EU only. Conditional on publishing a clinical trial online, EU trials are shorter and enroll fewer patients: On average, US clinical trials enrolled over 1,200 patients and lasted almost 3 years, while the EU-only products enrolled a third of the patients and the trials were 66 percent shorter in duration. This large difference in trial patterns would be expected based upon what we know about testing requirements in the two regions. That it persists in the available testing data demonstrates that private incentives do not generate further testing, making regulatory requirements binding and potentially important.

The bottom two thirds of Table 1, and also Figure 1, show how these pre-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 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

Figure 1: EU market share of products not available in US.



on the quality of these products. The goal of our analysis will be to resolve this tension and determine, for our sample of coronary stents 2004-13, whether the extra US testing provides information that the market values in terms of decreasing risk, the extent to which there is observational learning outside of clinical trials, and the value of access to more products earlier in the EU.

4.1 Evidence: Information, Learning, and Risk Aversion

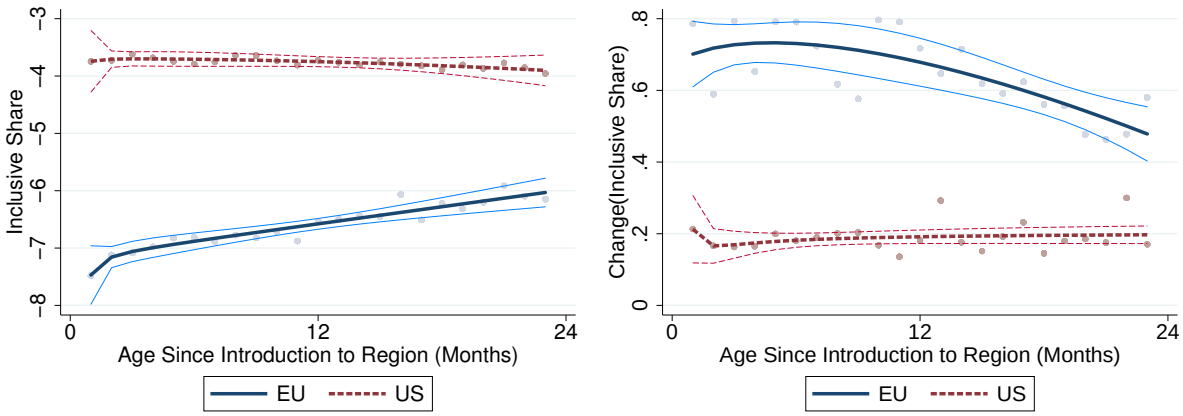
Figure 2 begins to explore whether EU consumers bear more risk than those in the US by introducing a larger number of devices earlier in their life cycles with less information imparted about the quality of those devices. Panel (b) plots mean absolute differences of the mean utility estimates over time, $\frac{1}{J_a}\sum_j |\ln(s_{ja}/s_{0a}) - \ln(s_{ja+1}/s_{0a+1})|$. As highlighted in Prediction 1, when consumers learn this value should asymptote toward zero as learning moves the quality estimate, Q_{ja}, toward true quality, Q_j. This statistic is decreasing in the EU and constant over the product lifetime in the US (though the difference of these differences is noisily estimated).[26] The gap remaining between the US and EU at $a = 24$ could be due to differences in the product mix or incomplete learning.

Panel (a) shows that in the EU the mean across products of a given age of the mean utility estimates $\frac{1}{J_a}\sum_j \ln(s_{ja}/s_{0a})$ is lower upon introduction and gradually increases with age, plateauing after approximately two years in the market. This is consistent with Prediction 2 of our theory regarding learning in the presence of risk aversion. Again, the pattern in the

[26]Month-to-month volatility is our preferred measure of learning for exposition because it is tightly linked to theories of convergence and decreases to zero with full learning about a fixed parameter. Appendix D.1 shows similar results for the standard deviation across products. Appendix B.2 provides further details on heterogeneity (good news and bad news) in the effect of learning on usage patterns at the product level which are concealed by the average patterns reported here.

Figure 2: Raw data comparison: EU vs. US.

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



	$x_{a=1}$	$x_{a=24}$	$x_{24} - x_1$	$(x_{24}^{EU} - x_1^{EU}) - (x_{24}^{US} - x_1^{US})$
$\text{Mean}_{j\|a}^{EU}\ \ln(s_{jt}/s_{0t})$	-7.48	-6.15	1.33	1.55
	(0.24)	(0.27)	(0.38)	(0.55)
$\text{Mean}_{j\|a}^{US}\ \ln(s_{jt}/s_{0t})$	-3.74	-3.96	-0.21	
	(0.28)	(0.37)	(0.46)	
$\text{Mean}_{j\|a}^{EU}\ \|\Delta_t \ln(s_{jt}/s_{0t})\|$	0.79	0.58	-0.21	-0.16
	(0.07)	(0.08)	(0.11)	(0.13)
$\text{Mean}_{j\|a}^{US}\ \|\Delta_t \ln(s_{jt}/s_{0t})\|$	0.21	0.17	-0.04	
	(0.05)	(0.06)	(0.07)	

$N = 1747$ product-month-region observations. Standard errors clustered by month $N_t = 114$ in parentheses.

US is different. There the mean of the mean utility estimate does not vary with product age and is higher on average. That the mean utility gap remains between the US and EU at $a = 24$ suggests the product mix in the US is of higher quality, but this particular comparison can not rule out the possibility of different preferences across the two regions.

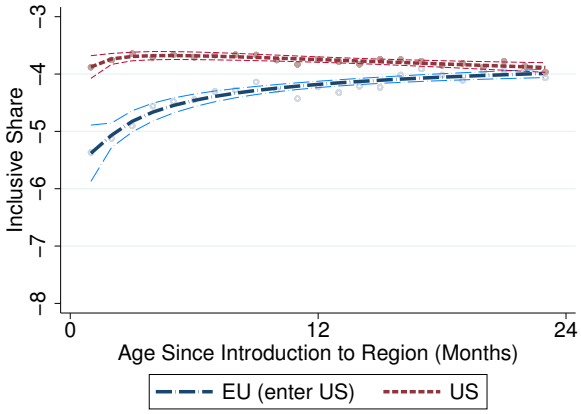
The patterns in the raw data in Figure 2 are consistent with our model predictions of learning by risk averse consumers while products are on the EU market. However, one might be concerned they are also potentially consistent with several alternative mechanisms: drivers of product diffusion other than learning, observational learning flowing from the EU to US, and product selection on uncertainty or on quality via early exit. Figures 3 and 4 rule out these alternatives in favor of the mechanism of EU learning through US clinical trials.

These figures are constructed after controlling for product fixed effects, so that all patterns are driven by within-product variation over time. In particular, this rules out any composition effect whereby increasing usage with product age is driven by worse products exiting the market at a younger age.

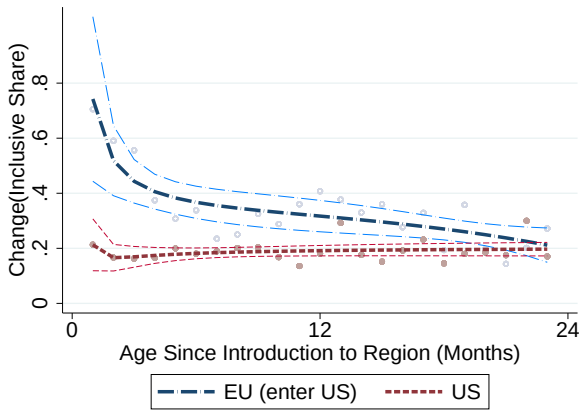
Figure 3 compares patterns for the exact same products marketed in both the EU and

Figure 3: Same products, EU vs. US.

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



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



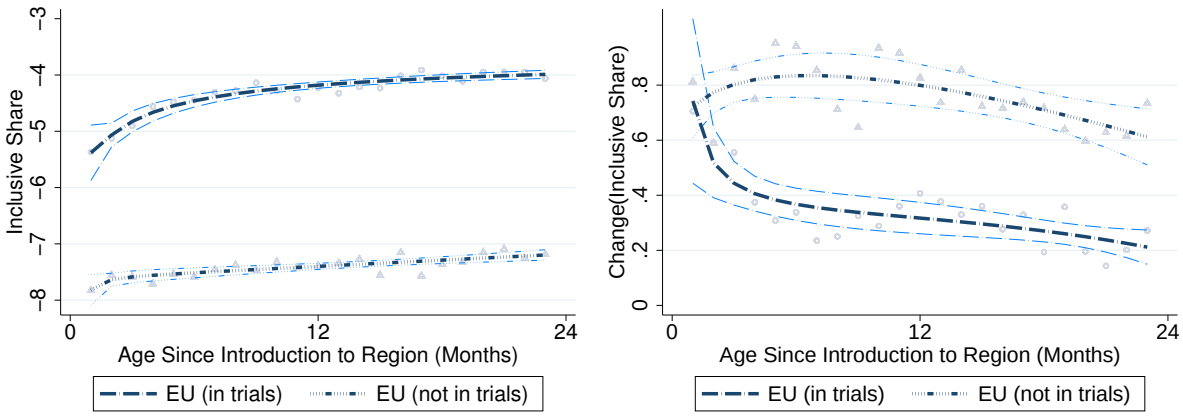
	$x_{a=1}$	$x_{a=24}$	$x_{24} - x_1$	$(x_{24}^{EU} - x_1^{EU}) - (x_{24}^{US} - x_1^{US})$
Mean$_{j\|a}^{EU\|_{trials}}$ $\ln(s_{jt}/s_{0t})$	-5.36	-4.06	1.30	1.39
	(0.27)	(0.09)	(0.28)	(0.29)
Mean$_{j\|a}^{US}$ $\ln(s_{jt}/s_{0t})$	-3.88	-3.96	-0.08	
	(0.12)	(0.14)	(0.18)	
Mean$_{j\|a}^{EU\|_{trials}}$ $\|\Delta_t \ln(s_{jt}/s_{0t})\|$	0.71	0.27	-0.43	-0.39
	(0.18)	(0.05)	(0.19)	(0.20)
Mean$_{j\|a}^{US}$ $\|\Delta_t \ln(s_{jt}/s_{0t})\|$	0.21	0.17	-0.04	
	(0.05)	(0.06)	(0.07)	

$N = 697$ product-month-region observations. Standard errors clustered by month $N_t = 114$ in parentheses.

US. The fact that mean of mean utility estimates in panel (a) is flat in the US rules out that the increasing usage with age is driven by marketing, sales or distribution efforts that are specific to product launch in a new region—if it were such a non-learning diffusion story the US would exhibit a similar pattern (and the learning effect could still be estimated as the difference between the shapes of US and EU curves, assuming learning is complete by the time products reach the US). Panel (b) again plots mean absolute differences, which which in the EU starts near 0.8 logit utils and converges fairly rapidly over the first 6-12 months on the market before leveling off around 0.2 (that it does not asymptote to 0 suggests either incomplete learning or some unobservable source of noise in the product usage data that we have not modeled). By contrast, in the US this same quantity stays level at 0.2 the entire time after introduction. That the volatility and mean usage in the EU converge to US levels when we restrict the sample to the same products in both regions is consistent with learning occurring during the time these products are in the EU, and risk-averse users responding to this learning. Figure 4 explores whether the source of this learning is from observational usage in the EU or from US clinical trials.

Figure 4: EU only, products in trials vs. not.

(a) Mean$_{j|a}$ $\ln(s_{jt}/s_{0t})$ (within product) **(b)** Mean$_{j|a}$ $|\ln(s_{jt}/s_{0t}) - \ln(s_{jt+1}/s_{0t+1})|$



	$x_{a=1}$	$x_{a=24}$	$x_{24} - x_1$	$(x_{24}^{EU} - x_1^{EU}) - (x_{24}^{US} - x_1^{US})$
Mean$_{j\|a}^{EU\|_{trials}}$ $\ln(s_{jt}/s_{0t})$	-5.37	-4.06	1.30	0.66
	(0.27)	(0.09)	(0.28)	(0.34)
Mean$_{j\|a}^{EU\|_{not}}$ $\ln(s_{jt}/s_{0t})$	-7.83	-7.19	0.64	
	(0.13)	(0.14)	(0.19)	
Mean$_{j\|a}^{EU\|_{trials}}$ $\|\Delta_t \ln(s_{jt}/s_{0t})\|$	0.71	0.27	-0.43	-0.36
	(0.18)	(0.05)	(0.19)	(0.25)
Mean$_{j\|a}^{EU\|_{not}}$ $\|\Delta_t \ln(s_{jt}/s_{0t})\|$	0.81	0.73	-0.08	
	(0.08)	(0.12)	(0.15)	

$N = 1430$ product-month observations (all in EU). Standard errors clustered by month $N_t = 114$ in parentheses.

Figure 4 compares products in the EU that (1) undergo clinical trials for US release (a set that is the same as the EU products that eventually enter the US shown in the previous figure, plus a few products that undergo US trials but are not introduced to the US); and (2) that never undergo trials beyond those required for EU introduction. The results show that all of the evidence of learning is driven by those products in US clinical trials. Both mean absolute differences in panel (b) and mean of mean utilities in panel (a) are flat for products not in trials, which not only rules out observational learning for these products, but also rules out diffusion driven by marketing, sales, or distribution of new-to-the-world products. The pattern of mean absolute differences also refute the argument of selection on uncertainty, as the EU products not in trials begin near 0.7 (the same as the products in trials), but remain flat near 0.7 over time, suggesting that there is plenty of uncertainty for these products, but no learning.

This last point is especially important because identification of a causal effect of learning from US clinical trials requires that unobservable factors driving product entry decisions into

the US are uncorrelated with the "amount to learn" about the device from clinical trials. The patterns in the last three figures offer empirical evidence that this condition is satisfied as US trial participation is correlated with higher expected quality (which we can control for using product fixed effects), but it is uncorrelated with uncertainty about that quality level (as measured by volatility in usage patterns).

4.1.1 Robustness of the Learning Evidence

We perform a series of robustness analyses and all of them confirm our basic findings that US trials generate risk reducing information that is valued by consumers and there is very little observational learning. These results are presented in detail in Appendix D and summarized here. First, we examine the possibility of differential selection into the US by reformulating Figure 4 for a set products that do and do not undergo US testing, and also have overlapping support on initial quality estimates at $a_j = 1$. Because all of these products have the same initial quality level, identification is 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). The pattern is essentially identical to that in Figure 4, indicating that our results are not driven by selection on initial quality/usage levels.

Next, we examine the impact of aggregating our analysis across all EU countries and hospitals. We perform our analysis country-by-country, and the conclusion that there is significant learning from US trials but little observational learning again holds. We also perform our analysis at the hospital level. Here the estimates are noisier (which is expected due to sampling variation for niche products and the fact that usage is conditional on hospital contracting and preferences), but again our main conclusions hold.

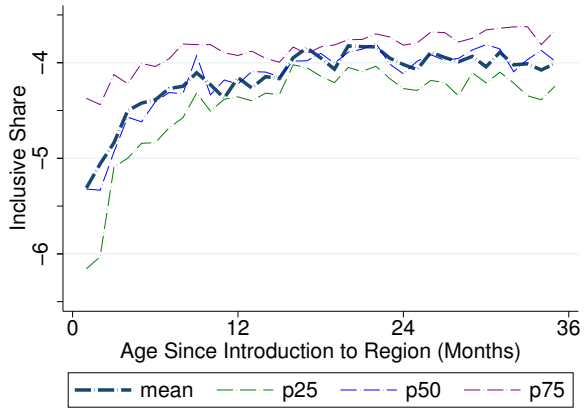
Finally, we perform a placebo test using percutaneous transcatheter coronary artery (PTCA) balloons, which face similar regulatory requirements in the EU and US, and thus should not display the differential signs of learning we document for stents if our proposed mechanism is true. The results show that 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, and there is no gap in time of entry on average. Most importantly, the usage patterns with age show no evidence of learning for PTCA balloons.

4.1.2 Empirical Test of Symmetric vs Asymmetric Information

Figure 5 offers further checks that the data is consistent with a model of learning over time with symmetric information between manufacturers and consumers. Here we focus only on

those products in the EU that are undergoing clinical trials (the products with evidence of learning). We present different quantiles of the $\ln(s_{jt}/s_{0t})|_a$ distribution. Under symmetric information, the distribution of product quality estimates should converge symmetrically to the true product quality distribution with the arrival of new information. 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 D.6 for more on this intuition). Learning in this way will cause the lower tail of the distribution to become truncated. Consistent with the symmetric informational framework, it is clear that 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.

Figure 5: Symmetry of changes in quality distribution (in EU; US trials).



	$x_{a=1}$	$x_{a=24}$	$x_{24}-x_1$
$\left(\frac{\mu-p50}{\sigma}\right)^{EU\|_{trials}}_{j\|a}\ln(s_{jt}/s_{0t})$	0.10	0.04	-0.06
	(0.24)	(0.17)	(0.29)
$\left(\frac{p75-p50}{p75-p25}\right)^{EU\|_{trials}}_{j\|a}\ln(s_{jt}/s_{0t})$	0.60	0.55	-0.05
	(0.24)	(0.17)	(0.28)

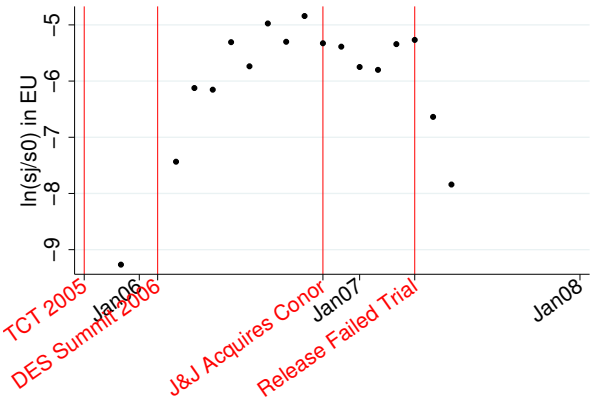
$N = 383$ product-months (in EU; US trials). Standard errors clustered by month $N_t = 114$ in parentheses.

4.1.3 CoStar Case Study and the Role of Bad News

The focus on averages across products thus far obscure the fact that information is not always good news for a product. The arrival of bad news will obviously reduce the posterior uncertainty, but it will also reduce the posterior mean quality estimate. Appendix B.2 provides more individual product summary statistics demonstrating these up and down

dynamics. 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 Feb 2006.[27] In Nov 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 trails as the stent was failing to meet its primary endpoints.[28]

Figure 6: 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 as Conor. And when trial results on efficacy were unfavorable, market share dipped and the product was pulled from the market.

4.2 Exploring Other US/EU Differences

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

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

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

have led to more and earlier entry in the EU. Further, we have presented evidence that this differential testing has led to different amounts of information generation, which can be observed in EU usage patterns.

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. However, all the evidence that we have been able to gather (detailed in Appendix B.3 and summarized here) indicates that these do not plausibly explain the patterns in the data described above. For example, the 2010 mortality rate in the US for ischemic heart disease was 127 deaths per 100,000; and the corresponding figure for the EU is 130 per 100,000.[29] This modest differential seems unlikely to account for the stark differences of entry rates between the two regions.

Prior to performing an angioplasty in which a stent may be inserted, the patient must undergo a diagnostic angiography 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. In our data, the distributions of the number of diagnostic angiographies performed and percent of those diagnostic procedures resulting in stenting procedure 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. Like the evidence on heart disease prevalence, this small difference seems unlikely to explain the large disparity in entry rates between the two regions.

It is also possible that other differences in preferences or willingness-to-pay might make the EU market more desirable for entry. Figure 12 in the Appendix 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. The same figure also shows that the prices per stent sold are lower in the EU. This is 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, actually making it a more attractive entry target than the EU.

4.3 Summary of the Evidence

The evidence from stent entry and usage patterns aligns with our model in which there is uncertainty about new product quality, learning occurs symmetrically to market players over time, and risk-averse decision makers factor uncertainty about quality into their product

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

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 not surprising given the lack of systemically collected data on device clinical performance after market entry.

We examine alternative plausible explanations and reject them all. Specifically, the patterns we observe are not consistent with differential marketing/diffusion, differential demand side factors, differential prices and lags in reimbursement determination, or selection into testing based on uncertainty.

5 Structural Identification, Estimation, and Results

The statistics presented in Section 4 align with the predictions of the model developed in Section 3 and suggest that the EU is indeed less stringent than the US in regulating the entry of new medical devices. In this Section we estimate the parameters of our model in order to better understand and quantify welfare implications. Using the data from the EU 2004-13, we estimate the distribution of product quality for innovations introduced, the rates of learning in and out of clinical trials, and risk aversion. We then use the estimated model to explore the access and risk issues separately before combining them in counterfactual analyses of optimal regulatory policy and post-market surveillance.

5.1 Identification and Estimation of Learning and Demand

The parameters of the utility function—and by extension the parameters of the device quality distribution and learning process—can be identified and estimated by a revealed preference assumption and data on device market shares in each month. We follow Berry 1994, setting the choice probabilities implied by utility maximization equal to the market shares in the data, inverting the system to be linear in mean utilities, and adding the econometric unobservable term ξ_{jt}:

$$\ln(s_{jt}/s_{0t}) = \delta_{jt} := Q_j - \theta_p p_{jt} - \frac{\rho}{2}\sigma_{jt}^2 - \gamma \ln(|\mathcal{J}_t|) + \xi_{jt}. \tag{11}$$

We use variation over time and across products in the first *and second* moments to estimate the product qualities Q_j (and from those the mean $\overline{Q}$ and variance σ_Q^2 of the product quality distribution), the signal variances σ_A^2 and $\sigma_{A^c}^2$, the risk aversion parameter ρ, the price sensitivity θ^p and the Ackerberg-Rysman congestion term, γ.

5.1.1 Identification of learning and demand

A simple and semi-parametric way to estimate Equation (11) 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 risk aversion and learning on utility. However, because we are interested in questions that involve market reactions to different learning rates and levels of risk, we need to add structure via the learning model to disentangle these forces. Comparison to the fixed-effect model in Appendix D.7 provides a useful benchmark for assessing the fit of the more parsimonious and parametric learning model.

As with all learning models, the identification of the signal precision depends on fitting the model to the *shape* of how choice behavior changes with the age of the product. The risk aversion parameter is then identified as the multiplicative shifter that best fits that shape to the observed choices. In our simple learning model, identification is even clearer because learning is identified by the fact that volatility 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.

This relates directly back to the reduced form evidence in Figure 4: For products in trials, the volatility decreases with age, identifying learning. As this variation decreases, the mean inclusive share increases, identifying risk-aversion. The fact that volatility is large and constant for products not in trials identifies a lack of observational learning. 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). The causal interpretation of learning from trials is based on the fact that selection into US trials appears to be based upon the level of expected US profits, not the amount of uncertainty about product quality.

5.1.2 GMM estimation algorithm

We implement the estimation based on the above intuition via a generalized method of moments algorithm, detailed in Appendix C and summarized here. For a given set of parameter estimates we construct empirical analogs of the product residuals, $\xi_{jt} = ln(s_{jt}/s_{0t}) - Q_j - \theta^p p_{jt} - \frac{\rho}{2}\sigma^2_{jt} - \gamma ln(\mathcal{J}_t)$. The estimator finds the parameters that set the empirical analog of $E(\xi' Z) = 0$. The instruments are $Z_{jt} = (\mathbf{1}_j, (\frac{1}{a_{jt}}, \frac{1}{a^2_{jt}}), (p_{jt-1}, \sum_{k\neq j} p_{kt-1}), ln(\mathcal{J}_t))$ where $(\frac{1}{a_{jt}}, \frac{1}{a^2_{jt}})$ are the instruments for volatility with age and $(p_{jt-1}, \sum_{k\neq j} p_{kt-1})$ are the instruments for price suggested by Grennan (2013) to capture price variation due to sticky contracts and changes in the competitive environment.

The key difference between our algorithm and the standard demand estimation algorithm is that in addition to matching the first moments of the model choice probabilities and market shares, we also need to match the second moments of quality uncertainty σ_{jt} in the model to the volatility in the data. We do this by fixing $\sigma_{jt|_{a=1}}$ equal to the standard deviation of $Q_{jt|_{a=1}}$ and using the learning model to predict the full sequence $\{\sigma_{jt}\}$.

Finally, we use the empirical distribution of the product fixed effects estimated from the EU data to estimate the mean $\overline{Q}$ and variance σ_Q^2 of the distribution of product qualities developed. This amounts to an assumption that all products that a firm might want to introduce to the market are in fact introduced in the EU. This is plausible as the EU has some products with very low market shares and profits (see Appendix B.4 for more on the distribution of profits for EU entrants), suggesting that the fixed cost of EU trials and introduction (conditional on having already developed the innovative product) are quite low.

5.2 Resulting Parameter Estimates

The parameter estimates from the model are presented in Table 2. The first observation is that there is meaningful underlying variation in product quality that exposes consumers to risk—at $\sigma_Q = 1.24$ the variation in product quality is nearly as large as the match-specific logit standard deviation of $\pi/\sqrt{6} = 1.28$. Thus without information, consumers selecting a new product for insertion face a significant probability that the product is significantly worse than the mean product quality.

Table 2: Structural parameter estimates of demand/learning model

$\overline{Q}$	σ_Q	$1/\sigma_A^2$	$1/\sigma_{A^c}^2$	θ^p (\$1000s)	ρ	γ
-5.02	1.24	0.01	0.70	0.25	4.54	0.12
(0.081)	(0.002)	(0.001)	(0.024)	(0.013)	(0.123)	(0.022)

$N = 4490$ product-months. Standard errors clustered by month ($N_T = 114$).

The second observation is that, consistent with the reduced form evidence, learning rates vary dramatically according to whether the product is in an FDA required clinical trial or not. The parameter estimates indicate that there is virtually no experiential market learning occurring—the estimate of $1/\sigma_A^2$ is an economical zero. By contrast, the precision of clinical trial learning $1/\sigma_{A^c}^2$ is 0.70, corresponding to approximately 18 months of clinical trials for learning to exceed 95 percent completeness.

The estimate of the price parameter is correctly signed though economically very small and similar to an appendicized demand system similar to ours in Grennan (2013) using US data 2004-07. The Ackerberg and Rysman parameter, γ, is also correctly signed and significant. Its magnitude implies that the logit assumption without this correction only modestly overstates the value of product variety.

5.2.1 Validation of model estimates

Risk-aversion: The model estimates provide several opportunities for validation with other data and research. One such comparison we find particularly reassuring is that the implied coefficient of risk aversion aligns with estimates of this parameter from the literature. Converting the estimate of ρ into a dollar equivalent by normalizing the total surplus per stenting procedure to \$5,000 (the estimated dollars in quality adjusted life years from the procedure relative to CABG), then the estimated risk aversion parameter is $\rho_{\$^{-1}} = 0.99 \cdot 10^{-3}$.[30] This is within the range of estimates of risk aversion in well-designed studies such as Cohen and Einav (2007).

Added values and prices: Another opportunity to validate our model with external data is to compare the price data to the added values implied by our estimated demand model, $AV_j := TS(\mathcal{J}) - TS(\mathcal{J} \setminus \{j\})$, the increase in total surplus from adding each product j to the choice set. Our estimates imply an average (quantity-weighted) added value of \$4629. This can be compared to average price of \$824 in the EU in our sample. Models of negotiated prices such as the Nash Equilibrium of Nash Bargaining used for stents in Grennan (2013, 2014) suggest that price should be equal to marginal cost plus a markup where the supplier receives a fraction of the added value that depends on his bargaining parameter vs. the bargaining parameter of the buyer. Though we do not specify a full supply side model, for the purpose of comparing our demand estimates to prices, we calculate that setting costs equal to the lowest observed price of \$168 implies bargaining splits where the supplier captures on average 14 percent of the added value. This is on the lower end the range estimated in Grennan (2013, 2014) for the US stent market 2004-07, but seems plausible with decreasing prices over time and lower EU procedure reimbursements.

***Q_j* and clinical trial data:** Finally, in our model, consumers formulate beliefs of quality of each stent, Q_j. We can compare these estimates from our model to the stent performance estimates in clinical trials for the 21 products for which we have trial data. For this analysis we focus on target lesion revascularization (TLR) rate but the same pattern also holds for MACE. Figure 7 panel (a) presents the scatterplot of Q_j versus TLR. Reductions in TLR correspond to better performance. Our measure of vertical stent quality is strongly and negatively correlated with the clinical quality measure, providing another piece of evidence on the validity of our estimated model.

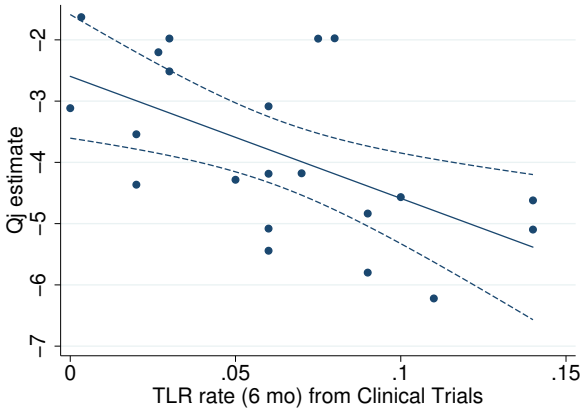
[30] We could also scale into dollars using the standard approach of the inverse of the price coefficient $\frac{1}{\theta^p} = 4057$, which comfortingly implies a similar total surplus per procedure of \$4,431. However, our estimated price coefficient is always economically small with the precise point estimate dependent on the specific demand model. Thus we take advantage of the fact that like many medical technologies, the procedure of angioplasty with a stent has been subject to numerous studies attempting to value the average quality adjusted life years added by the procedure in dollar terms. Source: Cost Effectiveness Analysis Registry (https://research.tufts-nemc.org).

5.3 Access: Rate of Technological Change

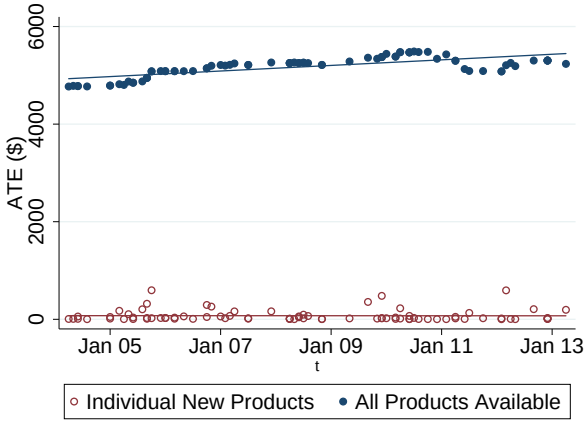
The rate of technological change is an important determinant of the optimal regulatory policy because it affects the value of access to the newest devices relative to those already available. We can compute the rate of technological change by calculating the *ex post* average treatment effect (ATE), i.e. the mean surplus (relative to the outside option) of having a stent implanted, $\ln(\sum_{\mathcal{J}_t} e^{Q_j})$. Figure 7 panel (b) presents these results. The top line is the mean consumer surplus per patient receiving a diagnostic angiography. Between 2004 and 2013, there is a meaningful increase of 9.6 percent in the utility consumers receive from access to coronary stents. Interestingly, though, the mean value to consumers of each new product at the time of its introduction, $\ln(1+e^{Q_j})$, (the bottom line in the figure) is flat over our sample period. This implies that the value of access in coronary stents during this time period is driven more by the increase in the variety in the choice set rather than increasing average product quality (relative to the next best alternative treatment) over time.[31]

Figure 7: Clinical Quality, Revealed Quality, and Technological Change.

(a) Correlation between Q_j and TLR Rate



(b) Rate of Technological Change



This finding is salient for the analysis of the optimal regulatory policy. If technological change is driven by increases in average product quality, changes in the 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. This is because niche products will by their nature have smaller market opportunities and thus find it more difficult to incur the fixed cost of greater testing.

[31] This is likely due at least in part to the increasing quality of alternative treatments, particularly less-invasive and beating-heart CABG, during this time. See Kalyanasundaram and Karlheinz (2014) for an overview.

5.4 Risk: Uncertain Quality and Market Outcomes

The potential welfare gain from access to new products is in tension with potential welfare loss due to the risk that those products may not be as effective as expected. The magnitude of this risk effect depends upon the mean quality level, the variance in that quality level, and the amount of information consumers possess.

Table 3 explores the role of uncertainty in the market by using the demand model to calculate total surplus per stent $\frac{TS}{1-s_0}$ and the percent of patients undergoing a diagnostic angiography who choose a stent over the outside good $(1 - s_0)$. Here we posit hypothetical markets where all products have uncertainty in their quality, varying from the unconditional variance of the quality distribution σ^2_Q (if there were no testing/learning at all), to the estimated uncertainty upon first entering the EU $\sigma^2_{a_{EU}=0}$ (after undergoing EU testing requirements), to varying lengths of US trials $\sigma^2_{T^c}$. In order to focus purely on the role of uncertainty, this is a partial equilibrium analysis in that we do not consider firms' strategic responses to these different parameters.

Table 3: The effect of uncertainty and mean quality on surplus per stent and the total number of stenting procedures. First set of rows for estimated quality distribution in the data, and also for the actual EU data (which has average uncertainty $\overline{\sigma^2_{jt}}|_{EU} = 0.29$, similar to that of $T^c = 3.2$ months of US clinical trials). The subsequent two sets of rows shift the product quality distribution one standard deviation of the logit distribution $\pi/\sqrt{6} = 1.28$ in each direction. All estimates differ at 95 percent confidence level except $T^c = 18$ and $T^c = 24$ (s.e. suppressed for readability).

		$\sigma^2_Q =$ 1.51	$\sigma^2_{a_{EU}=0} =$ 0.80	$\sigma^2_{T^c=6} =$ 0.18	$\sigma^2_{T^c=12} =$ 0.10	$\sigma^2_{T^c=18} =$ 0.07	$\sigma^2_{T^c=24} =$ 0.05	$\overline{\sigma^2_{jt}}|_{EU} =$ 0.30
$\overline{Q} = -5.02$:	$\frac{TS}{1-s_0}$ (\$)	4625	4830	5468	5620	5685	5722	5000
	$1 - s_0$ (%)	2	10	30	34	36	37	24
$\overline{Q} = -6.30$:	$\frac{TS}{1-s_0}$ (\$)	4591	4650	4850	4901	4924	4937	
	$1 - s_0$ (%)	1	3	11	13	14	14	
$\overline{Q} = -3.74$:	$\frac{TS}{1-s_0}$ (\$)	4746	5407	7084	7429	7574	7653	
	$1 - s_0$ (%)	7	29	61	65	67	68	

Table 3 makes several important points. First, holding the strategies of the firms constant, the stent market would fail without any learning. This can be seen in the first column of the table in which the percentage of consumers having a stent implanted is close to zero for all values of $\overline{Q}$. Clinical testing provides the necessary information to make this market operate.

Second, increasing the information available to consumers by a modest amount generates significant improvements in welfare. Moving from a world in which there are no clinical trials to one in which there is a FDA required clinical trial of 6 months leads to large increases in the number of procedures performed and the surplus created per procedure. Increasing the trial length another 6 months generates smaller increases.

Third, the impact of information is dependent upon the mean quality levels of the stent distribution. In particular, the higher the quality of the average stent, the more valuable is clinical trial information. At first blush, this is somewhat counterintuitive as one make think that if the mean stent is of low quality, avoiding that stent would generate significant improvement in welfare. However, the converse is also true. If the average stent is of high quality, the greater gain to selecting the right stent ex post and thus the more valuable is clinical trial information in our setting.

Note that all of these effects are driven by symmetric yet imperfect information and not by 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 risk-aversion and the amount of information provided can be just as important as solving asymmetric information problems.

6 Welfare Implications of Regulatory Policy

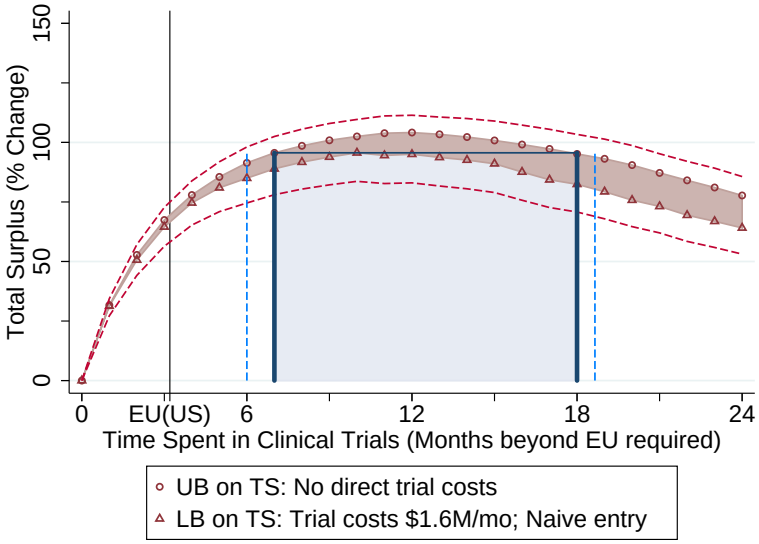
With the model and estimated structural parameters, we can examine the impact of different regulatory regimes on welfare. There is a longstanding debate over the appropriate device approval and clearance policy for medical devices which our estimates can help inform. We examine two different dimensions that could be influenced by regulatory policy: (1) the amount of information T^c that device manufacturers need to generate prior to marketing their products and (2) the precision of observational post-market learning $1/\sigma_A^2$. While the parameter values we explore are within the support of the EU and US data, the role of the model is in predicting the equilibrium responses of firms and consumers at intermediate values that we do not observe. As discussed in Section 3.3, we use relatively weak supply side assumptions to place upper and lower bounds on implications from a full equilibrium model.

6.1 Optimal Premarket Clinical Testing

Here we address a fundamental question facing any industry where new products are developed with uncertain quality: How much testing is enough?

Figure 8 plots our upper and lower bounds on expected total surplus versus the required length of time spent in clinical testing (relative to the current EU required clinical testing). Using these to then bound the optimal regulatory policy suggests that the optimal tradeoff of access vs. risk is reached between $T_c^* = [7, 18]$ (95 percent CI $[6, 19]$) months of premarket clinical testing. The width of these bounds is driven in part by the fact that the estimated total surplus is relatively flat for a wide range of trial lengths near the optimum.

Figure 8: Optimal Regulation: Red lines provide bounds on $TS(T^c)$. Blue vertical lines indicate the identified set of optimal trial lengths T^c. 95 percent confidence intervals, clustered by month, provided by dotted lines.



Outside of the flat range, however, surplus drops rapidly with zero month trials resulting in a nearly 100 percent drop relative to the optimal. At first this seems to suggest that the EU could make extremely large welfare gains by increasing its standards—until one realizes that the EU is able to free-ride off of the information being generated in trials for US entry. This makes the risk faced in the EU closer to a market with US trials of approximately 3.2 months. Thus, conditional on the US regulatory policy, the gain to the EU from increasing its standards is more modest but still substantial at 20 percent or more.

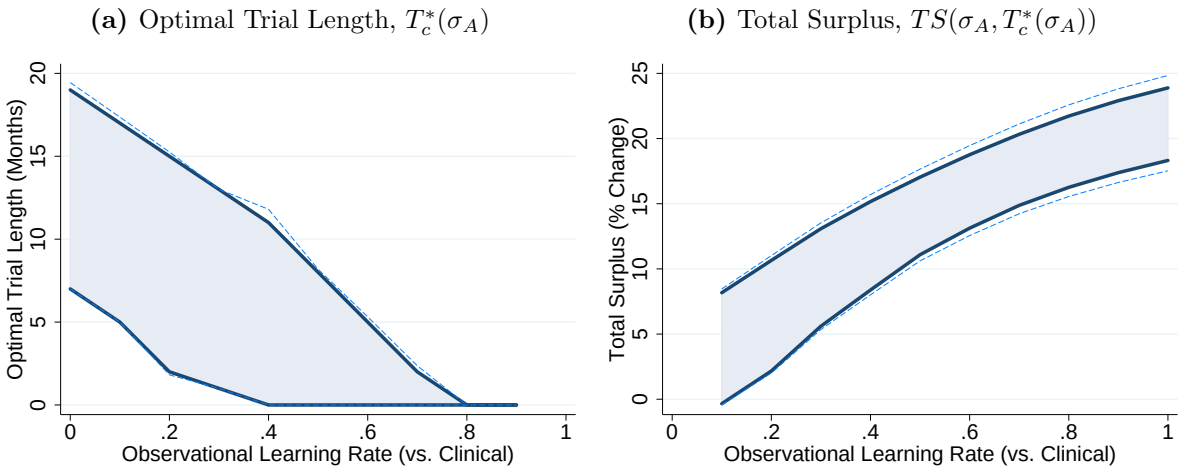
Recall that in our data the mean lag between US and EU entry is 10 months for all products and 17 months for DES. That is, we find that the current FDA policy for stents falls within our confidence interval for the optimal policy conditional on the rate of observational learning. This result speaks directly to the current policy debate over the FDA medical device approval pathway and supports the FDA argument that reductions in their standards for device approval will reduce consumer welfare. Though of course we are comparing a different time and product market, our results stand in contrast to the Peltzman (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 FDA oversight. He concludes that the Kefauver-Harris Amendments led to a significant decrease in welfare. Peltzman's analysis, however, does not speak to the optimal informational requirements pharmaceutical manufacturers should face when introducing a new molecular agent. To the best of our knowledge, the our analysis is the first that provides

an estimate of the optimal policy on the amount of information creation.

6.2 Valuing Increased Post-market Learning

One frequently proposed change to FDA regulatory policy is to relax requirements on pre-market clinical trials but increase requirements on post-market surveillance, including data collection, analysis, and reporting. This policy has a direct connection to our model in the sense that its intention is to increase the rate of post-market approval learning—in the language of our model, this means increasing the precision $1/\sigma_A^2$ of the signals that arrive outside of FDA required clinical trials. We estimated the post market learning rate is effectively zero for the set of products in our data. There are several potential reasons for this: For some products, observational learning from real world use may make it difficult to infer product quality (not having the randomization into treatment and control as in a clinical trial). For other products, though—and likely for those in our sample—the problem is simply a lack of systematic data collection and sharing of information. We analyze the potential of this policy by taking the estimated model, varying σ_A, and calculating the corresponding optimal trial length $T_c^*(\sigma_A)$ and total surplus generated $TS(\sigma_A, T_c^*(\sigma_A))$. Figure 9 displays the results, again using our bounds to generate a partially identified set of predictions.

Figure 9: 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 $1/\sigma_{A^c}^2$. 95 percent confidence intervals, clustered by month, provided by dotted lines.



When observational learning is as fast as clinical trial learning, there is no reason to run trials at all (in a symmetric information world). Total surplus is increased—up to 24 percent higher than with no observational learning—because there is no longer a tradeoff

between access and learning. The value of this increase is large. Using baseline estimates of utilization and a value of $5000 per treatment yields an estimate of $0.76 billion per year in increased welfare from this increase in post-market learning.[32]

Before reaching this extreme, as the precision of observational learning decreases (relative to clinical trial learning), it becomes optimal to require longer clinical trial periods prior to market access in order to take advantage of the faster learning rate of clinical trials. The lesson from this policy experiment is that there is merit to the argument that a requiring shorter trials with post-approval testing could improve consumer welfare. However, the gains from this policy critically depend on the rate and cost of learning via post-market surveillance.

6.3 Discussion and Implications

As discussed in the Introduction, there is an important literature measuring the optimal regulatory policies across settings as diverse as pharmaceuticals (Filson 2012; Budish, Roin, and Williams 2015), liquor distribution (Seim and Waldfogel 2013; Miravete, Seim, and Thurk 2014), and water management (Timmins 2002). In general, the estimates show that the regulator's behavior departs from the socially optimal policy by anywhere from 10 to 50 percent. Three of our main results add to this literature: First, to the best of our knowledge, our work is the first to find a regulator (the FDA, for coronary stents 2004-13) whose policies may in fact be maximizing total surplus. Second, our results are consistent with the previous literature in that we find an EU regulatory regime that is suboptimal. The EU could meaningfully increase welfare by increasing the informational criteria required to receive market access. Third, we show that adding another dimension of regulatory policies that improve market learning and reduce pre-market clinical requirements could dramatically increase social welfare.

It is important to note that our analysis models the decision to test and enter the marketplace, holding the new technology arrival process fixed and abstracting from the feedback effects of regulatory approval regime on firms' incentive to invest in the rate and direction of new product discovery. However, an important message from our analysis is that the value of a technological innovation to the marketplace depends to a large extent on the regulatory regime's informational requirements for product testing. In fact, our estimates of the value of information for medical technologies is large and comparable in magnitude to the estimates of the value of innovation itself. The welfare implications of medical technology innovation has long played a central role in health economics, but there is much less of an emphasis in the literature on the value of information regarding these new technologies. Murphy and

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

Topel (2006) show that medical technology innovation has lead to massive improvements in welfare over the twentieth century. They find gains on the order of $1.2 million per representative American in 2000. More recently, Budish, Roin, and Williams (2015) estimate that increasing effective patent lengths for cancer drugs would yield an $89 billion increase in welfare for patients diagnosed in 2003 by inducing more investment in treating cancers where the effective patent life on new drugs is relatively short. Coronary stents treat a narrower set of conditions than cancer drugs; but scaled for market size, our finding that increasing post-market learning rates can increase welfare by nearly a billion dollars per year suggest that the role of information can be comparable to the role of new technology innovation in affecting welfare.

Thus a broader lesson from our research is that the innovation process should be considered holistically from idea to consumer—the value of innovations can be significantly enhanced or diminished by the information regulators require technology firms to produce and disseminate. The availability of new medical technologies with uncertain quality can only achieve their welfare potential if firms undertake the necessary studies to document the product's clinical performance. For coronary stents, the market would shut down without some initial testing as required by the EU and benefits even more from the further testing required by the US. Thus, there are important complementarities between the value of new medical technologies and the regulatory approval product regime. Our work provides one of the first quantifications of that value.

7 Conclusion

The tradeoff between access and risk in regulating the market entry of new products is important in a variety of industries, and in particular in medical devices, where it is an active topic of policy debate in almost every country in the world. In this paper we develop a model with products introduced when quality is still uncertain, learning over time, and regulator (and manufacturer) decisions regarding product testing and market entry. We show that the empirical predictions of the model are borne out in market share data in the US and EU medical device markets and are consistent with the beliefs that the US regulatory environment is more restrictive than the EU. We then estimate the structural parameters of the model for use in welfare analysis of policy analyses affecting: (1) the length of clinical trials required before market entry and (2) observational learning after market entry.

For the set of devices on which we have data, we estimate that clinical testing is critical to market function. Without any testing, quality uncertainty plus risk aversion combine to keep most 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 at a comparable cost, then embracing recent calls for more active "post-market surveillance" could further increase total surplus by as much as 24 percent.

Firms that only enter the EU do not engage in testing beyond that required, suggesting a wedge between public and private incentives that makes regulatory requirements bind. Further understanding this wedge would require a fully specified supply side model and suggests an interesting area for future research.

Of course, our analysis is limited to coronary stents 2004-13, and extrapolating to policy for all devices should be done with care. The theoretical 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 clinical trials, and the observational learning rate for any type of device being considered.

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 play are important. Other product areas may also suffer from asymmetric information problems. Extending the model to allow for this and to further explore the extent to which certification solves asymmetry vs. amount of information problems offers another promising (and challenging) area for future research.

We also hope to have provided a building block that could be used to provide a more complete picture of how regulation affects market structure, innovation, and ultimately welfare. While estimating the welfare effects of the access/risk tradeoff for an exogenously given set of innovations is an important step towards better understanding this phenomenon, a more complete understanding would allow for the regulatory regime to effect the types of innovations firms develop for the market and vice-versa. Analysis of this type would require a significant extension to the theory and additional data on innovative activities of the firms in the market.

References

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

Ackerberg, D. A. and Rysman, M. (2005). Unobserved product differentiation in discrete-choice models: Estimating price elasticities and welfare effects. *RAND Journal of Economics*, 36(4):771–788.

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

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.

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

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

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.

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.

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.

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.

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.

Pakes, A., Porter, R., Ho, K., and Ishii, J. (2015). Moment inequalities and their application. *Econometrica*, 83(1):315–334.

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

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.

Reguant, M. (2016). Bounding equilibria in counterfactual analysis. working paper.

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.

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.

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

ELECTRONIC APPENDICES—NOT FOR PRINT PUBLICATION

Appendices

A Theory Appendix

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

A.1 Supply Bounds on Total Surplus

As discussed in Section 3.3 of the paper, developing a full supply model of entry/exit, contracting, and pricing would add frontier modeling efforts in both the bargaining and dynamic oligopoly literatures that would distract from the current focus on information and the trade-off between access and risk in regulating new products when quality of the innovations are uncertain. Instead, we develop bounds on the total surplus $TS(T^c)$ generated under any regulatory policy required pre-market clinical trials of length T^c.

Our bounds rely on relatively weak assumptions on supply side behavior:

Supply Assumption 1 (EU Entry Costs): EU entry costs are low enough such that all products developed with positive expected profits after EU testing enter.

Supply Assumption 2 (Entry Policy): The equilibrium entry policy of firms is increasing in own expected profits: $\pi_j > \pi'_j \Rightarrow Pr[Enter_{jt}|\pi_j] > Pr[Enter_{jt}|\pi'_j]$.

Supply Assumption 3 (Pricing): Prices are bounded by the marginal contribution of the product: $p_j - c_j \leq TS(\mathcal{J}) - TS(\mathcal{J} \setminus \{j\})$.

Under these assumptions, we can construct upper and lower bounds for total surplus.

Proposition UB (Upper Bound on $TS(T^c)$): The total surplus generated in equilibrium when there are no direct fixed costs of longer clinical trials ($\phi^e(T^c) = 0, \forall T^c \geq 0$) provides an upper bound for the surplus generated in equilibrium when entry costs are increasing in trial length ($\phi^e(T^c + 1) > \phi^e(T^c), \forall T^c \geq 0; \phi^e(0) = 0$).

Proof: In this case in any period t we have (note the fact that $Q_{jt}, \sigma^2_{jt}, p_{jt}, \mathcal{J}_t$ are all

functions of T^c is suppressed in the notation):

$$\begin{aligned} TS_t^{UB}(T^c) &:= \ln\left(\sum_{j\in\mathcal{J}_t(0,T^c)} e^{Q_{jt}-\theta^p p_{jt}^{\min}-\frac{\rho}{2}\sigma_{jt}^2}\right) + \sum_{j\in\mathcal{J}_t(0,T^c)} q_{jt}^{\max}(p_{jt}^{\max}-c) \\ &\geq \ln\left(\sum_{j\in\mathcal{J}_t(\phi^e(T^c),T^c)} e^{Q_{jt}-\theta^p p_{jt}^{\min}-\frac{\rho}{2}\sigma_{jt}^2}\right) + \sum_{j\in\mathcal{J}_t(\phi^e(T^c),T^c)} q_{jt}^{\max}(p_{jt}^{\max}-c) \quad (12) \\ &\geq \ln\left(\sum_{j\in\mathcal{J}_t(\phi^e(T^c),T^c)} e^{Q_{jt}-\theta^p p_{jt}-\frac{\rho}{2}\sigma_{jt}^2-\gamma|\mathcal{J}_t|}\right) \\ &\quad + \sum_{j\in\mathcal{J}_t(\phi^e(T^c),T^c)} q_{jt}(p_{jt}-c) - \phi^e(T^c)|\mathcal{J}_t\setminus\mathcal{J}_{t-1}| \quad (13) \\ &= TS_t(\phi^e(T^c),T^c). \end{aligned}$$

The inequality in (13) holds for any nonnegative ϕ^e, θ^p, γ. The inequality in (12) obtains from the fact that the choice set under nonnegative fixed costs is a weak subset of the choice set under no fixed costs $\mathcal{J}_t(\phi^e(T^c),T^c) \subseteq \mathcal{J}_t(0,T^c)$ (and the fact that in our model the utility parameters are not a function of the choice set itself). That this will hold weakly is certain due to the assumption that all potential entrants enter in the observed EU regime (which has cost zero and also $T^c = 0$), and so in expectation all of these same products would enter in the upper bound case. The subset will be strict in the case that $\phi^e(T^c)$ is large enough so that some product j finds it unprofitable to enter. *Q.E.D.*

Proposition LB (Lower Bound on $TS(T^c)$): The total surplus generated in equilibrium when entry costs are increasing in trial length ($\phi^e(T^c+1) > \phi^e(T^c), \forall T^c \geq 0; \phi^e(0) = 0$) is bounded from below by the total surplus generated in equilibrium with these same fixed costs, but where firms follow a naive entry strategy that assumes all other firms will enter in equilibrium.

Proof: In this case in any period t we have (note the fact that $Q_{jt}, \sigma_{jt}^2, p_{jt}, \mathcal{J}_t$ are all

functions of T^c is suppressed in the notation):

$$\begin{aligned}
TS_t^{LB}(T^c) &:= \ln\left(\sum_{j\in\mathcal{J}_t((\phi_j^e(T^c),0_{-j}),T^c)} e^{Q_{jt}-\theta^p p_{jt}^{\max}-\frac{\rho}{2}\sigma_{jt}^2-\gamma|\mathcal{J}_t(0_{-j})|}\right) \\
&\quad + \sum_{j\in\mathcal{J}_t((\phi_j^e(T^c),0_{-j}),T^c)} q_{jt}^{\min}(p_{jt}^{\min}-c) - \phi^e(T^c)|\mathcal{J}_t(0_{-j})\setminus\mathcal{J}_{t-1}(0_{-j})| \\
&\leq \ln\left(\sum_{j\in\mathcal{J}_t((\phi_j^e(T^c),0_{-j}),T^c)} e^{Q_{jt}-\theta^p p_{jt}-\frac{\rho}{2}\sigma_{jt}^2-\gamma|\mathcal{J}_t(0_{-j})|}\right) \\
&\quad + \sum_{j\in\mathcal{J}_t((\phi_j^e(T^c),0_{-j}),T^c)} q_{jt}(p_{jt}-c) - \phi^e(T^c)|\mathcal{J}_t(0_{-j})\setminus\mathcal{J}_{t-1}(0_{-j})| \quad (14) \\
&\leq \ln\left(\sum_{j\in\mathcal{J}_t(\phi^e(T^c),T^c)} e^{Q_{jt}-\theta^p p_{jt}-\frac{\rho}{2}\sigma_{jt}^2-\gamma|\mathcal{J}_t|}\right) \\
&\quad + \sum_{j\in\mathcal{J}_t(\phi^e(T^c),T^c)} q_{jt}(p_{jt}-c) - \phi^e(T^c)|\mathcal{J}_t\setminus\mathcal{J}_{t-1}| \quad (15) \\
&= TS_t(\phi^e(T^c),T^c).
\end{aligned}$$

The inequality in (14) obtains for any nonnegative θ^p. The inequality in (15) obtains from the fact that the choice set under the full equilibrium is a subset of the choice set under the naive equilibrium and that there is never over entry in either equilibrium: Under the assumption that all potential products entered in the zero cost case, potential profits (after fixed costs) will always be weakly higher in the full equilibrium, increasing the probability of entry. Under the additional assumption that prices are bounded by marginal contributions (which is the case, for example, in bargaining models previously used in the literature), the surplus gain is always greater than the fixed cost of entry. *Q.E.D.*

These upper and lower bound scenarios are equivalent to each other and to the full equilibrium at $T^c = 0$. Both will become further from the true equilibrium as the costs of entry increase. We go through the details of computing each bound in Appendix C.2.

A.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 risk and access in the model, it is helpful to consider a simple case

where testing and entry are 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:

$$\begin{aligned} TS_t(T^c+1) - TS_t(T^c) &= \ln\left(\frac{\sum_{j\in\mathcal{J}_t(T^c+1)} e^{Q_{jt}-\frac{\rho}{2}\sigma^2_{jt}(T^c+1)}}{\sum_{j\in\mathcal{J}_t(T^c)} e^{Q_{jt}-\frac{\rho}{2}\sigma^2_{jt}(T^c)}}\right) - \phi^e \left|\mathcal{J}^e_t(T^c+1)\setminus\mathcal{J}^e_t(T^c)\right| \\ &= \ln\left(\frac{\sum_{j\in\mathcal{J}_t(T^c+1)} e^{Q_{jt}-\frac{\rho}{2}\sigma^2_{jt}(T^c+1)}}{\sum_{j\in\mathcal{J}_t(T^c)} e^{Q_{jt}-\frac{\rho}{2}\sigma^2_{jt}(T^c)}}\right) \qquad (16) \\ &= \frac{\rho}{2}\left(\sigma^2(T^c)-\sigma^2(T^c+1)\right) - \ln\left(\frac{\sum_{j\in\mathcal{J}_t(T^c)} e^{Q_{jt}}}{\sum_{j\in\mathcal{J}_t(T^c+1)} e^{Q_{jt}}}\right) \qquad (17) \end{aligned}$$

where (16) follows from no fixed costs, and (17) follows from no observational learning. Then averaging over any period of time $t = 1, ..., T$ and recognizing $\phi^e = 0 \Rightarrow \mathcal{J}_t(T^c) = \mathcal{J}_{t+1}(T^c+1)$ so that the log sum term is telescoping yields (9).

A.3 Learning and Risk Aversion Predictions for Shares

Equation (10) showed the relationship between the estimated product qualities Q_{jt}, uncertainty about those estimates σ_{jt}, and the function of market shares implied by the demand system $\ln(s_{jt}/s_{0t})$. We reprint that equation here, with any other explanatory variables over which learning is not directly occurring (e.g. price, the Ackerberg-Rysman adjustment) partialed out, as they would be in the demand estimation.:

$$\ln(s_{jt}/s_{0t}) + \theta_p p_{jt} + \gamma \ln(|\mathcal{J}_t|) = \tilde{\delta}_{jt} := Q_{jt} - \frac{\rho}{2}\sigma^2_{jt} \ . \qquad (18)$$

We then make two claims about how this relationship can be used to examine the presence of learning and risk aversion. We provide proofs for each of those claims here.

Prediction 1 (Learning): If initial product quality is uncertain $\sigma^2_Q > 0$, then learning $1/\sigma^2_A > 0$ implies that expected volatility in product-specific quality estimates converge by decreasing with age $E_j|\tilde{\delta}_{ja} - \tilde{\delta}_{ja+1}| \searrow^{a\to\infty} 0$ or $Var_j(\tilde{\delta}_{ja}) \searrow^{a\to\infty} \sigma^2_Q$.

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 $Q_{ja} \longrightarrow^{a\to\infty} Q_j$ and the convergence of uncertainty about that estimate to zero $\sigma^2_{ja} \searrow^{a\to\infty} 0$. Our further claim is that evidence for this learning will be found by looking at measures of volatility of

the market shares. First, consider the month-to-month absolute volatility:

$$\begin{aligned} |\tilde{\delta}_{ja} - \tilde{\delta}_{ja+1}| &= |(Q_{ja} - \frac{\rho}{2}\sigma_{ja}^2) - (Q_{ja+1} - \frac{\rho}{2}\sigma_{ja+1}^2)| \\ &\geq |Q_{ja} - Q_{ja+1}| + \frac{\rho}{2}|\sigma_{ja}^2 - \sigma_{ja+1}^2| \end{aligned} \tag{19}$$

where the right hand side of the equation is clearly positive (strictly for $\sigma_Q^2 > 0$), and as an affine combination of two converging sequences must converge. $E_j[Q_{ja}] = \bar{Q}, \forall a$ and $\sigma_{ja}^2 \searrow^{a\to\infty} 0$ ensure that convergence of the expectation over products j is a decreasing sequence (strictly for $\rho > 0$).

Second, consider volatility across products:

$$\begin{aligned} \lim_{a\to\infty} Var_j(\tilde{\delta}_{ja}) &= \lim_{a\to\infty} Var_j(Q_{ja}) - 0 \\ &= \lim_{a\to\infty} Var_j\left(\frac{\sigma_0^2}{a\sigma_0^2 + \sigma_A^2}\sum_{\tau=1}^{a} Q_j + \epsilon_{j\tau} + \frac{\sigma_0^2}{a\sigma_0^2 + \sigma_A^2}Q_{j0}\right) \\ &= \lim_{a\to\infty} E_j\left[\frac{a\sigma_0^2}{a\sigma_0^2 + \sigma_A^2}(Q_j - \bar{Q}) + \frac{\sigma_0^2}{a\sigma_0^2 + \sigma_A^2}\sum_{\tau=1}^{a}\epsilon_{j\tau} + \frac{\sigma_0^2}{a\sigma_0^2 + \sigma_A^2}(Q_{j0} - \bar{Q})\right]^2 \\ &= E_j(Q_j - \bar{Q})^2 = \sigma_Q^2 \end{aligned} \tag{20}$$

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 ($E_j[\tilde{\delta}_{ja}] \nearrow^{a\to\infty} \overline{Q}$).

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

$$\begin{aligned} \lim_{a\to\infty} E_j[\tilde{\delta}_{ja}] &= \lim_{a\to\infty} E_j[Q_{ja}] - \frac{\rho}{2}\sigma_{ja}^2 \\ &= E_j[Q_j] - 0 \\ &= \bar{Q} \end{aligned} \tag{21}$$

where the second line follows from the convergence of $\{Q_{ja}\}$ and $\{\sigma_{ja}^2\}$. $\sigma_{ja}^2 \searrow^{a\to\infty} 0$ and $\rho > 0$ guarantee the convergence is increasing and strictly so. *Q.E.D.*

B Data Appendix

B.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 U.S. and E.U. 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. Finally, there is some misreporting in the survey. The following tables illustrate how key sample summary statistics compare across the cleaning steps for the E.U. and U.S. 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 E.U. and U.S. 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 (weakly) 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 at least two diagnostic angiography procedures per stent implant. The third row drops hospital-months if their total quantity of stents exceeds twice the number of diagnostic angiographies in that hospital-month. The angiography count itself could be suspect. 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. There are no products for which the U.S. approval dates could not be ascertained, so this row is missing from the U.S. table.

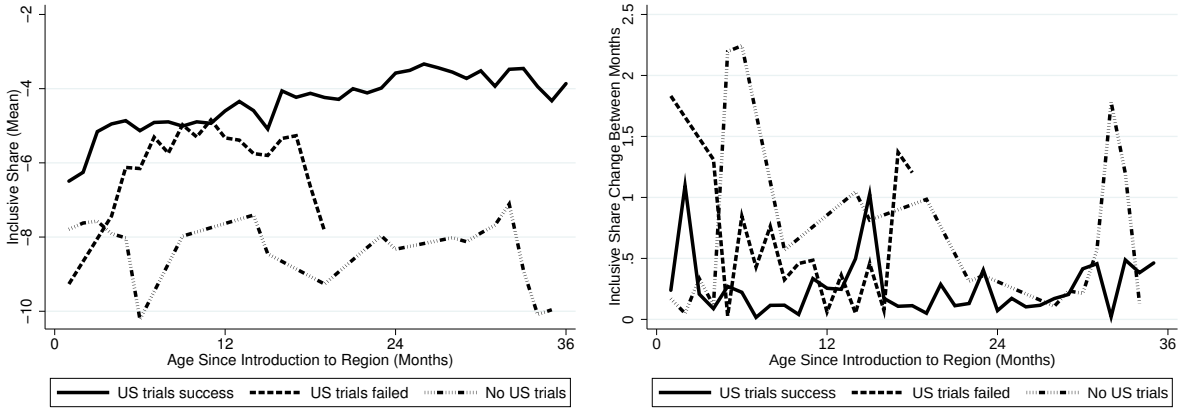
The final row in each table reports summary statistics for the final sample, which drops

all observations that meet one or more of the dropping criteria described above.

B.2 Evidence of learning from individual products

Figure 10: 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 10 provides two types of evidence of this variation. First, the figures in the panels provide patterns for a few individual products illustrate 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.

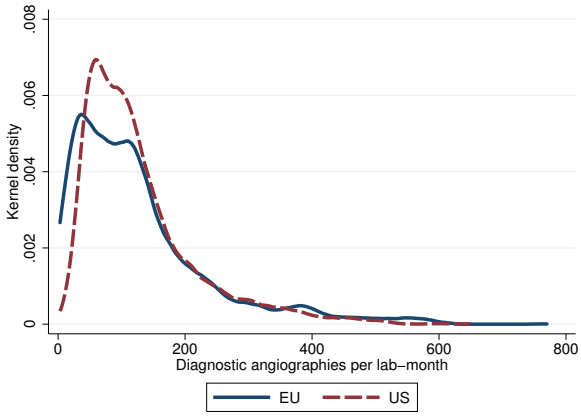
The patterns documented previously regarding decreases in volatility and increasing mean usage with age might be worrisome if they were driven by increasing usage for all product with age that then asymptotes as in a diffusion process. The table on the raw usage changes show this is not the case—there is a large fraction of changes that are "bad news" for products.

B.3 Regulatory Differences Are Not Driven by Differences in Disease Incidence or Treatment Preferences

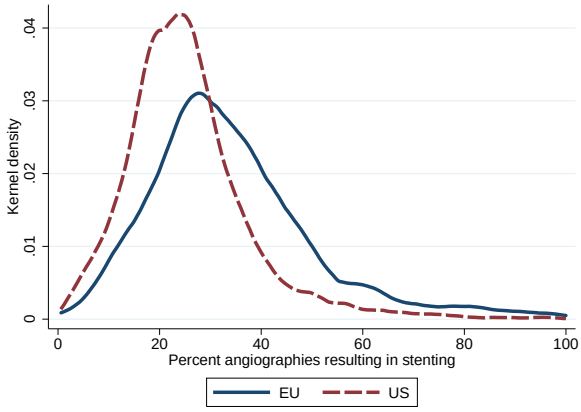
In theory it could be that the differences in usage patterns between the US and EU are driven by differences in disease incidence, preferences for angioplasty and stents, or variation in price setting regimes between the US and EU. However, all the evidence that we have been able to gather indicates 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.[33] This modest differential seems unlikely to account for the stark differences of entry rates between the two regions.

Figure 11: 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



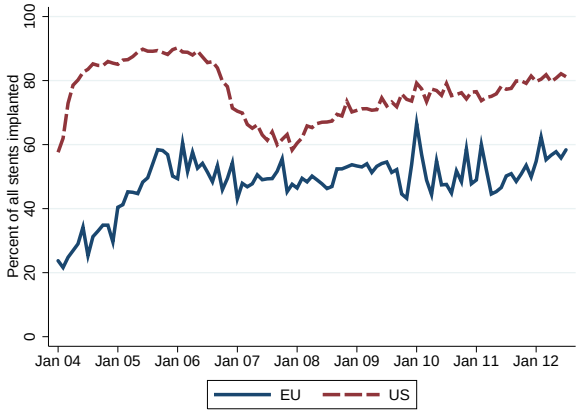
Prior to performing an angioplasty in which a stent may be inserted, the patient must undergo a diagnostic angiography. In this procedure, the blood flow through the coronary artery is visualized and this information is used to determine whether the patient should receive a stent or some other medical 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 11 documents the

[33] 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 12: Comparison of usage and price patterns EU vs. US.

(a) DES Share of All Stents



(b) Average Price Paid

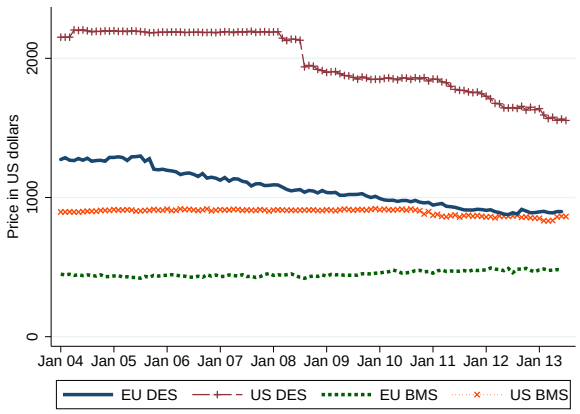


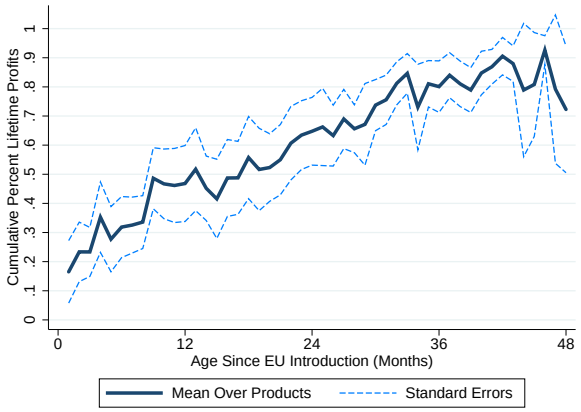
Figure 12 documents that DES usage as a percentage of all stents used is lower in the EU but follows similar patterns to the US over time. If the increased DES entry in the EU was driven by higher demand, we would expect the opposite pattern. Figure 12 also shows that the prices and hence profits per 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 is driven by differences in regulation and not underlying demand.

B.4 Distribution of Profits Over Product Lifetime and Across Products

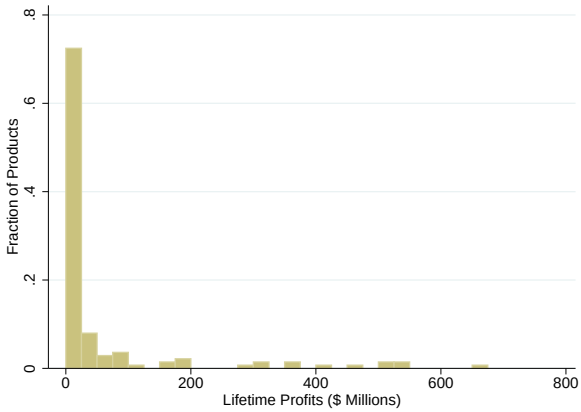
The counterfactual lower bounds with fixed costs of entry require calculation of expected lifetime profits under the assumption that all firms who enter in the EU do enter in equilibrium. This number can be directly acquired from the EU data for the 41 of 109 products that both enter and exit the market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

Figure 13: 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

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. 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. And further the products that are marginal in our counterfactuals, in the sense that they exit as entry costs increase, are also marginal in the computation of consumer surplus in that which enter (and even to some extent how many enter) does not

greatly affect total welfare.

The distribution of estimated lifetime profits makes it clear that many products with quite low profitability enter the EU, supporting 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.

C Estimation and Counterfactual Algorithms

C.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. Compute mean utilities $\delta_{jt} = \ln(s_{jt}/s_{0t})$ for all product-months.
2. Construct an initial estimator for uncertainty immediately after EU testing $\sigma^2_{ja=1}$ using the empirical equivalent from the distribution of δ_{jt}.
3. Guess initial values for learning precisions $\sigma_{\mathbf{A}} := (\sigma_A, \sigma_{A^c})$.
4. Compute the full vector of σ^2_{jt} implied by $\sigma^2_{ja=1}$, the learning precision parameters, and which products are in trials when.
5. Least squares then gives an estimator for ρ and the product qualities Q_j as a function of the learning parameters, where $[Q_j; \rho](\sigma_{\mathbf{A}}) = (X'X)^{-1}X'\delta$ with $X = [1_j, -\frac{1}{2}\sigma^2_{jt}]$. (Here Q_j represents the vector of coefficients on product dummy variables, and 1_j the matrix of product dummy variables.)
6. We need to make sure that the distribution of $Q_{jt} := \delta_{jt} + \frac{\rho}{2}\sigma^2_{jt}$ is consistent with the $\sigma^2_{ja=1}$ by recomputing $\sigma^2_{ja=1}(Q_{jt})$ and repeating 4-6 until $\sigma^2_{ja=1}$ converges.
7. Compute the residuals $\xi_{jt} = \delta_{jt} - Q_j + \frac{\rho}{2}\sigma^2_{jt}$.
8. Evaluate GMM objective function based on $E[\xi' Z] = 0$ where $Z = \left[\frac{1}{a_{jt}} \;\; \frac{1}{a^2_{jt}}\right]$.
9. Repeat 4-8 until we find the value of $\sigma_{\mathbf{A}}$ that minimizes the GMM objective function.

C.2 Optimal regulation counterfactual algorithm

The advantage of the upper and lower bounds we have defined on total surplus is that they can be calculated using only the data and demand/learning model estimates. For each potential $T^c = 0, 1, ..., 24$ we calculate the upper and lower bounds as follows:

Upper Bound

1. Given T^c, restrict sample to products that would be active in each month.
2. Use demand/learning model to compute total surplus over the sample period.

Lower Bound

1. Given T^c, restrict sample to products that would be active in each month.
2. Given $\phi^e = T^c \times 1.6E6$, restrict sample to products that would enter, under the naive assumption that firms assume other products enter as if $\phi^e = 0$.
3. Use demand/learning model to compute total surplus over the sample period, remembering to subtract fixed costs when products enter.

Note from the proofs that both of these algorithms require $p^{\max}$ and $p^{\min}$ vectors such that the total surplus bounds hold once price effects on the extensive margin are accounted for. In theory, these could be very extreme, e.g. $p^{\max} = c + (TS(\mathcal{J}) - TS(\mathcal{J} \setminus \{j\}))$ and $p^{\min} = c$, and provide valid bounds, but such bounds may not be tight enough to be informative. In our estimation routine, we use a reduced-form pricing equation $p_{jt} = c + \beta_{jt}(TS_t(\mathcal{J}) - TS_t(\mathcal{J} \setminus \{j\}))$ and assumed cost equal to the minimum price observed in the data to compute implied portions of the marginal contributions that go to each product β_{jt}. We then use the maximum and minimum marginal contributions that would occur under the full entry and naive entry cases considered in the bounds to compute maximum and minimum prices. We have also explored simple rules such as considering prices plus or minus 10 percent of those observed, and our results remain unchanged.

C.3 Observational learning counterfactual algorithm

Because we only specify bounds on total surplus for any trial length T^c, we obtain only bounds on the optimal trial length under any parameter values $[T^c_{LB}(\sigma_{\mathbf{A}}), T^c_{UB}(\sigma_{\mathbf{A}}),]$. Thus for each potential value of observational learning precision $1/\sigma^2_A = 0, 1/10\sigma^2_{A^c}, 2/10\sigma^2_{A^c}, ..., 1/\sigma^2_{A^c}$ we calculate the bounds on optimal trial time and surplus generated by these trial times as follows:

1. Given $1/\sigma_A^2$, calculate the upper and lower bounds on surplus generated for $T^c = 0, 1, ..., 24$ as done previously for the zero observational learning case.

2. $T^c_{LB}(\sigma_{\mathbf{A}})$ will be the maximum T^c such that the upper bound total surplus is less than the maximum of the lower bound total surplus (among the T^c below that at which the lower bound surplus is maximized).

3. $T^c_{UB}(\sigma_{\mathbf{A}})$ will be the minimum T^c such that the upper bound total surplus is less than the maximum of the lower bound total surplus (among the T^c above that at which the lower bound surplus is maximized).

4. The tightest bounds on surplus created in this case are simply the max of the upper bound surplus and max of the lower bounds surplus.

D Robustness Checks and Additional Analyses

D.1 Measuring Learning Using Variance Across Products, $Var(\delta_{jt})|_a$

In the paper we use the volatility measure based on within product usage over time $Mean_{j|a}$ $|\Delta_t \ln(s_{jt}/s_{0t})|$ to detect reduced form evidence of learning. We prefer this measure for the reduced form analyses because it is intuitively connected to learning and converges to zero as learning occurs. A drawback, however, is that this measure will only show evidence of learning if consumers are risk averse, and it could be more susceptible to mistaking some types of product diffusion for learning. Here we show evidence of learning using the more robust (though perhaps less intuitive) measure of across product variance in usage conditional on age $StdDev_{j|a}$ $\ln(s_{jt}/s_{0t})$.

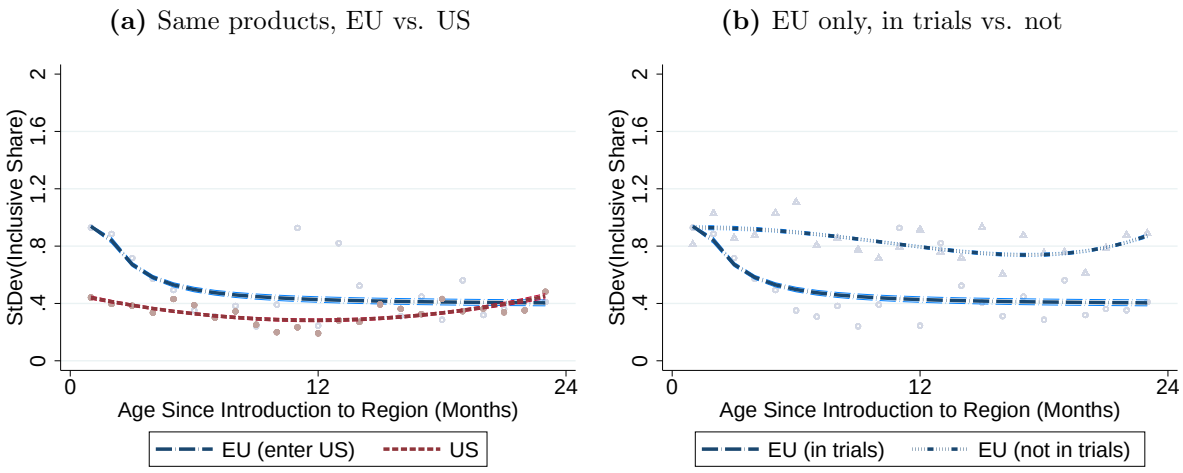
As demonstrated in the model predictions, this will converge to the unconditional product quality variation σ_Q with learning, independent of risk aversion. This is the variation we use to measure learning in our structural estimation. Figure 14 shows the reduced form patterns in this measure for the matched US / EU and EU in trials / not comparisons, which are the heart of our learning evidence.

D.2 Sample with Overlapping Support on $Q_{a_j=1}$

D.3 Looking at EU within Countries

At the beginning of the data section we noted that over seventy percent of all products released in the EU were observed in all EU countries for which we have reasonable sample sizes (France, Germany, Italy, Spain, and UK), but that even this might undercount country-level entry due to sampling error for little used products with samples of around twenty

Figure 14: Learning evidence using StdDev$_{j|a}$ $\ln(s_{jt}/s_{0t})$



hospitals. Because of this potential for sampling error, and the fact that regulation of entry is for the entire EU market, we prefer the EU as the level of analysis in the paper.

We have, however, run the same analysis at the country level for the five larger countries, and we find our results unchanged. We take this as further evidence that the patterns we observe at the EU level are indeed capturing learning and not a complicated diffusion process across countries. Figure 16 displays the results for Italy. Other country results are available from the authors upon request.

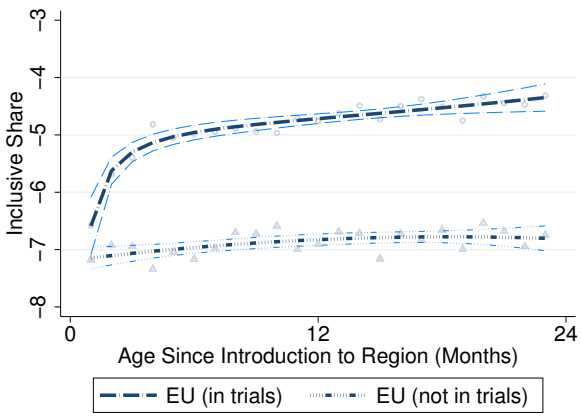
Figure 16 compares products in Italy that (1) undergo clinical trials for US release; and (2) that never undergo trials beyond those required for EU introduction. The evidence is qualitatively identical and numerically very close to that from the full EU sample. The mean of mean utility estimates in panel (a) reveal that all of the evidence of learning is driven by those products in clinical trials. The curve is flat for products not in trials, which not only rules out observational learning for these products, but also rules out diffusion driven by marketing, sales, or distribution of new-to-the-world products. The mean absolute differences in panel (b) reconfirm that the learning in the EU is driven by the products undergoing clinical trials. Importantly, they also refute the argument of selection on uncertainty, as the EU products not in trials begin the same as the products in trials, but remain flat over time, suggesting that there is plenty of uncertainty for these products, but no learning.

D.4 Looking at EU within Hospital

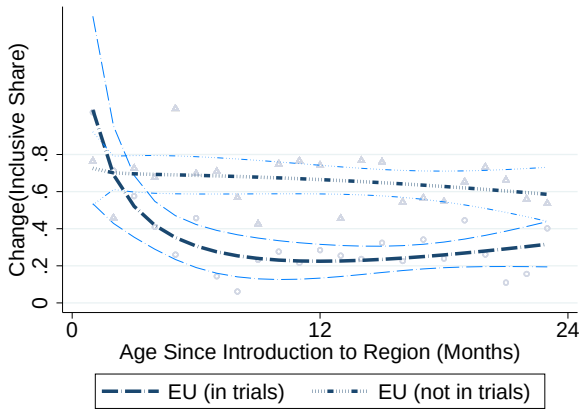
While our original data is at the product-hospital-month unit of observation, we aggregate to the EU level in order to account for the fact that because of physician preferences (and

Figure 15: EU only, products in trials vs. not. (overlapping samples) This robustness check corrects for the fact that many products in the EU may not be comparable to products that eventually enter the US, using samples for products in trials and not that *overlap in their initial quality estimates at* $a_j = 1$ (criteria $-8.92 < \ln(s_{jt}/s_{0t})|_{a_j=1} < -6.24$).

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



(b) $\text{Mean}_{j|a}\ |\ln(s_{jt}/s_{0t}) - \ln(s_{jt+1}/s_{0t+1})|$



	$x_{a=1}$	$x_{a=24}$	$x_{24} - x_1$	$(x_{24}^{trial} - x_1^{trial})$ $-(x_{24}^{not} - x_1^{not})$
$\text{Mean}_{j\|a}^{EU\|trials}\ \ln(s_{jt}/s_{0t})$	-6.58	-4.31	2.26	1.83
	(0.29)	(0.25)	(0.38)	(0.45)
$\text{Mean}_{j\|a}^{EU\|not}\ \ln(s_{jt}/s_{0t})$	-7.18	-6.74	0.44	
	(0.16)	(0.24)	(0.29)	
$\text{Mean}_{j\|a}^{EU\|trials}$ $\|\ln(s_{jt}/s_{0t}) - \ln(s_{jt+1}/s_{0t+1})\|$	1.03	0.40	-0.63	-0.40
	(0.38)	(0.14)	(0.41)	(0.41)
$\text{Mean}_{j\|a}^{EU\|not}\ \|\ln(s_{jt}/s_{0t}) - \ln(s_{jt+1}/s_{0t+1})\|$	0.76	0.54	-0.23	
	(0.11)	(0.10)	(0.14)	

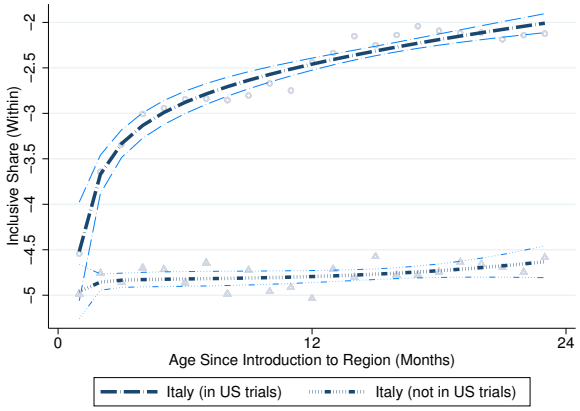
$N = 529$ product-month observations (all in EU; restricted to overlapping support of $Q_{j,a=1}$). Standard errors clustered by month $N_t = 114$ in parentheses.

possibly contracting concerns), not all hospitals use all stents. Thus hospital level revealed preference measures will be conditional on this selection. They will also fail to capture extensive margin changes in what is used at a given hospital and suffer the most severely from sampling error for rarely used products.

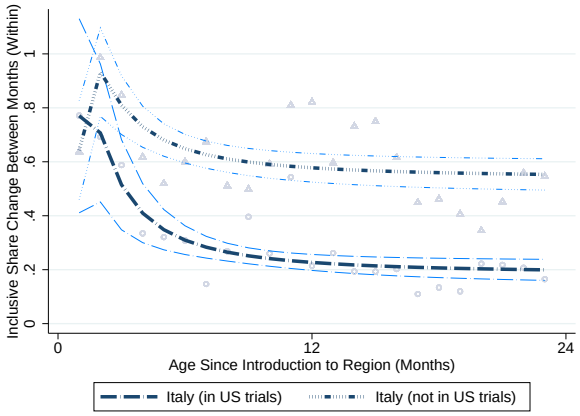
Despite these drawbacks to using the hospital level data to estimate welfare-relevant parameters, one would still hope to find (potentially muted) evidence of learning within hospitals for the products they purchase regularly over time. To explore this, we run the same analysis, instead using usage data at the hospital level to calculate $\ln(s_{jht}/s_{0ht})$ and take summary statistics by age over these product-hospital-month observations. Figure 17

Figure 16: Italy only, products in trials vs. not.

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



(b) $\text{Mean}_{j|a}\ |\ln(s_{jt}/s_{0t}) - \ln(s_{jt+1}/s_{0t+1})|$



	$x_{a=1}$	$x_{a=24}$	$x_{24} - x_1$	$(x_{24}^{trials} - x_1^{trials}) - (x_{24}^{not} - x_1^{not})$
$\text{Mean}_{j\|a}^{EU\|_{trials}}\ \ln(s_{jt}/s_{0t})$	-4.54	-2.12	2.42	2.02
	(0.31)	(0.11)	(0.33)	(0.39)
$\text{Mean}_{j\|a}^{EU\|_{not}}\ \ln(s_{jt}/s_{0t})$	-4.99	-4.59	0.40	
	(0.16)	(0.17)	(0.22)	
$\text{Mean}_{j\|a}^{EU\|_{trials}}\ \|\Delta_t \ln(s_{jt}/s_{0t})\|$	0.77	0.16	-0.61	-0.52
	(0.19)	(0.04)	(0.20)	(0.22)
$\text{Mean}_{j\|a}^{EU\|_{not}}\ \|\Delta_t \ln(s_{jt}/s_{0t})\|$	0.64	0.55	-0.09	
	(0.09)	(0.09)	(0.13)	

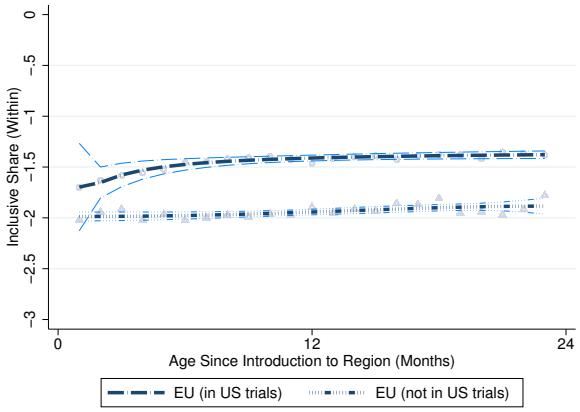
$N = 831$ product-month observations (all in Italy). Standard errors clustered by month $N_t = 114$ in parentheses.

displays the results.

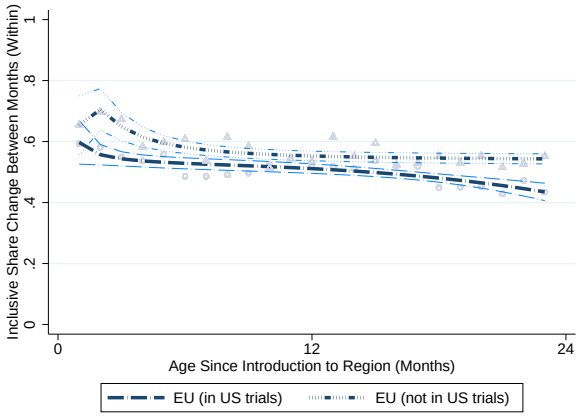
The evidence at the hospital level still shows signs of learning, though the evidence is muted and noisier than at the EU level. The numbers are difficult to compare directly to those at the EU level because of the selection on which hospitals are observed using which products (which is consistent is overall higher log share terms) and also the increased sampling variation (which could be the cause of the first differences not moving closer to zero). The differences-in-differences estimates have the right sign but are noisy with confidence intervals including zero. This is in part because of relatively noisier estimates at $a = 1$ (which itself is related to uncertainty)—while the in-trial vs. not groups are statistically indistinguishable at $a = 1$, the in-trial group has statistically lower volatility and higher usage at $a = 24$.

Figure 17: Hospital Level—EU only, products in trials vs. not.

(a) $\text{Mean}_{jh|a}\ \ln(s_{jht}/s_{0ht})$ (within product)



(b) $\text{Mean}_{jh|a}\ |\ln(s_{jht}/s_{0ht}) - \ln(s_{jht+1}/s_{0ht+1})|$



	$x_{a=1}$	$x_{a=24}$	$x_{24} - x_1$	$(x_{24}^{trials} - x_1^{trials}) - (x_{24}^{not} - x_1^{not})$
$\text{Mean}_{j\|a}^{EU\|_{trials}}\ \ln(s_{jt}/s_{0t})$	-1.70	-1.38	0.32	0.22
	(0.22)	(0.06)	(0.23)	(0.25)
$\text{Mean}_{j\|a}^{EU\|_{not}}\ \ln(s_{jt}/s_{0t})$	-2.02	-1.92	0.10	
	(0.09)	(0.05)	(0.11)	
$\text{Mean}_{j\|a}^{EU\|_{trials}}\ \|\Delta_t \ln(s_{jt}/s_{0t})\|$	0.59	0.43	-0.16	-0.06
	(0.03)	(0.03)	(0.04)	(0.07)
$\text{Mean}_{j\|a}^{EU\|_{not}}\ \|\Delta_t \ln(s_{jt}/s_{0t})\|$	0.65	0.55	-0.10	
	(0.05)	(0.03)	(0.06)	

$N = 14,500$ product-hospital-month observations (all in EU). Standard errors clustered by month $N_t = 114$ in parentheses.

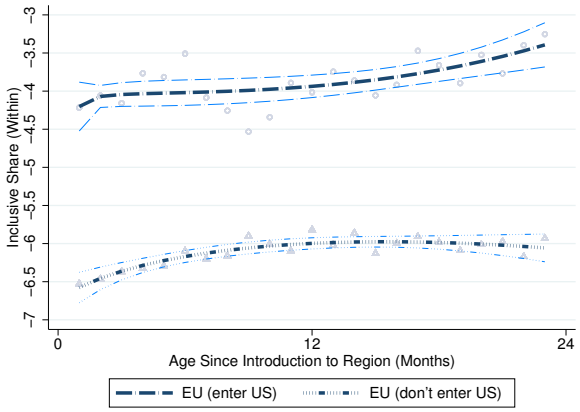
D.5 Placebo Check on Device With Similar US/EU Approval Requirements

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

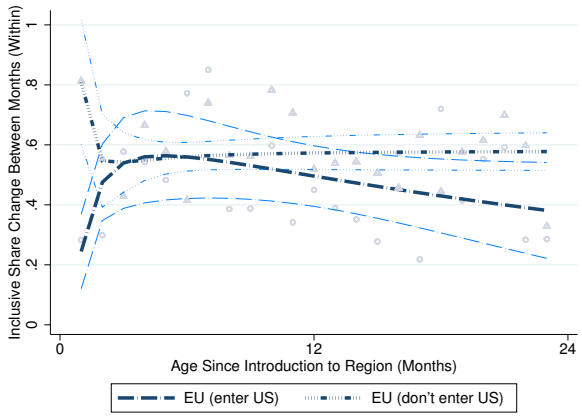
that only a few companies enter the US market for the purpose of selling balloons only. During our ten year sample, 40 manufacturers sell 113 different balloons in the EU and 6 manufacturers sell 40 different balloons in the US. Thus we can execute our same research design on balloons, with the expectation of no differential learning between products that are EU only versus those that enter the US as well.

Figure 18: PTCA Balloons—EU only, products that enter US vs. not.

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



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



	$x_{a=1}$	$x_{a=24}$	$x_{24} - x_1$	$(x_{24}^{trials} - x_1^{trials}) - (x_{24}^{not} - x_1^{not})$
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)
Mean$_{j\|a}^{EU\|_{not}}$ $\ln(s_{jt}/s_{0t})$	-6.52	-5.93	0.60	
	(0.16)	(0.18)	(0.21)	
Mean$_{j\|a}^{EU\|_{trials}}$ $\|\Delta_t \ln(s_{jt}/s_{0t})\|$	0.28	0.29	0.00	0.21
	(0.04)	(0.07)	(0.08)	(0.14)
Mean$_{j\|a}^{EU\|_{not}}$ $\|\Delta_t \ln(s_{jt}/s_{0t})\|$	0.81	0.60	-0.21	
	(0.11)	(0.13)	(0.11)	

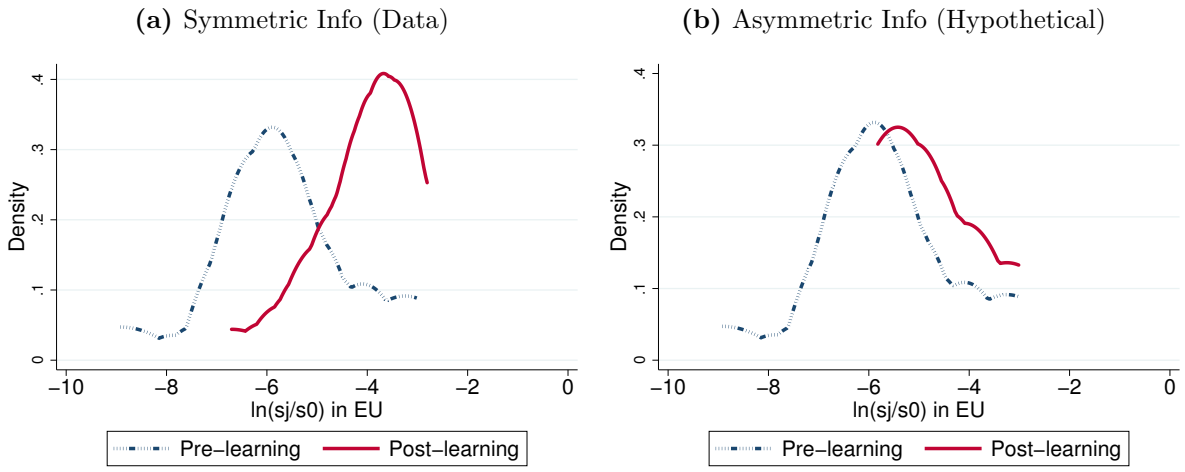
$N = 789$ product-month observations (all in EU). Standard errors clustered by month $N_t = 114$ in parentheses.

Figure 18 shows the results of this placebo test, comparing EU data for products that do 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. Except for what appears to be an outlier shock from month one to two for usage of EU only balloons, 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.

D.6 Testing Symmetric vs. Asymmetric Information

Our test of information symmetry in Figure 5 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 19 illustrates these ideas graphically.

Figure 19: 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 5 in the paper, 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 5 fails to reject the null hypothesis of symmetric learning.

D.7 Model fit and comparison with fixed effects least-squares estimates

Figure 20: Comparison of estimates from fixed effect and learning models. Left panel (a) plots the estimated distribution of product qualities from the parametric learning model and age fixed effects model. Right panel (b) plots the estimated discount due to uncertainty versus product age for the two models.

(a) Product Quality Density, $f(Q_j)$ **(b)** Uncertainty Discount, $-\frac{\rho}{2}\sigma_{jt}^2$

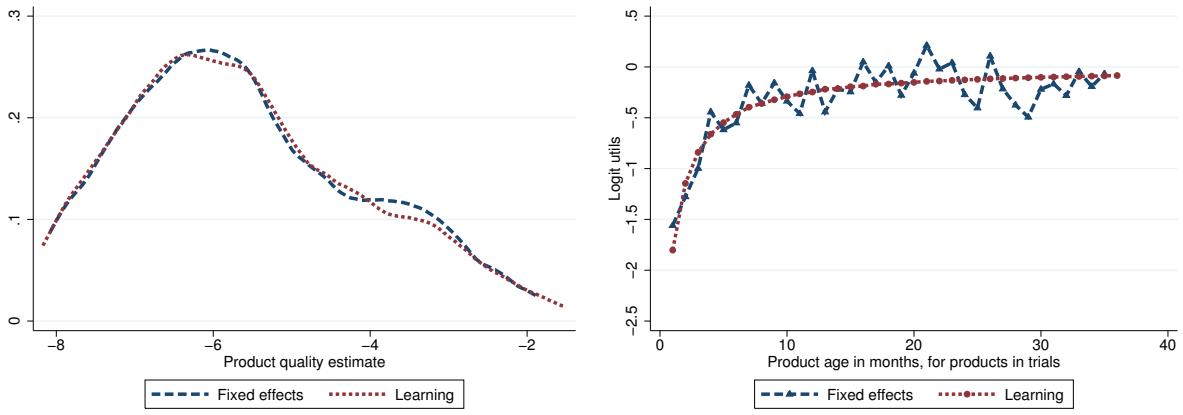


Figure 20 shows the estimated distribution of product qualities Q_j and uncertainty discounts $-\frac{\rho}{2}\sigma_{jt}^2$ for both the learning model and the more flexible model with product and age fixed effects. The message here is that despite its parsimony, the simple learning model fits the data well and its fit is comparable to the much more nonparametric fixed effects model (R^2 of 0.95 vs. 0.98), so our results are not driven by the functional form of the learning model.