Essays in the Economics of Health Care and the Regulation of Medical Technology

Dissertation Advisors:
Professor Amitabh Chandra
Professor David Cutler

Author:
Ariel Dora Stern

Essays in the Economics of Health Care and the Regulation of Medical Technology

Abstract

The first chapter of this dissertation explores how the regulatory approval process affects innovation incentives in medical technologies. While prior studies of medical innovation under regulation have found an early mover regulatory advantage for drugs, I find the opposite to be true for medical devices. Using detailed data on over three decades of high-risk medical device approval times in the United States, I show pioneer entrants spend approximately 34 percent (7.2 months) longer in the approval process than the first follow-on innovator. Back-of-the-envelope calculations suggest that the opportunity cost of capital of a delay of this length is upwards of 7 percent of the total cost of bringing a new device to market. I consider how different types of regulatory uncertainty affect approval times and find that a product's technological novelty is largely unrelated to time spent under review. In contrast, uncertainty about application content and format appears to play a large role: when objective guidelines for evaluation are published, approval times quicken for subsequent entrants. Finally, I consider how the regulatory process affects firms' market entry strategies and find that financially constrained firms are less likely to enter new device markets as pioneers.

The second chapter considers the voting behaviors of individuals on expert advisory committees at the U.S. Food and Drug Administration (FDA). Individuals on these committees sometimes have financial conflicts of interest, which may result in a principal-agent dilemma. Committee members also have institutional affiliations, a history of co-authoring relationships, and different areas of expertise, which may influence voting behavior. Using

data on over 1500 uniquely identified individuals at 110 new product meetings over a seven-year period, I find that in a simple analysis, financially conflicted individuals are 18 percent more likely to vote favorably for new medical devices, but no more likely to vote favorably for new drugs. This pattern is driven by individuals voting favorably for competitors' products and is consistent with a regulatory setting in which conflicted individuals help "pave the way" for subsequent entrants to move swiftly through the regulatory approval process. I then describe a preliminary model of individuals' voting behaviors which incorporates both direct conflicts of interest and peer effects. Using this framework, I find reduced form evidence that the composition of an advisory committee adds additional predictive power to a model of how individuals vote. Peer effects models suggest that at high (low) levels of in favor voting within a meeting, the simple analysis is likely to to understate (overstate) bias related to conflict of interest.

The final chapter considers drivers of regional variations in healthcare spending in the United States and is based on joint work with David Cutler, Jonathan Skinner, and David Wennberg. There is considerable controversy about the causes of regional variations in healthcare expenditures. We use a set of detailed vignettes from patient and physician surveys linked to Medicare expenditures at the level of the Hospital Referral Region to test whether patient demand-side factors or physician supply-side factors better explain regional variations in Medicare spending. We find patient demand is relatively unimportant in explaining variations. Physician organizational factors (such as peer effects) matter, but the single most important factor is physician beliefs about treatment: 36 percent of end-of-life spending, and 17 percent of U.S. health care spending, are associated with physician beliefs unsupported by clinical evidence.

Contents

List of Tables

List of Figures

Acknowledgments

Several wonderful people contributed to the completion of this dissertation. They rock.

First and foremost, I would like to thank my advisors Amitabh Chandra, David Cutler, Scott Stern, and Dan Carpenter for their unparalleled guidance and generosity during my doctoral studies. I could not be more grateful for their mentorship. The best part of my doctoral studies has been learning from and working with these warm, funny, kind, and thoughtful scholars.

Research in this dissertation has also benefited from conversations with Pierre Azoulay, Iain Cockburn, Jeff Furman, Larry Katz, David Studdert, and Heidi Williams. I am particularly grateful to a group of undergraduate professors who first encouraged me to pursue research and who remain mentors, friends, and role models. Amitabh Chandra, Jon Skinner, Doug Staiger, and especially Bruce Sacerdote taught me early on that a research career involves getting to do exciting work with great people (and in some cases, also involves spending a lot of time on sailboats).

I can't imagine the past five years without my colleagues and friends in Cambridge. I have learned so much from the NBER Pre-Doctoral Fellows in Aging and Health Economics, the NBER Productivity Lunch, and other academic classmates. A few folks were extra generous with their time: Steve Cicala, Johanna Möllerström, Sara Nadel, Tisamarie Sherry, Elizabeth Walker, Francisco Queiro and Clara Zverina, all provided feedback on my research. Innessa Colaiacovo is a wonderful friend and editor.

I am grateful for financial support from Harvard University, the National Institute on Aging (through Grant Number T32-AG000186 to the National Bureau of Economic Research), the NSF-IGERT Multidisciplinary Program in Inequality and Social Policy, and The Taubman Center for State and Local Government.

Last, but not least, thank you to my family. My mother Hannah was one of the toughest people I have ever known and her independent spirit remains an inspiration. My father, Paul has been a constant source of support and good questions – and he still has one more doctoral title than I do. My great uncle Hans, the first professor in the family, alternately

nagged and encouraged me in my research. My grandmother Eva's thirst for knowledge has always amazed me, and my brother, Ian is the most loving and supportive sibling I could imagine. Finally, my partner in life and crime, Sebastian, you are the best; thank you for everything. All remaining errors in this thesis are my own.

To my mentors

Chapter 1

Innovation under Regulatory Uncertainty: Evidence from Medical Technology

1.1 Introduction

When does regulation help or hinder pioneer innovators? On the one hand, first mover advantages in commercializing new technologies arise when firms can capture substantial market share, for example through exclusive patenting. On the other hand, early innovators may pay large fixed costs in order to establish regulatory precedents and in doing so, allow subsequent entrants to free ride. Thus, the effect of novelty on pioneer innovators is ambiguous.

Industry regulation, in turn, is often associated with delayed or reduced firm entry; all else equal, extended time between a new invention and its commercialization will reduce incentives to innovate. For example Roin et. al. (2013) find evidence of this phenomenon in cancer research and development (R&D). Reductions in firms' innovation incentives will, in turn, have a downstream effect on their strategies for entering new markets. This paper explores one determinant of these market entry choices by considering the costs of being a

first mover innovator in the context of new medical product regulation in the United States.

In the United States, all medical technologies are regulated by a single agency, the U.S. Food and Drug Administration (FDA). The FDA regulates two trillion dollars worth of products every year, including 80 percent of the U.S. food supply, cosmetics, animal products, and, importantly for this study, all ethical drugs and medical devices (Babiarz and Pisano, 2008). The FDA also regulates several emerging classes of medical products such as biologic drugs ("biologics"), nanomedicines, tissue engineered products, and the use and applications of cellular and gene therapies.

Previous studies of medical innovation under FDA regulation have focused almost entirely on the pharmaceutical drug industry (Goldman and Lakdawalla, 2012), where early mover regulatory advantages have been documented. For example, Carpenter et. al. (2010) find a small but statistically significant relationship between entry order into a drug market and approval times for new drugs: going from being first to second to enter a given market is associated with a regulatory approval process that is just over a week longer (approximately a 1.2 percent increase in the length of the approval process). Relatedly, Dranove and Meltzer (1994) show that more important chemical drugs are developed and approved more rapidly. However, newer classes of medical technology – in particular, medical devices – are characterized by a larger degree of product heterogeneity and significant regulatory uncertainty, changing the context of new product regulation.

I begin by comparing the dynamics of the well-established regulatory approval process for new chemical drugs to the less studied and more uncertain regulatory approval process for new medical devices, a category including products as wide-ranging as pacemakers, coronary stents, and silicone breast implants. I find that, in contrast to the early entrant advantages observed in drug regulation, first entrants in medical device markets experience a strong disadvantage in the regulatory approval process. Using data spanning three decades of regulatory approvals (1977-2007), I show that pioneer entrants in new device product categories spend 34 percent (7.2 months) longer in the approval process than the first follow-on innovator in that category. This represents 16 to 21 percent of the total period

of de facto market exclusivity a pioneer device innovator can expect to experience. Given the concentration of earnings in the earliest years a device is on the market, back-of-the-envelope calculations suggest that a delay of this length could mean a loss of approximately 8 percent of expected lifetime product revenues.

I then ask how different types of regulatory uncertainty are related to approval times in the medical device setting. I first consider *technological* uncertainty – uncertainty on the part of the regulator that involves a lack of technological or scientific understanding of a specific type of product which is used for a given function in the human body. Technological uncertainty arises most frequently in the evaluation of very novel medical devices, where the regulator needs to understand the scientific mechanisms through which a device works in the human body. Consider for example the first time that the FDA was asked to evaluate an implantable cardioverter defibrillator (ICD[1]) for approval. The first ICD was approved by the FDA in 1984 and at that time, the technological uncertainty faced by regulators was centered around understanding precisely how the device interacts with the heart and the surrounding tissues with which it is in contact.

Research and development on ICDs continued over subsequent years and to date, over two dozen later-generation ICDs have been approved by the FDA. Some of these ICDs were classified under the same product code as the originally approved device, but starting in 1997, some approved ICDs were given a new product code due to modifications in the design of the device (for example, one group of ICDs that has emerged since 1997 involves two electrodes inserted into the heart, rather than just one). While these later products were somewhat different than earlier models, the FDA had already established a good understanding of how ICDs function as well as an understanding of how to assess the technology involved in these devices by the time that later-generation ICDs began applying for regulatory approval.

[1]An ICD is a small device that is surgically placed in the chest or abdomen, which is used to treat irregular heartbeats called arrhythmias. An ICD uses electrical pulses to help control life-threatening arrhythmias – in particular, those that can cause sudden cardiac arrest and subsequent death (http://www.nhlbi.nih.gov/health/health-topics/topics/icd)

Exploiting the fact that some products with the same technical function are given a new nominal classification as a result of design changes, I ask how much of the longer regulatory approval times for first entrants can be explained by technological novelty vs. (nominal) categorical novelty. I find that once I control for the designation of being in a "new product code," knowing whether or not a device was technologically novel does not provide any additional explanatory power in understanding regulatory approval times. This suggests that the regulator's familiarity or lack of familiarity with the primary technology used in a new medical device is not the primary determinant of the length of the regulatory approval process. For example, the first ICDs in later-established ICD product codes still experienced a regulatory delay associated with being the "first entrant," despite the fact that the regulator was already familiar with the technology used in these devices.

If technological novelty is not the primary driver of longer regulatory approval times for first mover innovators, than what else might be at play? The results suggest that there is something particular about the administrative designation of being in a new product code that is of importance – that for some reason the categorical change associated with a new product code *itself* is predictive of longer regulatory approval times. With this in mind, I next consider the role of a different type of uncertainty: uncertainty about content and format of a new product application.

Content and format uncertainty occurs in the absence of clear guidelines for the protocol for evaluating a new product, leading to uncertainty on the part of the regulator as to how to assess the results of clinical studies and other (e.g. biocompatibility and engineering) tests. This type of uncertainty almost certainly co-occurs with technological uncertainty for new products, and without the establishment of clear evaluation standards, it will persist long into a product's development lifecycle. Content and format uncertainty is easiest to think of in a scenario in which a product and its functionality are known to the regulator, but evaluation criteria are not formally articulated or established. This can be seen in the case of drug eluting stents[2] (DESs), which were first submitted to the FDA for approval in 2002.

[2]Catheter-based procedures are frequently used to treat blockages in the arteries of the heart (coronary

It was not until 2008, however – after five different DESs had submitted applications for regulatory approval and four had already been cleared – that the FDA published a formal guidance document, detailing what criteria it would use to evaluate DESs moving forward.

I consider the release of FDA guidance on DESs and eight other unique medical devices. In each case, objective regulatory guidance was introduced for a group of already-established products (i.e. multiple approvals had already occurred). I find that on average, approval times for subsequent entrants fall by approximately 40 percent (6.1 months) after application content and evaluation procedures are made explicit through formal guidance. In contrast to technological uncertainty, uncertainty about content and format of new product applications appears to play a large role in explaining regulatory approval times for first movers, and overall.

This finding has implications for other emerging categories of medical technology including biologics, tissue engineered products, and the applications of cellular and gene therapies – all settings in which there is a large degree of uncertainty about the content and format of new product applications and as a corollary, around how to evaluate new products. This is the result of both a short regulatory history and dearth of established regulatory criteria. For these new product categories, regulatory approval times are similarly likely to be substantially protracted (relative to what is administratively required) until a time when objective product evaluation criteria are formalized and made available to innovators.

After showing the impact of uncertainty on review times, I consider how the implicit costs of the regulatory approval process affect firms' strategies for entry into new medical device product categories. I consider firm behavior under regulatory uncertainty, given likely additional costs of gaining regulatory approval in new product codes. I evaluate the behavior of all cardiovascular device firms in the data and find that financially constrained firms are less likely to enter new device markets as pioneers: the fraction of financially

---

arteries). Often a stent is used to prevent restentosis (renarrowing) of the diseased artery. Stents are small metal tubes that are inserted and expanded into the artery wall and used to keep the previously narrowed artery segment open. Drug eluting stents (DESs) are medication-coated stents that reduce the chance of renarrowing of the blood vessel (Maisel and Lasky, 2007)

constrained firms among pioneer entrants into device markets is between 25 and 52 percent lower[3] than among follow-on entrants.

The rest of the paper proceeds as follows: the next section describes the markets for drugs and medical devices and the institutions that regulate their entry. Section 3 lays out a model of regulatory delay and subsequent firm choice given large anticipated costs for pioneer innovators. Section 4 describes the data on new drug and device approvals used in the empirical analyses in Sections 5 and 6. Section 7 concludes.

1.2 Background: Markets and Regulatory Frameworks

1.2.1 Medical Products: Definitions and Markets

This paper considers two large categories of medical products: chemical drugs and medical devices. Chemical drugs are defined by the FDA[4] as "articles intended for use in the diagnosis, cure, mitigation, treatment, or prevention of disease" and "articles (other than food) intended to affect the structure or any function of the body." Examples of drugs include familiar ingestible or injectable products such as antibiotics and oral contraceptives. A medical device is defined by the FDA[5] as an "instrument, apparatus, implement, machine, contrivance, implant, in vitro reagent, or other similar or related article...intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease" and "which is not dependent upon being metabolized for the achievement of any of its principal intended purposes." Examples of medical devices range from stethoscopes to breast implants, to prosthetic limbs and pacemakers.

U.S. drug and device markets are large: at an annual $320b and $140b respectively, these markets make up a meaningful share of the $2.7 trillion that is spent annually on

[3] depending on the definition used; see Section 1.6 and Table 1.10 for detailed descriptions.

[4] FD&C Act, sec. 201(g)(1)

[5] http://www.fda.gov/medicaldevices/deviceregulationandguidance/overview/classifyyourdevice

health care in the United States[6]. Drug spending is greater – however, devices and other emerging medical technologies make up a growing share of national health expenditures: while spending on prescription drugs grew at an annual rate of approximately 3.3% over the five years ending in 2011, spending on medical devices grew at a rate of 6.0% (versus 4.5% overall health expenditure growth over the same period). In addition to representing large medical product markets in the United States, drugs and devices offer substantial research opportunities: detailed data are available across product classes and over the entire history of the FDA's regulation of these products.

Other emerging categories of medical technology also comprise an increasing share of health spending. One prominent example is that of biologics, a group of large, complex and heterogeneous proteins derived from living organisms, which are often the primary component of vaccines and cancer therapies. Because they are more complex and derived from living cells, biologic drugs are regulated separately from chemical drugs. Although biologics do not appear in the analysis below, they resemble devices in their heterogeneity and shorter regulatory history and are poised to increase in both economic importance and regulatory submissions over the coming years. In 2010, seven of the top 20 drugs in the US were biologics (Lancet, 2012).[7]

Drugs are a relatively homogeneous category of products with a century-long history of regulation. By comparison, medical devices and other non-drug medical products have a shorter regulatory history and are far more heterogeneous. As such, for devices and other newer categories of medical technology, it is more difficult to define detailed regulatory standards for new products *ex ante*. Given the greater degree of regulatory uncertainty for innovators in the medical device industry, I explore what types of incentives have been

[6]Source: *National Health Expenditures*, 2012

[7]Another example of an emerging medical technology is that of nanomedicine – a term used to define the application of nanotechnology in medicine. Nanomedicine involves the use of particles in the size range of 100 nanometres (nm) or less and includes liposomes, polymer conjugates, protein/antibody conjugates, block polymer micelles, cross-linked (nano)gels, bioactive synthetic polymers/vesicles, nanoparticles and nano-sized drug crystals. Nanomedicines are mainly anticancer, anti-infective or immunomodulator drugs. The global nanomedicines market was valued at $72.8 billion in 2011 and is expected to reach $130.9 billion in 2016 (Generics and Biosimilars Initiative, 2013).

created by the regulatory system in place.

1.2.2 Medical Product Regulation in the United States: The FDA

In the United States, all medical technologies are regulated by a single agency, the U.S. Food and Drug Administration (FDA). The FDA is an agency of the Department of Health and Human Services and is responsible for the oversight of two trillion dollars worth of products every year, including all over-the-counter and prescription drugs and medical devices (Babiarz and Pisano, 2008; Hamburg and Sharfstien, 2009). The FDA also regulates all other new and emerging classes of medical products. The precursor to the modern FDA was established through the Pure Food and Drug Act, which was signed by President Theodore Roosevelt in 1906. It was not until seven decades later, however, that the FDA's regulatory scope grew to include medical devices, which came under FDA regulation in 1976.

The FDA is organized into centers, each of which is tasked with the oversight of a different type of product. The two centers most relevant to the analysis below are the Center for Drug Evaluation and Research (CDER) and the Center for Devices and Radiological Health (CDRH), which regulate chemical drugs and medical devices, respectively[8]. Within the CDER, the Office of Drug Evaluation is responsible for the approval of new drugs and within the CDRH, the Office of Device Evaluation is responsible for the review and approval of medical devices. Other categories of products are also reviewed by specialty centers within the FDA (e.g. biologics and human cells, tissues, and cellular- and tissue-based products are reviewed by the FDA's Center for Biologics Evaluation and Research, CBER).

1.2.3 The FDA and the Regulation of Drugs

The foundation of the FDA's modern statutory authority to regulate medical products is the Federal Food, Drug, and Cosmetic Act of 1938 (FDCA), which requires that new drugs be tested for safety and that those tests be submitted to the government for marking

[8]The CDRH also regulates radiation-emitting products such as X-ray and ultrasound machines

approval (Babiarz and Pisano, 2008; FDA, 2013). The FDCA "endowed the FDA with sole authority to reject the ex ante marketability of any new pharmaceutical product" (Carpenter, 2010) and resulted in the establishment of the new drug application process (NDA), the "vehicle through which drug sponsors formally propose that the FDA approve a new pharmaceutical for sale and marketing in the U.S."[9] The goals of the NDA are to provide sufficient information on drug safety and effectiveness for proposed uses, to determine whether the contents of proposed labeling are appropriate, and to evaluate whether manufacturing methods used are adequate.

The NDA is organized into technical sections,[10] which are evaluated by specialized review teams of experts (Monahan and Babiarz, 2008). The components of the NDA are specific and well-defined for all types of drugs. For example, for the information required about the drug's manufacturing scheme, the applicant firm must describe the synthesis of the active ingredient, including details on all starting materials, solvents, reagents, intermediate substances and their compilations and analytical controls (Monahan and Babiarz, 2008). The results of randomized, typically placebo-controlled clinical trials[11] are also an important component of any NDA. During the FDA's in-depth review of the NDA, the sponsor may also be required to submit additional information supporting the drug application (Babiarz and Pisano, 2008). The average approval time for a new drug in this study is 23.5 months, although the average for a drug that is first in its disease group is shorter, at 19.3 months. Figures 1.2 and 1.2 provide additional information on the chronology and requirements of

[9] http://www.fda.gov/Drugs/DevelopmentApprovalProcess/HowDrugsareDevelopedandApproved/ApprovalApplications/NewDrugApplicationNDA

[10] Requirements are outlined in FDCA and Title 21 of the US Code of Federal Regulations part 314

[11] Typically three phases of clinical trials are required in order for the FDA to be assured of a drug's safety and effectiveness (although sometimes approval decisions are made early based on demonstrated need for a drug and very promising results in phase II trials). Phase I trials are typically very small (N=20 to 80) and are primarily for determining drug safety and establishing side effects. Assuming that Phase I trials don't reveal unacceptable levels of harm, Phase II trials are conducted in a greater number of healthy subjects (as many as a few hundred, with the exception of drugs for diseases like cancer) and the focus is on establishing a product's effectiveness. Phase III trials begin following evidence of effectiveness in Phase II and are usually very large studies (N= hundreds to 3000). Phase III studies are designed to have sufficient statistical power to confirm a product's safety and effectiveness in different populations and different dosages (FDA, 2012; http://www.fda.gov/drugs/resourcesforyou/consumers/ucm143534.htm)

the NDA process.

1.2.4 The FDA and the Regulation of Medical Devices

The FDCA of 1938 did not impose any pre-approval requirements on medical devices, which instead were regulated at the state level at the discretion of each state's legislature for nearly four subsequent decades. It wasn't until 1976, after a series of well-publicized medical device failures, that Congress passed the Medical Device Amendments Act (MDA), which gave the FDA primary authority to regulate devices sold in the United States (Sall, 2008; Kramer et. al., 2012; Munsey 1995).

Devices are diverse in their cost, invasiveness, function, and risk: they include products ranging from tongue depressors and stethoscopes (which the FDA classifies as "low-risk" devices) to hearing aids ("moderate-risk" devices) to pacemakers and prosthetic heart valves ("high-risk" devices). The MDA delineates these three risk groups and lays out the rules for regulating each differently. This paper focuses only on approval regulation of "high-risk" (Class III) devices which "support or sustain human life" and are of the highest risk (FDA, 2002).[12] Unlike moderate and low risk devices, high-risk devices are are subject to a rigorous regulatory process that is similar to that imposed on new drugs (Zuckerman et. al. 2011; Goldman and Lakdawalla, 2012), requiring detailed product information and evidence of safety and effectiveness from clinical trials. While high risk devices represent only about one percent of the devices that the FDA regulates each year (Redberg and Dhruva, 2011), they represent an out-sized fraction of medical device spending: In 2008, spending on the six highest-cost implanted devices alone was about $13 billion (Meier, 2009), or approximately 10 percent of total U.S. medical device spending.

The regulatory approval process for high-risk devices is called "premarket approval" (PMA) and is necessary when a medical device developer wants to market a new high-risk device. Importantly, once the first device in a product code is approved through the PMA

[12] http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/PremarketSubmissions/PremarketApprovalPMA/ucm2007514.htm

process, all subsequent devices in that product code go through the PMA process.[13] The average approval time for a new device is 18.1 months, although the average for a device that is first within a product code is longer, at 22.5 months. Figures 1.1 and 1.2 provide additional information on the chronology and requirements of the PMA process and Figure 1.2 highlights similarities and differences between the requirements for the NDA and PMA. Much like the NDA, the PMA is a complex document filed by the manufacturer that contains information about the product and results of clinical trials. As is the case for drugs, Section 515 of the FDCA requires that a PMA provide scientific evidence of safety and effectiveness, typically in the form of data from a pivotal study.[14] However, as the next section explains, the types of trials that can constitute a pivotal study for a new high-risk medical device are highly heterogeneous and to a large extent, open to interpretation – an important difference between the regulatory approval processes for drugs vs. devices.

1.2.5 Drugs vs. Medical Devices: Regulatory Differences

Importantly – and unlike drug trials – clinical trials for medical devices may take many different and often more flexible forms. In new drug studies, three phases of randomized controlled trials are the norm. In device trials, however, clinical evidence can come from a variety of sources: trials may take the form of well controlled investigations, partially controlled investigations, objective trials without matched controls, and other types of studies "from which it can fairly and responsibly be concluded by qualified experts that there is reasonable assurance of safety and effectiveness of a device under its conditions of

[13]The 510(k) process, which allows devices to be cleared for marketing on the basis of being "sufficiently similar" to other already-cleared devices was originally intended for use with medium-risk devices only. In recent years, some high-risk devices have also managed to gain clearance through this process, but these are not devices that have a history or precedent of PMA approval within the product code. The 510(k) process has been criticized for being used too freely and the Institute of Medicine has convened a committee to look at its use (Garber, 2010). This is certainly an important area for further research, however this paper focuses only on those device product codes that are explicitly designated for the PMA-track approval process.

[14]The clinical study report includes the study design and protocol, patient enrollment and exclusion data, primary and secondary endpoints of the study, data from all patients entered into the trial, and detailed statistical analysis of the results. Technical data on biocompatibility, stress and fatigue, shelf life, and other relevant non-clinical tests are also submitted (Zenios et. al., 2010)

Figure 1.1: *Regulatory Approval Processes: Drugs & Devices*

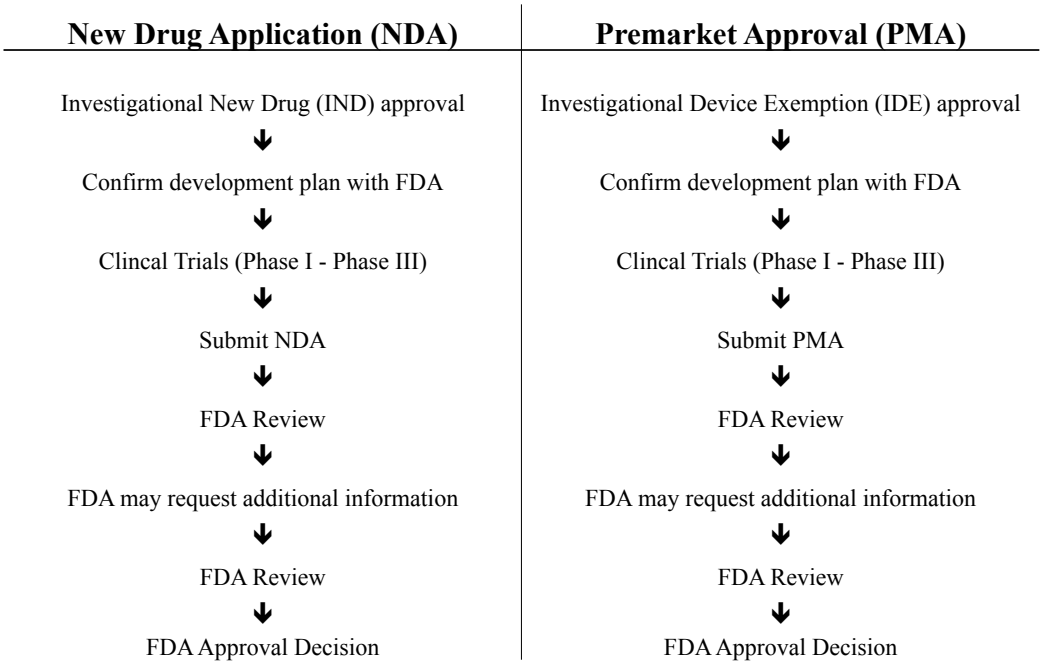
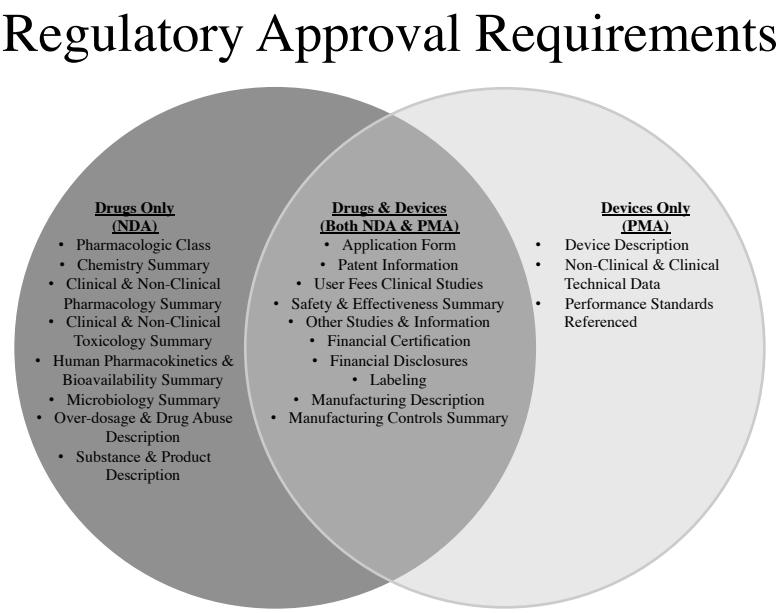


Figure 1.2: *Comparing Regulatory Approval Requirements: Drugs & Devices*



use" (21 CFR 860.7(c)(2)).

This lack of specificity about the type and execution of clinical trials is largely the result of product and delivery-method heterogeneity across medical devices. Given these sources of heterogeneity, regulators have been unable to articulate general rules or guidelines for medical device clinical trials and subsequent regulatory evaluation that are both sufficiently broad so as to be relevant to devices ranging from pacemakers to silicone breast implants, while still being sufficiently specific to guide the clinical trials and the regulatory evaluation of all types of devices.

While drugs are almost always delivered in one of just a few conventional ways (administered orally, injected intravenously or intramuscularly, inhaled, or administered topically), the insertion and delivery method of a new high-risk medical device is often a novel process with few (if any) related prior clinical trials to use as a precedent or guide. Thus, both the planning and execution of device trials are substantially more heterogenous than those for new chemical drugs. Devices can be used, implanted, or otherwise administered in hundreds of ways. Furthermore, how a device is used or the method by which an implantable device is put into the human body is often not only unique, but also critical to the success or failure of a trial (Sall, 2008).

The large degree of heterogeneity across medical devices and in the processes required for their evaluation combined with non-specific regulatory language about how clinical trials should proceed results in a much greater degree of uncertainty around content and format requirements for device regulation (vs. drug regulation) because the regulator's expectations are typically not clearly known or defined *ex ante*. Chatterji (2009) relays the anecdote of one extreme case of regulatory uncertainty: the company Acorn Cardiovascular "believed they were close to FDA approval in 2002 for their device that helps to shrink enlarged hearts, but the FDA instead recommended a much larger clinical trial that ended up taking three more years and costing the company $30 million."

While this represents an extreme example of delay due to regulatory uncertainty, it is also true that in general, FDA decisions are rarely made immediately after a PMA submission.

Indeed there are typically at least two cycles of requests and responses between the FDA and the applicant firm before a decision is made (Zenios et. al., 2010). This is because for most devices, the evaluation criteria that the FDA will use to assess a new product are not made explicit before the regulatory process begins. An important exception to this are cases where the FDA publishes regulatory guidance, a list of objective product evaluation criteria addressing application content and format that will be used to assess all devices of a certain type moving forward. The publication of such guidance is considered in detail in Section 5.3 of this paper.

Appendix A.1 presents additional case studies of firms' experiences with regulatory uncertainty and delay. Case 1 in Appendix A.1 presents the story of a heart failure monitoring system that has been under consideration at the FDA for three years. At the time of writing, the device has already been through one large-scale controlled clinical trial and one follow-up study, but the FDA has yet to come to an approval decision. Case 2 presents a typical occurrence for a new high-risk device: following the completion of a randomized, double-blind, sham-controlled pivotal clinical trial that yielded statistically significant results supporting the device's safety and efficacy, the FDA returned to the manufacturer with follow-up questions related to device testing and clinical data.

In sum, although device manufacturers need to present clinical trial evidence to the FDA, the lack of regulatory specificity about what types of data to collect and present to regulators as well as the content and format of a new product application makes the regulatory process for devices far more uncertain than that of drugs. In the sections that follow, I will explore how this uncertainty plays out in product approval times and firms' strategies for entering new markets.

1.2.6 A Note on Safety vs. Speed of Regulation

A long debate has engaged with the tradeoffs between regulatory speed and consumer safety. In a 2009 piece in the New England Journal of Medicine, FDA Commissioner Margaret Hamburg and Principal Deputy Commissioner Joshua Sharfstein discuss the balance that

the FDA must strike between risks to consumers and speed of regulation: "as a public health agency, the FDA should always ask whether delays in approval or safety problems can be prevented" (Hamburg and Sharfstien, 2009).

This paper does not assess or weigh in on the balance between regulatory speed and consumer safety in current policies. Rather, I consider factors that may affect delays in new product regulation on the intensive margin – that is, given the regulatory system as experienced by medical product innovators in the United States – and as such, the length of development times experienced by firms. My conclusions concern only the context of the regulatory system in place, given a regulatory agency that aims to protect both consumer safety and its own reputation. These dual goals are reflected in the model discussed in the next section and described in detail in Appendix A.2.

1.3 A Model of Approval Regulation and Firm Strategy

In many industries, government approval or licensing is a prerequisite for market entry. Examples include nearly all parts of the energy, health care and transportation industries. This paper considers the experiences of medical technology firms in their interactions with the FDA.

1.3.1 Framework and Regulator Decision-Making

The first part of my empirical work builds on Carpenter et. al.'s (2010), model[15] of the FDA drug approval process. In this model, a farsighted regulator discounts the future pipeline of device approvals and decides how rapidly to approve a new device in light of this. In such a setting, the regulator gets greater utility from quickly approving an earlier entrant into a given market than a later entrant. Appendix B presents details of this model of approval by a farsighted regulator.

[15]This framework is also related to Carpetner (2004). In this model, "early entrants" into an exclusive market niche (disease) receive shorter expected approval times than later entrants, even when later entrants offer known quality improvements over earlier products.

In the model, the regulator can also respond to political factors, which is consistent with existing evidence on the political economy of the FDA's regulatory behavior. For example, studies show that the FDA responds to the demands of lobby groups representing (potential) drug consumers, such as cancer or AIDS organizations (Olson, 1995; Carpenter 2002; Carpenter et. al., 2010).

Individual firms may also exert pressure on the FDA[16], although recent work on pharmaceutical drug approvals has found limited evidence of their influence on regulatory approval times (Carpenter et. al., 2010). In the model and analyses that follow, I account for firm and disease-specific factors that may influence the duration of the regulatory approval process without focusing on their relative importance (for example, Acemoglu and Linn (2004) find that potential market size has a strong influence on the entry of non-generic drugs and new molecular entities while Carpenter (2004) finds that firms submitting more new product applications may expect quicker and more likely approvals). In doing so, I deviate from Carpenter et. al. (2010) in defining a more general model of approval priorities for an uncertain regulator.

I begin with a simple, flexible model of regulatory approval times that includes known covariates, such as those factors identified above. Both firms and the regulator observe the relationship between regulatory approval times and application characteristics. Approval time (T) of product p, of entry order ϕ produced by firm f, in year t is observed as:

$$T_{p_\phi ft} = f(\beta \mathbf{X}) \tag{1.1}$$

where **X**s include:

- Entry order within a product code (devices) or disease group (drugs)
- Advisory panel (organized by medical specialty), product group, and firm fixed effects
- Year of review
- Applicant firm's cumulative regulatory experience

[16]Other work – e.g. Thomas (1990) has found that FDA regulations have heterogeneous effects on firm productivity by firm size.

- Eligibility for expedited review (e.g. product is for a rare/orphan/terminal disease)

Because the regulator discounts the future pipeline of products, it would prefer to approve earlier products more quickly (see Appendix B). Thus, review times should be increasing in entry order *ceteris paribus*. In other words, all else equal, earlier entrants should benefit from a shorter regulatory process (and later products should experience increasingly longer approval times) leading to early mover advantages in the approval process.

However, when there is regulatory uncertainty about how to evaluate a product, it will increase the time that a regulator spends on the approval decision. Further, because that uncertainty is likely to be inversely related to entry order (i.e. uncertainty is greater among the first products to seek regulatory approval), the presence of regulatory uncertainty could affect approval times in the *opposite* direction of the early mover regulatory advantages described above. Indeed, if regulatory uncertainty is great enough, it could lead to *longer* regulatory approval times for earlier entrants, even given the regulator's preference for getting more novel products to market quickly.

To account for entry-order specific uncertainty, I modify Equation 1. I relate review times to the set of determinants above as well as an uncertainty term:

$$T_{p_\phi ft} = f(\beta \mathbf{X}) + U_{p_\phi} \tag{1.2}$$

where

$$U_{p_\phi} = \begin{cases} D + \varepsilon & \text{if } \phi < \phi^* \\ 0, & \text{otherwise} \end{cases} \tag{1.3}$$

For simplicity, regulatory uncertainty, U_{p_ϕ}, can be thought of as generating a fixed delay during the regulatory approval process, on average D, although a more general framework would model $U_{p_\phi} = g(\phi) + \epsilon$ where $g'(\phi) < 0$. That is, among some set of the earliest entrants for whom the regulator is uncertain as to how to regulate the new product in question, approval times are D longer, on average. When D is large, expected approval times will increase. Thus, even when the regulator *prefers* faster approval for earlier entrants, a large value of D implies that approval times for the earliest entrants could be longer than

those of subsequent entrants.

In the empirical section of this paper, I ask when there is evidence that $D > 0$ and for which values of ϕ this is the case. By knowing the values of ϕ (entry order), for which there are additional costs of regulatory approval, one can evaluate which set of entrants are disadvantaged (in the form of extended approval times) in the approval process. In Section 5, I first focus on estimating the additional regulatory approval times associated with early entrants – i.e. the cost of pioneer entry that is directly observable in this data. However, there are other additional costs likely to accrue to early innovators such as additional legal fees and shortened periods of market exclusivity; these are discussed later in the paper.

In addition, the empirical section of the paper addresses the fact that U_{p_ϕ} likely has several components. I am unable to identify all of them, but note that a factor that increases U_{p_ϕ} should also lead to longer approval times. I consider two such factors – technological uncertainty and uncertainty about application content and format – which I am able to analyze separately by taking advantage of two unique sources of variation in the regulatory approval data. I test the model above and the role of different types of regulatory uncertainty in Section 5.

1.3.2 Firm Strategy

Finally, I present a testable hypothesis about firm strategy that emerges from the model described above. Both firms and the regulator observe to-date regulatory approvals, approval times, and the entry order of all prior products. Firms know that greater uncertainty increases time spent on regulatory approval and decide which markets to enter, given anticipated costs and benefits. The first dimension on which a firm makes a decision is whether to enter a novel or existing market. All else equal, this decision will be influenced by the relative cost of novel vs. established product regulation.

Assume that each firm, F, has capital K_F. Firms expect an uncertainty-driven delay of length D (as above) for innovating in a new market. For a firm, the implied cost of being a first mover is an increasing function of the length of the anticipated regulatory delay

and a decreasing function of firm capital (as financially-constrained firms will have less capital allocated for R&D and/or higher costs of borrowing) such that $C_F = c(D, K_F)$. Now consider two firms: Firm A has a large stock of available capital (e.g. Firm A as a large, publicly listed company with a large R&D budget), while Firm B is financially constrained (e.g. Firm B is small and has a finite amount of venture capital to deploy and faces high costs of borrowing or additional fundraising) such that $K_A > K_B$. Then in a given product area, the relative cost of innovating in a new market is greater for Firm B than for Firm A (i.e., $C_B > C_A$) because the expected value of D is the same for both firms.

Assume a distribution of the value of pioneer entry into new markets, such that there is a range of potential profits, π, that can be captured by the first entrant. Then each firm decides whether the expected marginal value of being the first entrant is greater than the marginal cost of being the first mover: $\pi > c(D, K_F)$. Since relative costs are greater for Firm B than for Firm A, Firm B will be willing to enter fewer new markets than Firm A. More generally, financially-constrained firms should be less inclined to enter new markets as pioneers when there are large delays associated with doing so.

Market Entry Hypothesis: In the presence of delays under regulatory uncertainty, financially constrained firms should be less likely to act as pioneer entrants

1.4 Data

The first two sources of data I use are FDA databases: the New Drug Approval (NDA) database and the Premarket Approval (PMA) database. Later in my analyses, I also use information from a detailed firm-level dataset, which was collected by hand from financial databases and firm websites and includes financial, ownership, and acquisition data for all cardiovascular device firms represented in the PMA data.

The FDA's NDA database includes a comprehensive list of all new drug approvals in the FDA's regulatory history.[17] For comparability in my empirical analyses and in order

[17] I am grateful to Daniel Carpenter for sharing the cleaned data from Carpenter, et. al. (2010) for this project. The raw data are available at http://www.fda.gov/Drugs/InformationOnDrugs/ucm135821.htm

to focus on contemporaneous regulatory periods for both drugs and devices, I limit the years of drug application data used to only those applications that were submitted after 1976 (when the FDA first began regulating medical devices) and through 2007. While later data are available, I truncate the approval data to avoid any bias that would be created by using a sample in which only the fastest approvals in more recent years would be observed.

I consider a final sample of 693 unique drug approvals that are indicated for 187 disease groups. "Disease groups" are specific product categories based on the function and target of a drug that are likely to be very good to excellent clinical substitutes for one another – for example, anti-inflammatory agents, contraceptives, or statins. The data also include detailed information about the date of NDA submission, date of FDA decision, the submitting firm's identity, and an indicator for whether a product received "priority" or expedited review (e.g. a drug could be eligible for expedited review because it is used for a rare or late-stage/terminal disease). I observe approval times as elapsed days or months from the date of the NDA submission to FDA decision.[18] Summary statistics are presented in Table 1.1.

The data on high-risk device approvals come from the FDA's PMA database,[19] which includes an exhaustive record of all PMA approvals since the 1976 Medical Device Amendments to the Federal Food, Drug and Cosmetic Act. As with the NDA data, I include all submissions starting in calendar year 1977 and truncate the data to include submissions through the year 2007.[20] The medical device approval data summarized in Table 1.1 include 847 unique device approvals in 249 product codes. Product codes are specific definitions

[18]One reader noted that it may be harder to recruit patients for clinical trials for non-first-in-class drugs, and that this could make the clinical trials last longer and extend commercialization lags for non-pioneers. While this may be true, it would represent an effect above and beyond what I observe in the FDA's data on approval times, which measure time between submission and an approval decision and not pre-NDA-submission phenomena such as the duration of clinical trials.

[19]The raw data are available at http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpma/pma.cfm

[20]For example: a device application that was submitted in 2007 and approved in 2010 would be included in the dataset. A device application submitted in 2011 and approved in 2012 would not because its submission occurred after the end of calendar year 2007.

Table 1.1: *Summary Statistics*

New Drug Applications (NDAs) - Drugs: N=693 Premarket Applications (PMAs) - Devices: N=847 Premarket Applications (Cardiovascular Devices): N=241						
	Drugs		Devices		CV Devices	
Variable	Mean	S.D.	Mean	S.D.	Mean	S.D.
Approval Time (Months)	23.54	17.67	18.12	15.84	17.31	12.96
Approval Time (1st Product)	19.31	14.40	21.48	16.77	23.07	18.16
Entry Order	13.64	17.81	6.37	8.79	5.06	4.32
Priority Review	0.44	0.50	0.10	0.30	0.10	0.29
New Applications (Current)	7.68	7.78	15.32	20.63	27.57	25.92
Submission Year	1991	7.42	1994	8.46	1995	7.98
Firm	57 FEs	–	32 FEs	–	15 FEs	–
Disease Group / Product Code	187 FEs	–	249 FEs	–	55 FEs	–

Summary statistics for the 693 drugs and 847 medical devices used in the empirical analyses, as well as separate descriptive statistics for the subset of (241) cardiovascular devices alone. Approval Time measures months from PMA/NDA submission until FDA approval. Entry Order is based on the chronological ordering of PMA or NDA submissions. Priority Review is an indicator for whether a product was eligible for expedited FDA review. New Applications (Current) is a firm-specific, time-varying count of successful new product applications that the applicant firm has completed at the time of a given submission. Submission year is the calendar year in which an application was sent to the FDA. Firm contains a set of dummy variables for each firm in the data set or a dummy indicator for being a "small" firm – i.e. one with fewer than five new applications over the entire period of observation.

based on design and function that "delineate [a device's] technology and indication,"[21] such as drug-eluding stents or silicone breast implants. As an analog to drug disease groups, device product codes are likely to be very good to excellent clinical substitutes for one another. A list of example device product code names as well as an example of a device product code definition from the FDA can be found in Appendix C. The PMA database also includes detailed information about the date of each application's submission, date of FDA decision, the submitting firm's identity, and an indicator for whether a product ever received "priority" or expedited review.

Table 1.1 highlights several similarities and important differences between the drug and device approval data. While average approval times in the sample are longer for drugs (22.5 months) than devices (18.1 months), the average approval times for the *first* product in a given category are shorter for drugs (19.3 months) than for devices (21.5 months). Drugs tend to have more entry per product category (13.6 products on average) than devices (6.4 products on average) and drugs are also far more likely to be eligible for "priority" (expedited) review (44 percent of drugs vs. 10 percent of devices).

I focus many analyses on understanding medical device approval times and for a subset of the exercises that follow, I focus only on high-risk cardiovascular devices, which are those reviewed by the Circulatory System Devices Panel. Table 1.2 shows the distribution of medical device approvals by specialty. Cardiovascular (circulatory system) devices make up by far the largest speciality area, comprising 241 out of the 847 applications in the data, or approximately 28.5 percent of the total device sample.

Finally, for the set of firms that produce the high-risk cardiovascular devices in the PMA database, I collect detailed firm-level financial and ownership data. These include data on firm size (as measured by annual revenues), firm ownership (public vs. private), and whether and when a firm was acquired by another company – as well as the identity of that company and the year of acquisition, if relevant. Financial data were collected from Google Finance, NASDAQ, NYSE Euronext, and from firm websites.

[21] http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments

Table 1.2: *New Devices by Advisory Committee (Specialty)*

Advisory Committee	New Devices	Percent
Circulatory System	241	28.45
Opthalmic	160	18.89
Microbiology	74	8.74
General and Plastic Surgery	60	7.08
Gastroenterology-Urology	53	6.26
Orthopedic and Rehabilitation	53	6.26
Immunology	38	4.49
Obstetrics and Gynecology	33	3.90
Radiology	28	3.31
General Hospital and Personal use	23	2.72
Clinical Chemistry and Toxicology	17	2.01
Dental	15	1.77
Ear, nose and throat	13	1.53
Neurology	13	1.53
Anesthesiology	12	1.42
Physical Medicine	8	0.94
Hematology and pathology	6	0.71
(Total)	847	100.00

This table shows the distribution of all 847 new devices analyzed in this study by FDA (specialty-specific) Advisory Committee.

Throughout the paper, I observe data on new product approvals, not on innovation and other decisions prior to the regulatory approval process. This means that I do not observe those products that are abandoned before or during the Premarket Approval process, based on unpromising clinical results. As such, the approval time phenomena I observe and the effects that I calculate represent the *effects of regulation on the regulated*, and not the effect of regulation on those products that do not make it into (or through) the approval process.[22]

1.5 Empirical Estimation

I proceed with a series of estimates from the models above. I first compare drugs and devices using the model of regulatory approval times presented in Section 1.3. I then test the hypothesis about firm market entry strategies in detailed firm-level data.

1.5.1 Approval Times and Entry Order

The first part of this analysis is grounded in the literature on the determinants of FDA approval times for new drugs, notably Carpenter et. al. (2010) and others. I account for potential political and institutional factors that may affect approval times while estimating the relationship between product entry order and approval times for both drugs and devices. Carpenter et. al. (2010) define "entry order" as the order in which a drug within a given disease group submits an application for FDA approval. I extend this definition to its closest analog for medical devices: the order in which a medical device within a given product code submits an application for FDA approval.

[22]While the fraction of PMAs that are rejected following the PMA process is negligible (zero in recent years), the fraction of devices that are granted investigational device exemptions and then never apply for approval through the PMA process is likely higher. This data is not currently available to the public. I have requested it through the FDA's Division of Freedom of Information and hope that future versions of this paper will be able to shed additional light on the existence and nature of selection that may be involved in understanding which devices (and which type of *firms'* devices) are most likely to make it to the stage of regulatory approval.

Table 1.3: *Entry order and Approval Times*

Outcome = Ln Time to Approval				
	(1)	(2)	(3)	(4)
	(Log-Linear)	(Log-Linear)	(Cox Hazard)	(Cox Hazard)
	Drugs	Devices	Drugs	Devices
Entry Order	0.0200**	-0.0098*	-0.0207***	0.0265***
	(0.0074)	(0.0047)	(0.0057)	(0.0063)
Controls	X	X	X	X
N	693	847	693	847
R^2	0.3587	0.1048		
F-test	P[$(\beta_1) = (\beta_2)$] = 0.000		P[$(\beta_3) = (\beta_4)$] = 0.000	

* p<0.05, ** p<0.01, *** p<0.001

This table shows the average relationships between product entry order and approval times for drugs and devices.

Columns 1 and 2 represent the results from a (parametric) log-normal model. Columns 3 and 4 present the results form a (semi-parametric) Cox hazard model. Columns 1 and 3 consider new drug approvals and columns 2 and 4 consider new device approvals. The dependent variable in all models is the natural log of approval time from submission.

All models include firm and product type fixed effects and a time trend (year) Results presented are robust to the exclusion/inclusion of firm fixed effects and to the use of year fixed effects rather than a time trend. All models also include controls for whether a product was granted "priority" (expedited) review as well as a count of the applicant firm's total approved applications at the time of submission. Standard errors are clustered at the product level.

I begin my analysis by replicating Carpenter et. al.'s (2010) results on the set of 693 new chemical drugs described above. Columns 1 and 3 of Table 1.3 show the results of both a parametric (log-normal) model and a semi-parametric (Cox proportional hazard) model.[23] As previously observed, I find evidence of a positive, statistically significant entry order gradient in approval times for new drugs that is persistent, robust to multiple statistical specifications, and tantamount to early mover advantage in the drug regulatory approval process. On average, a one unit increase in entry order is associated with approximately a 2 percent increase in regulatory approval times for new drugs within a disease group (e.g. among statins, oral contraceptives, etc.). The results are statistically significant at conventional levels and robust to the inclusion of firm and year fixed effects or a time trend, disease group fixed or random effects, and a time-varying indicator of a firm's "expertise" in navigating the regulatory process (for which I use a firm-specific, time-varying count of successfully approved NDAs at the time of a given new application as a proxy).

I then conduct a parallel analysis for the approval times of new medical devices. In Columns 2 and 4 of Table 1.3, I repeat both the parametric (log-normal) and semi-parametric (Cox proportional hazard) analyses on the dataset of 847 new medical devices. In the medical device sample, I document a statistically significant relationship, which is oppositely signed compared to that estimated for drugs: on average, a one unit increase in entry order is associated with approximately a 1 percent *decrease* in regulatory approval times for new medical devices. That is, the later a product enters a given market, the shorter the average time to regulatory approval. These medical device approval models also present results that are statistically significant at conventional levels and robust to the inclusion of firm and year fixed effects or a time trend, product code fixed or random effects, and a time-varying indicator of a firm's "expertise" in navigating the regulatory process (for which I use a firm-specific, time-varying count of successfully approved PMAs at the time of a given new application as a proxy). F-tests comparing the drug versus device coefficients reject

[23]The log-normal model can be interpreted as the percentage change in approval time associated with a one unit increase in entry order. The Cox proportional hazard model (Cox, 1972) reports the effect of a unit increase in entry order with respect to the hazard rate of exiting the approval process.

the equivalence of the relationships between entry order and approval times for these two categories of products at the 0.001 percent level in both sets of models.

Having found evidence of early mover regulatory advantages (on average) in drug approval times and early mover regulatory disadvantages (on average) for device approval times, I turn to understanding the drivers of these patterns. If the relationship seen in device approvals is a result of early entrant regulatory disadvantage, there are three phenomena that should be observable in the data. First, if observed patterns are being driven by early entrants, one should expect to see stronger relationships in samples that include these entrants and should not expect to see the same patterns in samples that do not include early entrants. Second, it should be possible to identify those entrants for whom there are additional delays associated with entry order and third, to quantify their magnitude.

Table 1.4 tests the first implication above. Column 0 replicates the two sets of log-linear results in Table 1.3: on average, approval times are decreasing in entry order for devices and increasing in entry order for drugs. Subsequent columns of Table 1.4 then ask the question: "what is the relationship between entry order and approval times when considering only entrants *beyond the Zth* product?" While the positive entry order gradient documented in the regulatory approval of new drugs is relatively stable over the product development lifecycle of a category of drugs, this is not the case for devices: the negative entry order gradient disappears as soon as the first entrants are excluded from the sample. The device results in Table 1.4 thus suggest that delays accrue mostly to the first entrant in a device product code and that the inclusion of these early entrants drives overall averages in the data.

To explain the first entrant effects further, the first column of Table 1.5 uses dummy variable indicators for a product being first, second, third, fourth, or greater than fifth in a product code, rather than a linear indicator of entry order. Column 2 estimates the same model as Column 1, but uses months as the independent variable. Column 3 compares only the first entrant to the first unambiguous follow-on entrant (i.e. the first PMA submitted in a product code vs. the first PMA submitted *after* the first PMA had been approved) and finds that relative to the first follow-on entrant, a pioneer entrant spends approximately

Table 1.4: *Truncated Samples and Approval Times*

Outcome = Ln Time to Approval for Products of Entry Order >Z					
	(0)	(1) Z=1	(2) Z=2	(3) Z=4	(4) Z=6
		Drugs			
Product Entry Order	0.0200**	0.0195*	0.0236**	0.0295***	0.0303**
	(0.0074)	(0.0076)	(0.0072)	(0.0084)	(0.0092)
Controls	X	X	X	X	X
N	693	581	497	394	337
R^2	0.3587	0.3824	0.4365	0.4938	0.4927
		Devices			
Product Entry Order	-0.0098**	-0.0054	0.0014	-0.0050	-0.0069
	(0.0047)	(0.0062)	(0.0069)	(0.0090)	(0.0148)
Controls	X	X	X	X	X
N	847	608	479	330	234
R^2	0.1048	0.1253	0.1400	0.1478	0.1682

* p<0.05, ** p<0.01, *** p<0.001

Column 1 replicates the log-linear results in column 3 of Table 1.3. Subsequent columns answer the question: "what is the relationship between entry order and approval times when considering only entrants *beyond the Zth* product?" Columns 2 - 5 show results for an increasingly later group of entrants into a product code as one reads from left to right.

All models include firm and advisory committee fixed effects and a time trend (year). Results presented are robust to the exclusion/inclusion of firm fixed effects and to the use of year fixed effects rather than a time trend. All models also include controls for whethera product was granted "priority" (expedited) review as well as a count of the applicant firm's total approved applications at the time of submission. Standard errors are clustered at the product level.

Table 1.5: *Quantifying Early Mover Disadvantage*

Outcome = Device Approval Time (Months)				
	(1) Ln Approval Time	(2) Approval Time (Months)	(3) Ln Approval Time	(4) Approval Time (Months)
First in Product Code	0.2157**	5.7158***	0.3376***	7.1993***
	(0.0890)	(1.5015)	(0.0914)	(1.3238)
Second in Product Code	-0.0705	0.1781		
	(0.0887)	(1.3966)		
Third in Product Code	0.1208	4.7995		
	(0.1235)	(3.4273)		
Fourth in Product Code	0.0039	1.6371		
	(0.0694)	(1.7781)		
Greater than 5th in Product Code	-0.0536	0.9762		
	(0.0732)	(1.0754)		
Full Sample	X	X		
Restricted Sample (1st + 1st Follow-on Only)			X	X
N	847	847	342	342
R^2	0.0934	0.1073	0.1105	0.0986

* p<0.05, ** p<0.01, *** p<0.001

Column 1 shows the relationship between the listed entry order dummies and the log of approval time. Column 2 converts these results into months. Column 3 considers only the difference in approval times between the first applicant (the pioneer) and the first unambiguous follow-on innovator in the same product code. Column 4 converts these results into months.

All models include firm and advisory committee fixed effects and a time trend (year) and are robust to the exclusion/inclusion of year fixed effects rather than a time trend. All models also include controls for whether a product was granted "priority" (expedited) review as well as a count of the applicant firm's total approved applications at the time of submission. Standard errors are clustered at the product level.

34 percent longer in the regulatory approval process. Column 4 converts this result into months, indicating that a pioneer spends an average of 7.2 months (approximately 219 days) longer in the regulatory approval process than the first unambiguous follow-on entrant into the same product code. The results from Tables 1.4 and 1.5 also allow me to put an upper bound on the value of ϕ^*: approval delays are only statistically significant for the first entrant into a product code, suggesting that the value of ϕ^* is close to 1.

With the brunt of the costs of delay borne by the first entrant, one might wonder about the financial implications of pioneer innovation. Consider the estimated value of D, the 7.2 month longer approval times estimated for pioneer entrants: how large is this? One benchmark is the length of delay relative to the length of the period of *de facto* market exclusivity that a pioneer can expect to have. In the full medical device sample, the first entrant into a product code has an average of 3.8 years as the sole product with regulatory approval before the second product is approved for market entry – that is, the pioneer can expect an average of 3.8 years of *de facto* market exclusivity. For high-risk cardiovascular devices, this period is just 2.8 years. Thus as a ratio, the additional time a pioneer medical device can expect to spend in regulation is between 15.8 and 21.4 percent of the total period of time it can expect to spend alone on the market.

For medical devices, earnings are often concentrated in the first few years in which a product is marketed, making the role of approval times especially important in determining a device's profitability. According to the 2013 Annual Report from Medtronic (the world's largest medical device company) 38 percent of 2013 revenues were from products introduced in the last three years (Medtronic, 2013). While it is only a rough estimation of lost revenue, it is illustrative to think about what a 7.2-month regulatory delay means in this context: 7.2 months represents 20 percent of three years. If, on average, a medical device makes 38 percent of its total profits over its first three years on the market (as the Medtronic average would indicate), then a 7.2 month delay in getting to market would translate into a decrease of approximately 8 percent of lifetime revenues per new device.

A final way to think about the observed delay is in the context of the implied opportunity cost of capital. In medical product industries, the opportunity costs of capital are large. Assuming a typical discount rate used for the biotechnology industry of 11.5% (DiMasi and Grabowski, 2007), one can calculate the opportunity cost of a 7.2 month delay. Makower et. al. (2010) survey roughly 20% of firms in the medical device industry and find that the average cost of bringing a high-risk medical device to market is about $94 million. Assuming a discount rate of 11.5%, the results suggest that the opportunity cost of capital of the delay associated with being the first entrant in a product code is probably at least $6.7 million, or more generally, over 7 percent of the total cost of device development.

1.5.2 Sources of Uncertainty Part 1: Is Technological Novelty Associated with Longer Approval Times?

Given evidence of longer regulatory approval times for the earliest innovators in a medical device product code, I next explore some potential explanations. Regulatory delay has many potential components. One of the most obvious is technological uncertainty about the workings of a new product. Technological uncertainty broadly encompasses uncertainty on the part of the regulator due to a lack of scientific familiarity with or understanding of a specific type of product used for a given function.

When a product is very novel – i.e. the regulator has never seen anything that performs its function before – technological uncertainty is high. An example can be seen in the historical approvals of implantable cardioverter defibrillators (ICDs) described in Section 1.1. However, when the technological uncertainty around a certain type of device has been largely resolved – for example through multiple assessments and approvals of that type of technology – one would expect to see a decrease in that component of approval delay associated with technological uncertainty for subsequent product approvals.

I use the information embedded in FDA-defined, detailed device product names[24] to

[24]The FDA has 16 independent panels for device classification. These panels are found in 21 CFR 862-892. For each of the devices classified by the FDA the CFR gives a general description including the intended use, the class to which the device belongs (i.e., Class I, II, or III), and information about marketing requirements.

measure product "novelty" in a subsample of high-risk cardiovascular devices. I look within cardiovascular devices because this is by far the largest specialty area in the data, representing over 28% of all new device approvals and because this speciality includes the greatest number of unique product codes.

I identify eight "functional categories" of devices, each of which contains multiple unique device product codes, but all of which share a common cardiovascular function, making each category a natural setting for comparing highly related products. Examples include functional categories for stents, implantable cardioverter defibrillators (ICDs), and replacement heart valves. Each of these functional categories includes multiple products that have the same general function in the human body, but some variation in the materials from which they are made, their method of delivery, and/or the product design, resulting in administrative classifications of multiple product codes within each functional category. Figure 1.3 provides a guide to functional category construction for the subsample. The eight functional categories analyzed and the number of products and product codes in each are listed in Table 1.6.

To evaluate how technological uncertainty affects approval delays, I consider whether a prior device approval within the same functional category is associated with reduced approval times for subsequent new devices in that functional category. Devices in a functional category will, by definition, be highly similar to one another. Moreover, the prior approval of the first of a particular device (e.g. catheter) should lead to a technological understanding of that type of product among reviewers for subsequent products of that type. Thus I ask: *when a device is first in its product code, but its primary technological function and components are already known to the regulator, are regulatory times shorter?* In other words, I control for the designation of being first within a product code and then ask empirically how much additional explanatory power (if any) can be gleaned from knowing that a device was scientifically novel.

I identify the earliest entrant in each functional category of products and then look for

(http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/Overview/ClassifyYourDevice)

Figure 1.3: *Functional Category Construction*

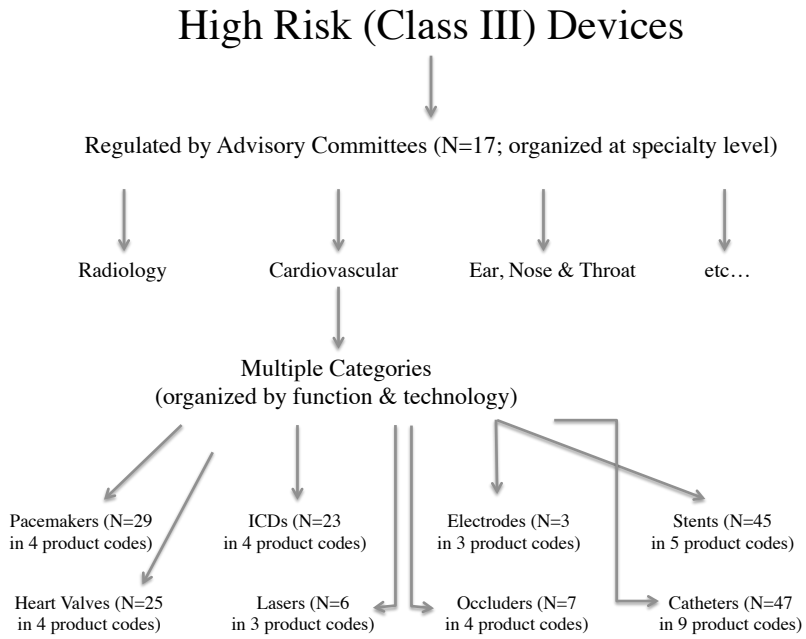


Table 1.6: *Functional Category Composition*

(Cardiovascular Devices)

Device Function (Category)	Number of Unique Product Codes	Number of Unique Devices (Total)
1. Pacemaker	4	29
2. Catheter	9	47
3. ICD	4	23
4. Electrodes	3	3
5. Stents	5	45
6. Valves	4	25
7. Laser for Angioplasty	3	6
8. Occluder	4	7

This table presents the eight functional categories evaluated in Section 1.5.4. Each of the categories contains multiple unique product codes, making each a useful setting for comparing technologically similar products.

subsequent entrants into that category. These subsequent entrants are the clear beneficiaries of reduced technological uncertainty because the first product of that kind had necessarily already being approved. This is true regardless of entry order within the relevant device product code – which may or may not be different from entry order within the functional category.

Because this analysis is limited to a smaller sample of only cardiovascular devices, I first repeat the product-code-level analyses of Table 1.5 for the subset of cardiovascular devices alone. The results of this analysis are presented in Panel A of Table 1.7 and yield coefficients of a very similar magnitude and statistical significance to those seen in Table 1.5: being first within a product code is associated with a regulatory approval process that is 5.1 to 6.8 months longer. I next proceed with the analysis at the functional category level. I find little evidence of the importance of reduced technological uncertainty in explaining subsequent approval times (Panel B). The results suggest that on average, being first within a functional category is associated with a regulatory approval process that longer, but these results are not statistically significant at any conventional levels.[25]

In Panel C, I ask how much – if any – of the delay seen for a new entrant in a product code is reduced when a highly related product has already completed the regulatory approval process. Specifically, I control for the resolution of a large degree of technological uncertainty (at the functional category level) and then look at the residual relationship between product code entry order and approval delay. The statistically and economically non-significant coefficients on "First in Category" suggest a very limited role for technological uncertainty in explaining regulatory delays. However, in this specification, being first within a product code is associated with a regulatory approval process that is 5.3 to 7.2 months longer and these coefficients are statistically significant at the 1 percent level. Indeed, it seems that the delineation of a new product code *itself*, rather than the novelty of the technology involved

[25]In models not presented, I also perform a "placebo test" in which I randomly assign each of the devices to one of eight arbitrary dummy categories and then run the same set of regressions. As would be expected, a prior approval of another randomly selected and unrelated cardiovascular device does not help in predicting approval times for subsequent cardiovascular devices.

Table 1.7: *Technological Novelty in Cardiovascular Devices*

	(A) Ln Approval Time	(B) Approval Time (Months)	(C) Approval Time (Months)
Panel A: Cardiovascular Subsample Only (by Prod. Code)			
First in Product Code	0.2334*	5.1143**	6.8224**
	(0.1292)	(2.4862)	(2.6205)
N	183	183	163
R^2	0.5009	0.4372	0.4118
Panel B: Devices in 8 Functional Categories			
First in Category	0.1624	2.7185	9.0857
	(0.2767)	(5.3434)	(5.8699)
N	183	183	179
R^2	0.4899	0.4206	0.4218
Panel C: Controlling for Technological Uncertainty			
First in Product Code	0.2327*	5.2872**	7.1890**
	(0.1376)	(2.6466)	(2.8121)
First in Category	0.0041	-1.1056	-2.1300
	(0.2934)	(5.6446)	(5.7774)
N	183	183	163
R^2	0.5009	0.4374	0.4125

* p<0.05, ** p<0.01, *** p<0.001

This table looks at first entrants and their respective approval delays in a) product codes b) functional categories and c) both in the same model.

All models include firm and year fixed effects. Models also include controls for whether a product was granted "priority" (expedited) review and a count of the applicant firm's approved applications at the time of submission.

Column 1 presents a log-linear model, while Column 2 translates the result into months. Column 3 restricts the sample to only the first entrant plus those subsequent entrants who submitted applications *after* the first entrant's approval decision was finalized.

in that product's primary function is the strongest predictor of longer regulatory approval times.

1.5.3 Sources of Uncertainty Part 2: Reduced Uncertainty about Application Content and Format through Publication of Objective Regulatory Criteria

This section addresses cases in which procedural uncertainty about new product application content and format is resolved through the publication of formal guidance documents. This type of uncertainty occurs in the absence of clear guidelines about the protocol for evaluating a new product, leading to uncertainty on the part of the regulator as to how to evaluate the results of clinical studies and other (e.g. biocompatibility and engineering) tests and uncertainty on the part of firms as to what information to submit to the regulator and in what format. An example of the resolution of procedural uncertainty can be seen in the publication of FDA guidance documents related to the regulation of drug eluting stents, which is described in Section 1.1.

The publication of formal FDA guidance about a specific product or class of products[26] is the primary way in which protocols for evaluating a new medical device are formally established. In 1997, the FDA announced that it would formalize its *Good Guidance Practices* in order "to provide transparency and consistency in policy development" moving forward (FDA, 2007).[27] Examples include documents that describe the:

- *design, production, ...manufacturing, and testing of regulated products*
- *processing, content, and evaluation or approval of submissions*

[26]The history of FDA guidance dates back to the 1970s, when the FDA began to issue "guidelines" for clinical trials, a regulatory norm (less stringent than formal rule-making) that would lead to an important role of "guidance documents" in communicating structures of clinical experiment and drug development to the pharmaceutical industry moving forward (Carpenter, 2010). Guidance documents continue to shape the FDA's regulation of medical products to this day and their scope has expanded with that of the FDA to include medical devices and other products.

[27]Guidance documents are issued by the FDA, however their standardization in the 21st century has been governed by a formal congressional regulation: on September 19th, 2000, Congress approved regulation (21 CFR 10.115), which outlined the FDA's policies and procedures for developing, issuing, and using guidance documents. While the FDA had released various medical device guidance documents prior to 2000, they were not standardized and so their interpretability and significance were more limited.

Table 1.8: *Case Studies, Publication of Objective Regulatory Guidance*

Product Type	Date Published	Product Code(s) Affected	Pre-Guidance Approval Time (Months)	Post-Guidance Approval Time (Months)	N (obs)
Drug-Eluting Stents	(3/1/2008)	1	15.38	8.75	9
Intravascular Stents	(4/18/2010)	4	13.50	8.02	42
Heart Valves	(1/20/2010)	3	11.83	9.00	6
Catheter Ablation Devices	(8/5/2008)	1	14.29	9.36	7
			(N=49)	(N=15)	(N=64)

This table summarizes four recent cases in which objective regulatory guidelines were published by the FDA for major categories of cardiovascular devices. In each of the cases, regulatory delays fall substantially in the period after guidance is published. The data are raw and un-adjusted for potentially relevant covariates.

- *inspection and enforcement policies*

In recent years, the FDA has released several pieces of guidance related to medical devices, which are available from the Office of Device Evaluation (ODE). Of the 162 pieces of guidance released since the approval of GGPs, the vast majority deal with Class II (moderate-risk) devices and several others relate to general evaluation practices, rather than focusing on specific technologies. I consider a set of high-risk device policy changes around the publication of four pieces of formal guidance. These pieces of guidance directly outline objective evaluation criteria relating to the PMA process for nine specific product codes of high-risk (Class III) cardiovascular devices. These guidance documents and the dates of their publication are listed in Table 1.8.

In each of these cases, uncertainty around application content and format was largely resolved through the release of formal content and evaluation guidelines for new product applications and in each of these cases, average approval times subsequently decreased. In the analysis that follows, I define "post-guidance" applications as those that were submitted one month or more after the release of guidance for a given product code or set of products. This ensures that all post-guidance applications were able to incorporate information from the FDA guidance into their application prior to submission.

Table 1.8 shows that (without any controls), following the publication of regulatory guidance, an average decrease in regulatory approval times of 2.8 to 6.6 months was observed

Table 1.9: *Publication of Objective Regulatory Guidance*

Outcome = Approval Time				
	(1)	(2)	(3)	(4)
Post-Guidance	-10.0515**	-8.3711†		
	(4.7666)	(4.9293)		
ATE (Post-Guideance)			-6.0696***	
			(0.8577)	
ATT (Post-Guideance)				-8.5193***
				(3.0972)
Controls	X	X	X	X
Excluding first 2 Entrants		X		
Pre-Post Analysis	X	X		
Matched Analysis			X	X
N	64	51	192	192
R^2	0.3401	0.3944		

†p<0.10 * p<0.05, ** p<0.01, *** p<0.001
Controls = submission year, priority review eligibility, applicant firm, applicant firm's to-date total approvals, product code, entry order. Column 3: nearest neighbor matching on observables to identify two similar "untreated"/control applications.

in affected groups. Table 1.9 includes statistically appropriate control variables and estimates the covariate-adjusted average decrease in approval time associated with the publication of guidance. All models in Table 1.9 include product code fixed effects and controls for whether a product was granted "priority" (expedited) review, year of submission, and a count of the applicant firm's total approved PMAs at the time of submission. The first column of Table 1.9 presents a covariate-adjusted pre-post analysis of approval times with respect to the publication of regulatory guidance for all applications in affected categories. Column 2 excludes the first two entrants in each group so as not to bias the results by including applications in the pre-guidance average that are known to have longer approval times.

Although these results are consistent with the conclusion that uncertainty about application content and format is an important driver of first mover disadvantage in the medical device regulatory process, one might be concerned about likely endogeneity in

the FDA's decision to publish guidance for these particular devices. For example, it may be the case that more "popular" categories of medical devices were more likely to get regulatory guidance. To address potential selection, Column 3 presents results from a nearest neighbor matching analysis in which each device in a "treated" product code (i.e. one in which guidance was at some point published) is matched to two other "untreated" devices (other high-risk cardiovascular devices in product codes in which guidance was not published) based on *ex ante* observables about the application and relevant product code including entry order, submission year, total PMA submissions in the product code at the time of a given application, and average approval times in the product code. Both the average treatment effect (ATE) and average treatment effect on the treated (ATT) of the introduction of regulatory guidance are presented. Even the most conservative estimate (the ATE presented in Column 3), suggests that the resolution of procedural uncertainty through the publication of formal guidance is associated with a 6.1 month (approximately 185 day) reduction in regulatory approval times. In this subsample, that represents a 41 percent reduction in regulatory approval times.

The results above complement existing research on the determinants of entrepreneurial success in the device industry: Chatterji (2009) finds evidence that for venture capital funded companies, familiarity with protocols is more important then technical knowledge for predicting firm successes. My results in turn, suggest that uncertainty about the content and format of a new product application is more important then technological uncertainty about a product for predicting regulatory approval times.

1.6 Entrant Type and Strategy

The final empirical section of this paper considers the relationship between firm type and market entry strategies. The market entry hypothesis in Section 1.3.2, suggests that in the presence of delays under regulatory uncertainty, financially constrained firms should be less likely to enter new device markets as pioneers. Looking within the ownership and financial data assembled for all cardiovascular device firms in the data, I identify those firms

that are most likely to be financially constrained. For this exercise, I define a financially constrained firm as one that a) is not publicly listed, b) does not have revenues of more than $500 million per year, and c) is not a subsidiary of firms of type a or b. This leaves a set of small, privately held firms, none of which are subsidiaries of larger companies.

Using the criteria above, Table 1.10 considers how the proportion of financially constrained firms varies with the application of the above definition. The most conservative definition ("Definition 1") looks only at those firms that were defined as "financially constrained" at least one year before an application was submitted. The next definition ("Definition 2") excludes those firms that were or became subsidiaries of established firms within a five year window of a given PMA submission (for example, Irvine biomedical's percutaneous cardiac ablation catheter was submitted to the FDA for approval in 2004, acquired by St. Jude in the same year, and received approval in 2005. This product would count as "financially constrained" under Definition 1, but not under Definition 2, which is broader). The third definition ("Definition 3") broadly classified "financially constrained" firms as those that never met criteria a, b, or c above – that is, they were never part of a more established (less financially constrained) company.

I find that financially constrained firms make up 6.9 to 17.2 percent of the sample among pioneer entrants but 14.3 to 23.0 percent of the sample among follow-on entrants. The difference between these two samples is statistically significant at the 10% level for Definitions 2 and 3 in two-sample t-tests of means with unequal variance. The difference between the two samples is not statistically significant based on Definition 1, likely a result of the small sample sizes used to calculate the averages, however the average differences between proportions of financially constrained firms among pioneers and follow-on entrants are consistent with the hypothesis's predictions in all cases.

1.7 Discussion and Conclusion

I have considered how regulatory uncertainty is related to first mover advantages and disadvantages in the regulatory approval process for new chemical drugs and high-risk

Table 1.10: *Financially Constrained Firms' Market Entry Strategies*

	(1) Pioneer Entrants	(2) Follow-On Entrants	(3) P[(1) = (2)]†
Definition 1	17.2%	23.0%	0.3169
Definition 2	10.3%	19.4%	0.0657
Definition 3	6.9%	14.3%	0.0751

† P-values are from a 2-sided t-test with unequal variances.

A financially constrained firm as one that is not a) publicly listed, b) does not have revenues of more than $500 million per year, and c) is not a subsidiary of firms of type a or b.

Definition 1: only those firms that were defined as "financially constrained" at least one year before an application was submitted.
Definition 2: excludes those firms that were or became subsidiaries of established firms within five years of a given PMA submission.
Definition 3: firms that never met criteria a, b, or c above.

medical devices. The data on FDA drug approvals show that earlier entrants into drug markets experience a slight advantage over later entrants in the regulatory approval process. However, the data on FDA medical device approvals reveal large fixed costs of early entry into new device markets: pioneer entrants in new device product codes spend 34 percent longer in the approval process than the first follow-on innovator in that product code. I estimate that the magnitude of the additional approval time faced by pioneer innovators is approximately 7.2 months, a large delay relative to the 2.8 to 3.8 years of *de facto* market exclusivity that a pioneer innovator can expect. Back of the envelope calculations suggest that a delay of this length could translate to a loss of approximately 8 percent of expected lifetime product revenues and that the opportunity cost of capital of a delay of this length is upwards of 7 percent of the total cost of bringing a high-risk medical device to market. I find that financially constrained firms are less likely to act as pioneer innovators. This result is consistent with the prediction that firms with more capital should be better able and/or more willing to bear the additional regulatory costs of pioneer entry.

I analyze regulatory approval times under two sources of uncertainty by looking at

settings in which either technological or procedural uncertainty are greatly reduced. I find that large delays for the first entrant in a product code persist even when a great degree of technological uncertainty has been resolved. In contrast, I find that the resolution of uncertainty about application content and format through the publication of formal regulatory guidance to clarify product evaluation criteria is associated with substantially reduced approval times thereafter. These results complement other research into the importance of understanding regulatory protocols in the medical device industry; for example, Chatterji (2009) finds that regulatory and procedural knowledge is more important then technical knowledge for predicting firm success.

This paper contributes to a broad literature about the relationship between regulatory uncertainty and innovation incentives – in particular, with respect to medical devices and other emerging categories of medical products, where methods for effectively incentivizing innovation remains poorly understood (Xu et. al., 2013). Generally speaking, incentives for engaging in R&D activity are negatively influenced by increases in the costs and risks of developing new products (Grabowski et. all, 1976). This study is therefore related to research on how R&D incentives affect the pipeline of innovation. Budish et. al. (2013) find evidence that private firms' incentives to innovate have meaningful impact on the level and composition of R&D investments. Moreover, they find that increases in R&D – in particular in cases where there may be underinvestment – have the potential to generate large improvements in patient health. This paper suggests that under regulatory uncertainty, the nature of the approval process for new medical products can create disincentives for pioneer entry by meaningfully increasing the length of the product development period for novel products. This, in turn, affects firms' strategies for entering new markets and could lead to under-development of new medical technologies, although a definitive statement about the extent to which this occurs in medical device innovation is beyond the scope of this paper.

The results also suggest that the regulation of medical technologies could be made more efficient through the earlier resolution of uncertainty about new product application content

and format whenever possible. This could be done, for example, through the earlier release of guidance documents and/or by encouraging firms to work with the FDA very early in the new product development process in order to help the FDA develop evaluation standards or formal guidelines for a new medical technology before a regulatory approval application is officially submitted.[28]

This study could be expanded in a number of ways. First, it would be interesting to know more about regulatory delays themselves: what happens over the period between PMA submission and the FDA's ultimate approval decision? Relatedly, how much of an observed delay is due to the FDA requesting additional information from device companies and what types of information are requested? And finally, are certain types of information requests (e.g. additional product manufacturing specifications) faster to execute and/or evaluate than others (e.g. additional biocompatibility tests)? The data that I use in this study do not allow me to satisfactorily answer these questions. In my conversations and interviews with medical device companies, it has frequently been expressed that the FDA's requesting of additional clinical or technical information is a major source of uncertainty for device firms entering a regulatory process in which the regulator's expectations are unknown *ex ante*. Unfortunately, no quantitative data that I know of are able to shed light on the relative frequency or size of these types of delays. As such, this would be a very fruitful area for future data collection and aggregation – both within and beyond the context of medical device regulation.

The results do not address the onerous process of regulatory reform. While it seems likely that earlier engagement and articulation of regulatory requirements on the part of the FDA could decrease regulatory approval times and minimize delays, the process for implementing any large changes to the formal FDA regulatory policy is complex, time-consuming and institutionally entrenched. Yet, my findings suggest that there may be room for regulatory process efficiency improvements in the regulation of medical devices and other categories of products with a high degree of regulatory uncertainty, as delays are

[28]Interviews with regulatory consultants revealed that this is a strategy that they often recommend.

most prominent in cases where evaluation procedures are poorly defined and delays can be substantially reduced through clear articulation of the regulator's expectations. FDA Commissioner Margaret Hamburg has noted that "these challenges are not the FDA's alone." Indeed, she argues that in order "to truly leverage advances in science and technology, there must be a collaboration of all relevant stakeholders, including government, academia, and industry. The FDA must work with its partners to promote innovation and creativity at various points throughout the development process" (Hamburg, 2010).

New medical technologies are poised to continue to grow in importance over the coming years and earlier engagement and clear communication between regulators and innovators may be able to accelerate their regulatory approval. The goal of such communication should be to mitigate uncertainty about regulatory protocols as early as possible in the regulatory approval process. By minimizing content and format-related regulatory delays for entrants into new product markets, the FDA can also increase its overall efficiency and free up reviewer resources, potentially improving the process of regulatory approval not only for early entrants, but also for later ones as well.

Chapter 2

Conflicts of Interest and the Economics of FDA Advisory Committee Voting

2.1 Introduction

The influence and importance of expert recommendations is clear in many health care settings, where the recommendations of skilled medical professionals guide economic activity and shape patient and societal welfare. Yet in the health care industry and beyond, the structure of expert advising relationships may result in principal-agent issues.

At its most general, an agency dilemma can arise when the expert (agent) has incentives that are different from those of the individual or society (principal) on whose behalf they are making a recommendation. This principal-agent problem could take many forms ranging from "physician induced demand" for health care services (McGuire, 2000) to the acceptance of industry funding for academic research (Brennan et. al., 2006). In the health care setting, when we talk about a "conflict of interest," we are almost always talking about a setting in which a principal-agent problem could potentially arise, due to competing incentives on the part of an agent. This paper focuses on a specific set of conflicts of interest that may arise in

one important expert advising setting: committee voting on new medical product approvals at the U.S. Food and Drug Administration (FDA).

In the United States, the FDA is tasked with regulating all medical products, including drugs and medical devices. The FDA also acts as a gatekeeper to all medical product markets: without the agency's marketing approval, it is illegal to market a medical product in the United States. In making its decisions about whether or not to grant marketing approval for new drugs, devices, and other products, the FDA relies heavily on the use of expert "advisory committees," that come together to make approval recommendations for new high-risk products. These committees are staffed by scientific experts and their recommendations are highly predictive of agency approval decisions: in a recent study, the FDA went on to approve 88 percent of original new drug applications (NDAs) and biologics license applications (BLAs) that were endorsed by advisory committees and of those applications that were not endorsed by FDA committees, 86 percent were subsequently denied marketing approval (Smith et. al., 2012).

Individuals on FDA advisory committees are typically experts in their fields (e.g. cardiology, anesthesiology, or the evaluation of biostatistics). When committee members have existing or recent financial conflicts of interest, these conflicts of interest must be disclosed to the FDA and publicly declared at the start of a meeting in which a conflicted member participates. Financial relationships present one obvious set of potential biases that may affect experts' recommendations, and are analyzed here.

Committee members often also have longstanding academic careers: individuals who serve on advisory committees have both past and current institutional affiliations, a career-long history of co-authoring relationships, and varying degrees of expertise. These diverse professional characteristics also present additional sources of potential voting bias and will be discussed below and addressed empirically in future work.

Of course, financial and professional relationships may also add valuable information to committee members' decision-making: individuals may better understand industry-sponsored research if they have been involved in similar studies in the past, may be

better able to assess the value of a co-advisory committee member's opinions given prior professional experiences with that individual, or may be able to make better assessments of a new type of technology's potential, given their expertise in that area of research. For example, Li (2013) finds that reviewers of NIH grants are biased in favor of projects in their own area of expertise, but also better informed about those projects.

In this preliminary analysis, I consider the voting behavior of 1545 uniquely identified individuals at 110 new product meetings over a seven-year period. The results presented focus on individuals' voting propensities vis-à-vis their own and their co-committee members' financial conflicts of interest, however future research will consider other potential sources of bias, such as institutional affiliations, academic expertise, and co-authoring relationships.

I find that financially conflicted individuals – broadly defined to include anyone who has had a financial relationship with industry in the past 12 months – are far more likely (about 18 percent) to vote favorably for new medical devices, but not any more likely than financially unconflicted individuals to vote favorably for new drugs. I present a basic conceptual framework for assessing how this may affect panel-based recommendations and outcomes and what kinds of financial conflicts of interest are predictive of voting patterns.

I find that the average propensity of financially conflicted individuals to vote more favorably for new medical devices is not driven by individuals with a direct financial conflict of interest with the firm whose product is under consideration. Rather, I find that the pattern is driven by individuals with another financial relationship with industry voting favorably for a competitor's product – that is, an individual who has a financial relationship with device Firm A is more likely to vote favorably for a new medical device produced by Firm B.

This result initially seems counterintuitive: it would appear that individuals who work for Firm A are voting against their own financial interests when voting to endorse their competitor's product. However, voting incentives in the medical device setting are dynamic: as I have shown in other work (Stern, 2014), there are significant costs associated with being a "first mover" in a new medical device product market. One corollary of this first

mover regulatory disadvantage is that by facilitating the approval of early entrants, advisory committee members could help "pave the way" for subsequent entry by competitors (including the firms in which conflicted individuals have an interest).

I look for evidence of "paving the way" behavior by asking whether or not favorable voting patterns among financially conflicted individuals are related to product entry order.[1] Specifically, I look for evidence of an interaction effect that would suggest that favorable voting propensity among those with a financial conflict of interest is even stronger when the product in question is an early entrant within a device category. I find preliminary evidence in support of this type of behavior, although the statistical power of these models is limited by sample size.

An important aim of this paper is to move beyond a framework in which it is assumed or implied that conflict of interest biases individuals identically regardless of the composition of their peers on a panel. I introduce a conceptual model that includes basic peer effects: how an individual votes is modeled as a function of characteristics of that individual (financial conflict of interest, academic expertise, gender, institutional affiliation) as well as characteristics of their fellow panel members, interacted with the reference individual's characteristics and voting behavior. This allows me, for example, to ask if a financially conflicted person's propensity to vote in favor changes with the fraction of other in favor voters or the fraction of other conflicted voters on the panel.

Similarly to Li (2013), the identification strategy for this exercise will rely on the fact that the relative composition of panels (e.g. in terms of fraction of conflicted members, members' academic expertise, gender, and institutional affiliations) changes over time in a way that is arguably independent of the relationship between panelist characteristics and voting behavior. This is due to the rotating terms of permanent panel members: committee members are staggered in a set of 4-year rotating terms, such that the same individuals will

[1] A second explanation for the observed voting patterns could be that there is an important unobserved variable that could be thought of broadly as something like "optimism about medical technology." This particular type of optimism could drive the results seen if it makes certain individuals both more likely to take money from industry as well as more likely to take a favorable view of new medical technology. This could also certainly occur in the presence of other phenomena.

appear multiple times in the data with different peer groups.

In a model in which individuals' voting behaviors are influenced not only by their financial conflicts of interest, but also their panel peers, I find that there is a positive association between the fraction of other panel members' in favor votes and an individual's probability of voting in favor of a new product, regardless of whether or not that individual is conflicted.[2] Further, I estimate a positive interaction effect between other panel members' in favor votes and a conflicted individual's probability of voting in favor, suggesting that at very high rates of within-meeting in favor voting, the difference between conflicted and unconflicted individuals' in favor voting propensities will be the largest (the difference is nearly 25 percentage points at 3 standard deviations above the sample average).

More generally, allowing for peer effects, I estimate that the bias associated with a financial conflict of interest will vary positively with the fraction of other panelists (or other conflicted panelists) voting favorably for a new product. Future work will focus on simulating meeting outcomes if the same individuals were to sit on panels with different compositions of colleagues. Future work will also expand the set of characteristics of individuals and peers used in predictive models of voting behavior.

2.2 Background

2.2.1 Medical Product Regulation in the United States: The FDA

In the United States, the FDA regulates all medical technologies. An agency of the Department of Health and Human Services, the FDA is responsible for the oversight of two trillion dollars worth of products every year, including all over-the-counter and prescription drugs and medical devices (Babiarz and Pisano, 2008; Hamburg and Sharfstien, 2009).[3]

The FDA is organized into centers, each of which is tasked with the oversight of a different type of product. The two centers most relevant to the analysis here are the Center

[2]This result is consistent with herding behavior discussed by other authors (e.g. Smith, 2012).

[3]The FDA also regulates all other existing and emerging classes of medical products.

for Drug Evaluation and Research (CDER) and the Center for Devices and Radiological Health (CDRH[4]), which regulate chemical drugs and medical devices, respectively.[5] Within the CDER, the Office of Drug Evaluation is responsible for the approval of new drugs, and within the CDRH, the Office of Device Evaluation is responsible for the review and approval of medical devices.

These centers are often tasked with assembling committees of "experts with recognized expertise and judgment in a specific field" for the purpose of assessing the safety and efficacy of new products (FDA, 2008). In general, FDA review begins with regulators completing an initial review of a marketing approval application for a new product and identifying those questions that require external input. As necessary, the relevant center at the FDA then convenes an advisory committee meeting to obtain input from experts through presented information, discussion, questioning, and voting by committee members (Smith et. al., 2012). Committee members are not employees of the FDA itself, but rather "special government employees for the days they participate as members of a panel." As in other areas, these committees make recommendations only, and all final regulatory decisions regarding the approval of medical products are made by FDA officials.

Which products are ultimately reviewed by an expert advisory committee? Reviewing publicly available materials, Smith et. al. (2012) analyze FDA drug advisory committee meetings over the decade ending in 2010 and find that "new biologics, priority status applications and orphan drugs were the subject of more meetings, on a percentage basis, than new chemical entities, standard applications and non-orphan drugs." That is special categories of products were more likely to be sent to advisory committees for expert opinions than more standard types of products.

Smith et. al. (2012) note that 87 percent of meetings resulted in a clear positive or negative vote. The authors suggest that this may be a result of committee expertise and/or

[4]The CDRH also regulates radiation-emitting products such as X-ray and ultrasound machines.

[5]Other categories of products are also reviewed by specialty centers within the FDA (e.g. biologics and human cells, tissues, and cellular- and tissue-based products are reviewed by the FDA's Center for Biologics Evaluation and Research, CBER).

herding behavior in voting.

2.2.2 Conflicts of Interest in Medicine

It has been suggested that in recent decades, the drug industry "has gained unprecedented control over the evaluation of its own products" (Angell, 2008). Brennan et. al. (2006) write about the "extraordinary challenges to the principles of medical professionalism" posed by physicians' conflicts of interest and by their relationships with pharmaceutical companies and medical device manufacturers. Indeed, in recent years, many academic medical centers, which include medical schools and their affiliated hospitals have taken steps to limit the influence of pharmaceutical companies on their physicians. Moreover, many professional organizations, such as the American Medical Association, the American College of Physicians, and the Accreditation Council for Continuing Medical Education have issued or revamped their own guidelines for physicians' interactions with pharmaceutical companies (Studdert et. al., 2004). The practice of disclosure is a commonly prescribed remedy for conflicts of interest and is used commonly in medicine and the health care industry: the Medicare Payment Advisory Commission and the Accountable Care Act both impose disclosure requirements as do most medical journals (Loewenstein et. al., 2012).

Others have pointed out that disclosure itself may lead to unintended behavioral consequences on the part of the disclosing agent. One such consequence is "increased bias," which has two potential mechanisms: 1) strategic exaggeration – providing more biased advice to compensate for the fact that the audience will discount that advice if it knows about an agent's conflict of interest – and 2) moral licensing – feeling justified in providing biased advice because the principal (advisee) has been warned that the agent has a conflict of interest (Loewenstein et. al., 2012).

Indeed, it has been demonstrated that physicians themselves believe that conflicts of interest create biases and they adjust their assessments of their peers' research when given information about financial conflicts of interest. Kesselheim et. al. (2012) look at how disclosure of industry funding affects internists' assessment of the rigor of clinical studies

for hypothetical drugs. They find that an industry funding disclosure results in physicians downgrading their assessment of the rigor of a trial, decreases their confidence in the results, and lowers their willingness to prescribe the drug. In particular, they find that "physicians were half as willing to prescribe drugs studied in industry-funded trials as they were to prescribe drugs studied in NIH-funded trials," suggesting a meaningful behavioral adjustment on the part of physicians in response to their perception of peers' financial conflicts of interest.

However, there are also important reasons to believe that physician cooperation with industry could facilitate better patient outcomes. Physicians can provide an essential understanding of medicine and technology that can be incorporated into new medical products. Further, physician involvement in industry activities that generate conflicts of interest (e.g. clinical trials) has been cited as a way in which physicians can learn about new technology (Chatterji et. al, 2008). For example, the diffusion of new cancer treatment technologies has been linked to precisely those researchers who are involved in running clinical trials for industry (Agha and Molitor, 2014).

Industry engagement may not only accelerate learning, but also facilitate product innovation. Chatterji et. al. (2008) find that those medical device patents filed by physician-innovators "had more influence on subsequent inventive activity than non-physician patents" and argue for an open environment for physician-industry collaboration in the medical device discovery and development process. Relatedly, Cockburn and Henderson (1998) find that pharmaceutical firms that are more "connected" to basic research perform better at drug discovery. One component of this connectedness is co-authoring relationships between industry researchers and researchers at public institutions: higher levels of co-authorship are correlated with private sector research productivity, implying potential welfare benefits of industry-funded (or partially industry-funded) research initiatives. Relatedly, Zinner et. al. (2009) find that "life science faculty with industry research support were more productive than faculty without such support on virtually every measure."

This paper is thus related to an important set of policy questions about balancing the

drawbacks and potential biases inherent in financial conflict of interest against the potential for additional expertise and innovative activity that comes with industry relationships.

2.2.3 FDA Advisory Committees and Financial Conflicts of Interest

While financial conflicts of interest have been studied in the context of clinical trials and the prescription of drugs, little research has considered the role of conflicts of interest in expert committee decision making. An expert committee setting of particular interest is that of FDA advisory committees, which are convened to make recommendations on the safety, efficacy, and subsequent approval of new medical products such as new chemical and biologic drugs and new high-risk medical devices. Those who study conflicts of interest in medicine have identified this setting as one where financial conflicts of interest could also affect important high-stakes decisions about new product approvals (Angell, 2008).

FDA committees must be composed of independent members without any financial conflicts of interest, except in cases where a conflict of interest waiver is granted by the FDA. The FDA's policy[6] on participation of financially conflicted individuals is as follows:

"When [the] FDA determines that an advisory committee member has a financial conflict of interest, the agency may grant a waiver that allows the member to participate in an advisory committee meeting if certain criteria and policies are met. In general, FDA may grant a waiver if the member's expertise is considered essential to the committee's discussions and recommendations. FDA must also take into consideration a cap on the number of waivers that can be granted each year. FDA searches for experts who have the necessary expertise without conflicts of interest; yet, in some cases, many of the top authorities in specialized scientific fields may have a conflict of interest. When FDA grants a waiver, the financial interests associated with the waiver are posted on FDA's website along with the reasons for granting the waiver."

In short, the FDA may grant a waiver when an expert's knowledge is arguably required for the committee's understanding of the product in question. As indicated above, there is a cap on the fraction of waivers that can be granted each year, although in aggregate, this cap is far above the actual fraction of waivers granted during the period of time I consider below.

[6] http://www.accessdata.fda.gov/FDATrack/track?program=advisory-committees&id=AdvComm-waivers&fy=all

Financial conflicts of interest themselves vary along several dimensions, such as the total sum of money involved and the type of financial relationship in question. Reasons for conflicts include a host of profit-generating activities such as direct employment, current investments, patents, contracts, grants, cooperative research and development agreements (CRADAs), consulting agreements, and honoraria for speaking and writing agreements within past 12 months. These conflicts are divided into types: "index conflict" arises when a committee member has financial ties to the product's sponsor (the firm working to bring that product to market), while "competitor conflict" arises when a committee member has financial ties to a competitor of the sponsor. Both are broadly categorized as constituting "any financial conflict of interest" by the FDA and separately categorized into one or more (non-mutually-exclusive) categories.[7]

The primary study that considers financial conflicts of interest and voting at the FDA, Lurie et. al. (2006), looks only at four years worth of committee voting data and only at drug meetings. The authors find that while disclosures of financial conflicts of interest are common, only a weak relationship can be detected between conflicts and favorable voting behavior and a simulation exercise suggests that conflicts of interest are unlikely to influence overall panel outcomes (majorities) in the meetings considered. It has separately been noted that "there is much more to an advisory committee meeting than votes on product approval," including the flow of information to the public and to the regulator (Moffitt, 2012). This could also include the influence of conflicted members on their colleagues, a scenario that will be discussed below.

Moffitt (2012) finds some evidence that the participation of conflicted individuals on new drug advisory committees may be related to the probability of subsequent safety issues. In comparing the outcomes of drugs recommended by committees both with and without (a binary measure of) participation by financially conflicted individuals, she finds that "drugs reviewed by committees with no conflict of interest waivers were associated with

[7]The data used also include an indicator for "other" financial conflicts of interest which are industry relationships that are not directly associated with the reference product's company or one of it's competitors.

significantly fewer subsequent drug safety alerts." Yet she also notes selection bias arising from the use of observational data and a lack of consistent statistical significance.

2.3 Conflict of Interest: Conceptual Framework

2.3.1 Voting Propensities and Bias

The behavior of individuals in voting situations is observed, as are the presence or absence of financial conflicts of interest. FDA committees make recommendations based on a simple majority,[8] so if the total proportion of in favor ("yes") votes for a new product is greater than 0.5, a product will be recommend. Thus proportion of in favor votes, $p(y)$ can be estimated easily as:

$$p(y) = \frac{1}{N}\sum_{i=1}^{N} v_n$$

where v_n is observed voting behavior and $v_n = 1$ if individual n votes in favor and $v_n = 0$ if individual n is not in favor (against). When $p(y) > 0.5$, the committee makes a recommendation in favor of a new product.

However, the acknowledgement of the potential principal-agent problem that arises with the participation of individuals with a financial conflict of interest implies that individuals with a conflict (c) may have different voting propensities *ceteris paribus* than individuals who are unconflicted (u), where each could be calculated separately as:

$$p(y|c) = \frac{1}{A}\sum_{i=1}^{A} v_a \;\; and \;\; p(y|u) = \frac{1}{B}\sum_{i=1}^{B} v_B$$

where A is the total number of conflicted individuals and B is the total number of unconflicted individuals.

More broadly, population estimates of P(y) could be thought of as weighted averages of conflicted and unconflicted individuals propensities to vote favorably for a new product

[8] In the case of a tie, the committee chair will break the tie; this is discussed later.

given the same observable data about that product, Ω:[9]

$$P(y|\Omega) = \alpha P(y|c,\Omega) + (1-\alpha)P(y|u,\Omega)$$

where α is equal to the ratio of conflicted individuals to total voters and Ω is the observable information about a given product that all voting members can see (e.g. results from clinical trials). In subsequent sections, I will change this simple framework to one that includes peer effects, but this simple set-up is instructive for preliminary exercises.

2.3.2 Bias

The first question of interest pertains to the existence of voting bias that is related to financial conflict of interest: when is $P(y|c,\Omega) \neq P(y|u,\Omega)$ in a way that would suggest bias associated with financial conflict of interest? To simplify, we can start by asking: when are voting patterns different among conflicted vs. non-conflicted individuals? In calculating probabilities, there are five possible scenarios:

1. $P(y|c,\Omega) = P(y|u,\Omega) = 1$: unanimous votes in favor, no evidence of bias
2. $P(y|c,\Omega) = P(y|u,\Omega) = 0$: unanimous votes not in favor, no evidence of bias
3. $0 < P(y|c,\Omega) = P(y|u,\Omega) < 1$: identical (average) voting behavior between conflicted and unconflicted individuals, no evidence of bias
4. $P(y|c,\Omega) < P(y|u,\Omega) \neq 0$: unconflicted individuals are more likely to vote favorably for new product than conflicted individuals. This is a) not consistent with empirical evidence (presented in Section 5) and b) more generally not a scenario of interest when we think about the specific principal-agent problem (and subsequent societal risk) that could be generated by the presence of direct financial conflict of interest.

[9]Future versions of this paper will relax the assumption that information is the same across all voting individuals, but for the current analyses, I assume that all individuals have information Ω.

5. $P(y|c,\Omega) > P(y|u,\Omega) \neq 0$: what would be expected if an individual with a financial conflict of interest is an imperfect agent: this is bias that is aligned with the direct financial conflict of interest and has potential to lead to the approval of riskier products.

Thus, to learn about the existence of bias in voting behavior, we are interested in whether the different components of the inequality in scenario 5 are statistically different from one another, which I will test empirically in the data.

An obvious limitation of existing research on the voting behavior of conflicted individuals is the ubiquitous implicit assumption that conflict (potentially) affects individuals in the same way, regardless of who their peers are. That is, existing studies assume that the composition of the panel on which an individual sits is irrelevant to her voting behavior. This paper proposes a framework in which not only the presence or absence of a financial conflict of interest, but also the presence or absence of conflicted peers (or peers of varying levels of expertise and varying academic relationships with other panelists) impacts each individual's voting behavior.

2.3.3 Peer Effects: Which Factors Predict Voting Behavior?

In order to move beyond a framework in which it is assumed or implied that conflict of interest biases individuals identically regardless of their peers, I introduce a conceptual model that includes basic peer effects. In this model, how an individual votes is a function of her characteristics (financial conflict of interest, academic expertise, gender, institutional affiliation) as well as characteristics of her fellow panel members (the fraction of others with a financial conflict of interest, others' academic expertise, whether or not an institutional affiliation or professional relationship is shared with another panel member) interacted with those individuals' voting behavior.

The identification strategy I will use relies on the fact that the relative compositions of panels (e.g. in terms of fraction of conflicted members, members' academic expertise, gender, institutional affiliation) changes over time and does so in a way that is arguably independent of the relationship between those variables and voting behavior. This is due to

the rotating terms of permanent panel members:[10] committee members are staggered in a set of 4-year rotating terms, such that the same individuals will appear multiple times in the data with different peer groups. Several temporary members also appear in the data in multiple meetings, although there is more potential for endogeneity in their peer group – for example because they are brought in because of a specific area of product expertise that is missing among permanent members.[11]

Below I present one way of thinking about an empirical model for voting with peer effects. At its most simple, the probability of an in favor vote by individual i on panel z can be thought of as a function of both the individual i's characteristics as well as the characteristics of the other voting members of the panel.

$$p(y_i) = f[c_i, e_i, X_i, \sum_{j=1}^{N-1} \beta_{j \neq i} * (c_j + e_j + a_{ij}), \ z, t]$$

- c_i = an indicator of an individual's conflict of interest
- e_i = academic expertise of individual i, measured, e.g., as a binary indicator of whether or not an individual is an academic superstar or by publication count
- X_i = other observables about an individual – e.g. gender, whether or not they are currently employed at a (top) research institution
- $\beta_{y|\neq i}$ = a binary indicator of an in favor vote by individual j (where j $\neq$ i)
- c_j = an indicator of a co-panel member's conflict of interest
- E_n = Binary indicator of whether or not individual j is an academic superstar or publication count
- a_{ij} = an indicator for whether individual j shares a past or current institutional affiliation or co-author relationship with individual i
- z and t = panel and year fixed effects, respectively

The expression after the summation sign in the expression above should relate to the *additional* influence of another in favor vote when that individual a) has a financial conflict of interest, b) is a scientific expert or c) shares a professional relationship or institutional

[10] This identifications strategy is similar to Li (2013), who does a similar exercise using NIH study sections.

[11] future analyses will quantify the extent of this selection issue to the extent it exists

affiliation with person i. Respectively, these terms should answer the questions: a) are individuals more/less likely to vote favorably when another conflicted individual votes favorably? b) are individuals more/less likely to vote favorably when a scientific expert votes favorably? and c) are individuals more/less likely to vote favorably when an individual with whom they have shared an institutional affiliation or a professional relationship votes favorably?

Given the above framework, we can then ask first how much additional explanatory power is gained from the inclusion of peer effects. Second, we can analyze the relative importance and magnitude of influence of different types of peer effects (e.g. financial conflicts of interest vs. professional relationships). Finally, we can consider how financial conflict of interest and professional relationships are likely to affect voting propensities and subsequent panel outcomes in this modified framework.

2.4 Data

The database used in this analysis was assembled in two stages. The first version of the database was used for the empirical analysis in Lurie et. al. (2006), which is described above. Following the publication of the Lurie et. al. (2006), the FDA commissioned the Eastern Research Group (ERG) to assess "the relationship between financial conflict-of-interest disclosure and voting patterns at FDA advisory committee meetings" (ERG, 2009). ERG was granted access to the data from Lurie et. al. (2006), which included data on all new drug meetings held by CDER between January 2001 and December 2004. ERG researchers then expanded the original database to include drug meetings through the first quarter of 2008 and to include meetings on new devices over the same period of time. While the findings of this commissioned study are not statistically rigorous and were never published in a peer-reviewed venue, the study finds that individual and meeting-level patterns were similar to those published in Lurie et. al. (2006).

The primary data set for this project was provided to me by the FDA Office of Planning

in November of 2012.[12] The data include both meeting and participant-level data for meetings that took place between the beginning of 2001 and the end of the first quarter of 2008. The meeting data include information on the relevant FDA Center responsible for the recommendation, the committee name, meeting date and meeting topic. The data were limited to "particular matters involving specific parties" (PMISP), which are those meetings that consider specific products. Other meetings (e.g. those about "matters of general applicability" where no recommendations were made about specific products) were not included.

Participant data include the participants' names, participation type (e.g. permanent or temporary committee member), disclosed financial conflicts of interest, and individual votes. Individuals whose conflicts were so significant as to preclude their participation in meetings were excluded. Participant data, which are publicly available from advisory committee meetings, also include additional details about an individual such as their institutional affiliation. An example of a publicly available panel roster is included as Appendix B.2 and an example of a publicly available panel agenda is included as Appendix B.3.

For every meeting, each individual's voting record (a binary indicator of voting in favor or against a new product) is observed. Additionally, the meeting outcome – i.e. the majority recommendation from the entire voting panel – is known. Other product-specific outcomes (to the extent that they occur) are also observable; for example, whether or not a product was recalled. The current version of this paper presents results exploring the association between financial conflicts of interest and voting patterns on the panels observed. Future versions of the paper will present results that explore the overall relationships between observable characteristics of individuals and their voting behaviors and the influence of other professional relationships among panel participants on voting behavior.

Table 2.1 summarizes the total number of drug (CDER) and device (CDRH) meetings observed in each year of the data from the FDA. Table 2.2 shows the distribution of panel members and their observable characteristics across those meetings: of the 186 meetings in

[12] I am grateful to Clark Nardinella, Economics Staff Director, for helping me to secure access to the data.

Table 2.1: *Summary of Meetings in Sample*

Meetings By Year		
Year	Drugs	Devices
2001	20	19
2002	18	16
2003	14	14
2004	7	14
2005	15	9
2006	9	9
2007	10	9
2008(Q1)	3	0

Table 2.2: *Meeting Sample Summary Statistics*

	All	Drugs	Devices
Total Product Meetings	186	96	90
Total Panels (Focal Areas)	26	13	13
Av. Voting Members on Panel	11.90	14.25	9.39
...of which, Permanent	5.42	5.33	5.52
...of which, Temporary	6.47	8.92	3.87
Meetings with Any Conflict	79.35%	71.60%	84.14%
Members with Any Conflict	28.69%	33.33%	21.18%
...of which, Index Conflict	5.20%	5.26%	5.09%
...of which, Competitor Conflict	17.44%	24.05%	6.75%
...of which, Other Conflict	2.80%	0.66%	6.27%

the data,[13] just over half (96) consider drugs and the remainder (90) consider devices. These meetings included 13 drug and 13 device panels (medical specialty areas).

Table 2.2 shows that the average number of voting members on drug panels (14.25) was larger than the average number of voting members on device panels (9.39). The average fraction of meetings with *any* conflicted member was higher among device panels (84.14%) than among drug panels (71.60%), however, the percentage of individuals with "any conflict"

[13]Data were also collected on a small number of meetings about new biologics, but there were too few meetings to be considered in this study for reasons of statistical power; future versions of this paper may try to re-incorporate meetings about biologics

was actually lower among among device panels (21.18%) than among drug panels (33.33%).

Device and drug panel members were similarly likely to have "index conflict" – a financial relationship with the sponsoring firm whose product is under consideration (5.09 and 5.26 percent respectively) – but drug panel members were far more likely to have "competitor conflict" – a financial relationship with a firm that is considered a competitor of the sponsoring firm – with 24.05% of drug panel members reporting competitor conflict, but only 6.75% of device panel members reporting competitor conflict. Device panel members were more likely than drug panel member to have "other" forms of financial conflicts of interests (6.27% vs. 0.66%) – i.e. financial relationships with industry, but not with the firm sponsoring the product in question or one of its direct competitors.

2.5 Estimation

I begin by looking at meeting-level data in greater detail. One implication of the conceptual framework section above is that meetings with unanimous voting behavior will not be helpful in identifying bias. Summarizing within-meeting votes, I find that 23.5% of all meetings resulted in a unanimous vote in favor of approval while 8.9% of meetings resulted in a unanimous vote against the approval of a new product. Thus nearly 1/3 of the total meetings in the data do not provide directly usable information about expert bias and its influence.

The fraction of in-favor meetings are similar for drugs and devices, but the incidence of unanimous votes against a new product is over twice as likely in drug meetings vs. device meetings (11.2% vs. 5.2%; p-value <0.01). Data on unanimous voting frequencies is presented in Table 2.3. After excluding votes from unanimous meetings, data remains on a set of 1495 votes from 110 non-unanimous meetings (48 drug meetings and 62 device meetings). Figure 1.1a presents the frequency of "in favor" votes for all non-unanimous meetings over the years observed. Figures 1.1b and 1.1c present the same results for drug and device meetings respectively.

Figure 2.1: *Non-Unanmous Meetings*

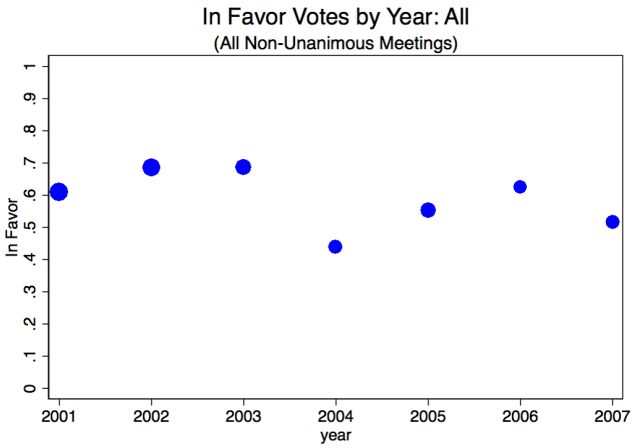


Figure 1b

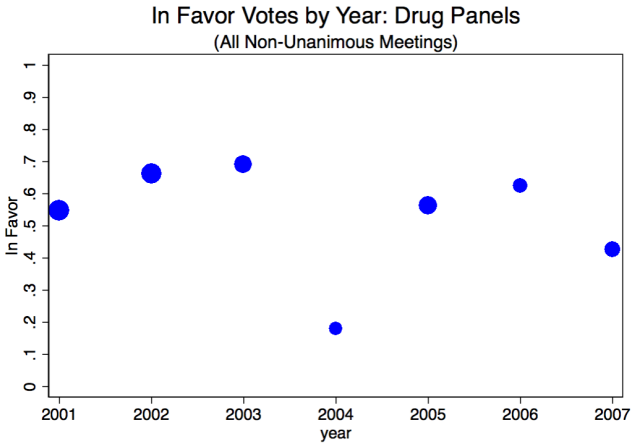


Figure 1c

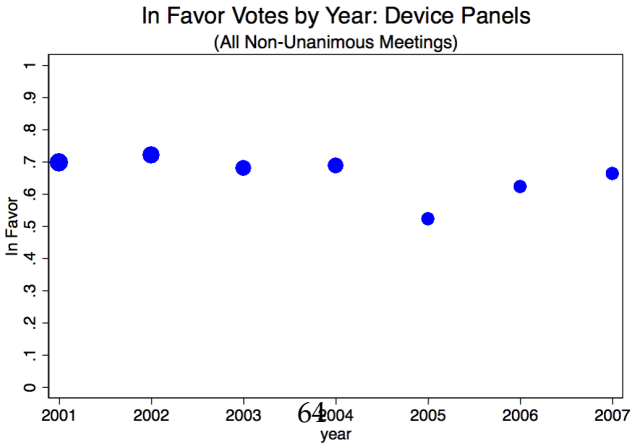


Table 2.3: *Summary of Unanimous Meetings*

	All Meetings	Drug Meetings	Device Meetings	P-value (Drug $\neq$ Device)
Unanimous In Favor	23.5%	22.8%	24.7%	0.29
Unanimous Against	8.9%	11.2%	5.2%	0.00

2.5.1 Is There Evidence of Bias?

On average, are individuals with a financial conflict of interest more likely to vote favorably for a given new product than non-conflicted individuals? Table 2.4 presents marginal effects of probit regressions for all meetings in the data (panel a) and for drug and device meetings separately (panels b and c, respectively).

Analyzing all drug and device meetings together does not reveal any statistically significant relationships between observable characteristics about individuals' conflicts of interest or participation type and their probability of voting favorably for a new product. The same is true when considering drug meetings separately. However, when device meetings are considered alone, a set of interesting and statistically significant associations emerges.

First (column 1 of panel c), individuals with any financial conflict of interest are, on average, 13.6 percent more likely to vote favorably for a new medical device. This relationship is robust to the inclusion of year and panel (specialty) fixed effects (column 2) and if anything, becomes more pronounced with controls for additional covariates: controlling for the year of the meeting and the (medical specialty-focused) advisory panel making the recommendation, individuals with a financial conflict of interest are, on average, 18.5 percent more likely to vote favorably for a new medical device.[14]

I also consider the potential for differences between temporary and permanent members. Column 3 asks if temporary voting members have different probabilities of voting favorably for new products above and beyond the relationships observed for financial conflict of interest and controlling for year and advisory panel fixed effects. Temporary members

[14] As a robustness check, I have also confirmed that the results are highly similar in magnitude and statistical significance when voting data from unanimous meetings are reintegrated into the data set.

Table 2.4: *Patterns in Conflicted Member Voting*

Outcome = Individual "In Favor" Votes			
	(1)	(2)	(3)
(a) All Meetings			
Any conflict	0.025	0.032	0.033
	(0.048)	(0.038)	(0.038)
Temp. member			0.039
			(0.033)
Year Fixed Effects		x	x
Panel Fixed Effects		x	x
N	1495	1495	1495
(b) Drug Meetings Only			
Any conflict	0.012	-0.013	-0.015
	(0.058)	(0.042)	(0.042)
Temp. member			-0.038
			(0.048)
Year Fixed Effects		x	x
Panel Fixed Effects		x	x
N	903	903	903
(c) Device Meetings Only			
Any conflict	0.136*	0.185**	0.181**
	(0.056)	(0.054)	(0.051)
Temp. member			0.123**
			(0.039)
Year Fixed Effects		x	x
Panel Fixed Effects		x	x
N	592	592	592

+<0.10, * p<0.05, ** p<0.01
Reported coefficients are marginal effects at sample means from probit models; standard errors are clustered at the meeting level

Table 2.5: *Voting on Device Panels by Conflict Type*

Outcome = Individual "In Favor" Votes (Device Meetings Only)			
	(1)	(2)	(3)
Index conflict	-0.002		0.010
	(0.101)		(0.102)
Competitor conflict		0.206*	0.207*
		(0.083)	(0.083)
Year Fixed Effects	x	x	x
Panel Fixed Effects	x	x	x
N	592	592	592
+<0.10, * p<0.05, ** p<0.01			
Reported coefficients are marginal effects at sample means from probit models; standard errors are clustered at the meeting level			

do indeed have a higher probability of voting favorably, but this appears to be largely orthogonal to the relationship between conflict and voting behavior. One explanation for this pattern might be that temporary members are called into device panels because they are experts on a specific type of medical technology. Although these members are statistically no more likely to be conflicted than permanent members, they may have greater expertise. Future versions of this paper will explore this possibility.

2.5.2 A Closer Look at the Nature of Conflict of Interest on Device Panels

Before moving on from the simple individual voting model, I decompose financial conflicts of interest on device panels – where they are predictive of voting behavior – into their constituent parts. Table 2.5 considers voting behavior in the sample of device panels analyzed in Table 2.4 but separates out two specific types of financial conflict of interest: index conflict (a financial relationship with the sponsoring firm whose product is under consideration) and competitor conflict (a financial relationship with a firm that is considered a competitor of the sponsoring firm).

In doing so, it becomes clear that the overall pattern in which conflicted individuals

on device panels are more likely to vote favorably for a new product is largely driven by individuals with a *competitor* conflict: these individuals are about 21 percent more likely to vote favorably for a new medical device, controlling for the year of the meeting and the advisory panel making the recommendation (columns 2 and 3).

2.5.3 Evidence of "Paving the Way" for Follow-On Innovation

Why might individuals be inclined to nudge products from a *competitor* of their employer (or a firm in which they hold stock) toward regulatory approval? At first, this finding would seem counterintuitive: individuals have no direct gain (and indeed, if they are stock owners, may be immediately harmed) by voting in favor of competitors' products. However, voting incentives in the medical device setting are dynamic: as I have shown in other work (Stern, 2014), there are significant costs associated with being a "first mover" in a new medical device product market. One corollary of this first mover regulatory disadvantage is that individuals may be able to "pave the way" for subsequent entrants – e.g. from their own firm – by facilitating the approval of first or early entrants into a new product category.[15]

I look for evidence consistent with this type of "paving the way" behavior in Table 2.6. Here, the coefficients of interest are the interaction effects between having a competitor conflict and some measure of a product's "newness" within a category. Unfortunately only about half of the device meetings could be directly linked to information about entry order based on the product names provided by ERG. Future versions of this paper will aim to capture a larger subsample of meetings.

Column 1 reproduces the results on the direct observed association between competitor conflict and in favor voting for the subsample of devices for which entry order information was available. While the result is only statistically significant at the 10% level (likely a result

[15] Another explanation is that some individuals simply may be more optimistic about the value of medical technology, a trait that would lead them to be both more likely to take industry consulting jobs as well as more likely to vote in favor of the approval of new products. Similarly, the very act of taking money from industry might make individuals more optimistic about other medical technologies, making them more likely to vote in favor of the approval of new products. In either of these cases, the observed conflicts and voting behavior would be endogenously related to one another and/or an omitted variable. While I cannot rule such an explanation out, I can look for evidence of an alternative scenario.

Table 2.6: *Do Conflicted Individuals "Pave the Way"?*

Outcome = Individual "In Favor" Votes (Device Meetings Only)			
	(1)	(2)	(3)
Competitor conflict	0.170+	0.142	0.307*
	(0.097)	(0.109)	(0.154)
First		-0.044	
		(0.074)	
First * Comp. Conflict		0.047	
		(0.202)	
Order			0.017+
			(0.010)
Order * Comp. Conflict			-0.043+
			(0.023)
Year Fixed Effects	x	x	x
Panel Fixed Effects	x	x	x
N	289	289	289

+<0.10, * p<0.05, ** p<0.01
Reported coefficients are marginal effects at sample means from probit models; standard errors are clustered at the meeting level

of the decreased sample size), I estimate a coefficient that is similar to those estimated in Tables 2.4 and 2.5: a competitor conflict is associated with a 17% greater propensity to vote in favor of a new product, controlling for year and panel-level fixed effects.

Columns 2 and 3 look for preliminary evidence of "paving the way": column 2 includes a binary indicator for whether or not a product was the first to be approved in a given category and shows that while, on average, first products are about 4.4% less likely to get in favor votes, those who have a competitor conflict are 4.7% more likely to vote in favor of first products, although neither of these results is statistically significant. Column 3 uses a direct (integer) measure of entry order into a product category and finds that later products are, on average, 1.7% more likely to garner in favor votes ($p<.10$), however earlier products are, on average, 4.3% more likely to garner in favor votes from conflicted individuals ($p<.10$). These results are also consistent with industry insiders "paving the way" for subsequent entrants in more novel product categories, although the collection additional data will help to establish a definitive pattern.

2.5.4 Peer Effects and Voting Behavior

The results presented so far do not account for potential peer effects in voting behavior. This section introduces a few types of potential peer effects and presents their associations with observed voting behaviors in reduced form models. The potential peer effects added here are a) the fraction of in favor votes at the meeting level, b) the fraction of in favor votes at the meeting level for all individuals except the reference individual (i.e. a number is calculated for each voter as the average of her panel without her vote included), and c) the fraction of in favor votes among conflicted panelists (only) at the meeting level. The results of regressions accounting for these potential sources of peer influence and interacting each with an indicator for financial conflict of interest are presented in Table 2.7.

An important note with respect to Table 2.7 is that the coefficients presented were recovered from linear probability models for ease of interpretation. So far all regression output tables have presented marginal effects from probit regressions. For the models

Table 2.7: *Preliminary Analysis of Peer Effects in Voting Behavior*

(A)			
Any conflict	0.180**	0.104*	-0.131*
	(0.051)	(0.045)	(0.054)
Meeting Vote In Favor (Fraction)		0.983**	0.924**
		(0.011)	(0.018)
Conflict * Meeting Vote In Favor			0.378**
			(0.076)
Year Fixed Effects	x	x	x
Panel Fixed Effects	x	x	x
N	592	592	592
(B)			
Any conflict	0.180**	0.142**	-0.187*
	(0.051)	(0.049)	(0.081)
Meeting Vote In Favor (Fraction) for All $j \neq i$		0.607**	0.530**
		(0.074)	(0.077)
Conflict * Meeting Vote In Favor for All $j \neq i$			0.537**
			(0.143)
Year Fixed Effects	x	x	x
Panel Fixed Effects	x	x	x
N	592	592	592
(C)			
Any conflict	0.180**	0.133**	-0.225**
	(0.051)	(0.047)	(0.045)
Vote In Favor (Fraction) among Conflicted Panelists		0.609**	0.506**
		(0.046)	(0.053)
Conflict * Vote In Favor among Conflicted Panelists			0.518**
			(0.052)
Year Fixed Effects	x	x	x
Panel Fixed Effects	x	x	x
N	592	592	592

+<0.10, * p<0.05, ** p<0.01
Reported coefficients are from a linear probability model with standard errors clustered at the meeting level

presented in Table 2.7, results from probit regressions using the same dependent and independent variables are of the same sign and statistical significance in all cases, but for this exercise, I focus on the linear probability model results, for ease of comparability across specifications and exercises involving the summing of coefficients.

The models in Table 2.7 consider the relationships between each of the additional factors presented and an individual's probability of voting in favor of a new product. Panel A considers the peer effects of a higher/lower fractional vote in favor among all panelists at a meeting. Panel B considers the fractional vote in favor among all panelists except the reference individual at a meeting; this is likely a better measure because it is not mechanically biased toward finding a positive result and the one I focus on in interpretation. Panel C considers the average vote in favor among all conflicted panelists at a meeting.

While absolute probabilities are difficult to calculate without adding in averaged year and panel fixed effects as well as a constant, the differences between conflicted and unconflicted individuals *ceteris paribus* can be calculated easily algebraically. Coefficients in a linear probability model are interpreted as the percentage point change in y associated with a one unit change in x. Thus, the difference between conflicted and unconflicted individuals is estimated by summing the coefficient on "any conflict" and the coefficient on the interaction between "any conflict" and the specific meeting measure of interest, multiplied by, e.g., the meeting average vote, or a 1 standard deviation change in the meeting-level vote.

For example, in panel B, the difference in probabilities between conflicted and unconflicted individuals at the sample average for the fraction of in favor votes would be calculated as -0.187 (the coefficient on "any conflict") + 0.537 (the coefficient on "any conflict * meeting average vote in favor for all $j \neq i$") * 0.601 (the average meeting vote in favor for all $j \neq i$). That is, at the meeting level average of in favor votes, conflicted individuals are 13.6 percentage points more likely to vote favorably for a new product than unconflicted individuals. This is notable because the implied difference between conflicted and unconflicted panelists voting propensities with peer effects is actually smaller in magnitude than the naive estimate without peer effects in column 1 (13.6 vs. 18 percentage points).

However, a further implication of the positive coefficient on the interaction term is that in a setting with very high rates of other panelists voting in favor, the difference between conflicted and unconflicted individuals' in favor voting propensities will increase. For example using the estimated coefficients from Panel B again, we can calculate that, at just 1 standard deviation above the mean fraction of co-panelists' in favor votes, a conflicted individual is 28.1 percentage points more likely to vote favorably for a new product than an unconflicted individual. Thus allowing for peer effects suggests that the bias associated with a financial conflict of interest will vary positively with the fraction of other panelists voting favorably for a new product. A corollary to this, of course, is that the difference between conflicted and unconflicted individuals' voting behaviors is expected to decrease as the fraction of in favor votes among other panelists shrinks; at just one standard deviation below the mean fraction of other panelists' in favor votes, the estimated difference between conflicted and unconflicted individuals shrinks to roughly zero.

Another interesting fact implied by Table 2.7 is that the fraction of in favor votes among other panelists is also positively predictive of how unconflicted individuals vote (panel B). This is consistent with peer effects that lead to "herding behavior" – even for individuals without a financial conflict of interest. Moreover, panel C implies that a greater fraction of in favor votes among conflicted panelists (who represent just over 16 percent of the total voters) is associated with a higher probability of in favor voting among unconflicted individuals as well.

The data suggest that financially conflicted individuals are likely to have a bias toward in favor voting and the size of that bias is statistically increasing in the fraction of in favor votes on the panel itself and separately, among conflicted individuals only. Moreover, although these are only reduced form estimates, they are generally supportive of a meaningful relationship between panel composition and individual voting behaviors, a framework that has not yet been introduced into research on conflict of interest and voting. Lastly, I note that Table 2.7 uses data on both temporary and permanent members, but coefficients of similar magnitude and statistical significance are estimated when using the (smaller) sample

of permanent members only.

A policy question raised by these estimation exercises is then: how often might bias impact meeting outcomes? Extensions of this preliminary work will focus on simulating meeting outcomes if the same individuals were to sit on panels with differing compositions of colleagues.

2.6 Conclusion and Next Steps

In this preliminary analysis, I have explored the relationship between declared financial conflicts of interest and individuals' voting behaviors on expert advisory committees at the FDA. While I do not find any evidence that financial conflicts of interest lead to more favorable voting in the setting of drug meetings, I do find evidence that they are associated with more favorable voting in the setting of device meetings. The differences between conflicted an unconflicted individuals' probabilities of voting in favor of new devices are large: in a simple analysis, I find that conflicted individuals are about 18 percent more likely to vote favorably for a new medical device than their unconflicted peers.

I decompose the conflicted votes on device panels into two sub-types: votes by individuals with "index conflict" and votes by individuals with "competitor conflict." I find that an indicator of "competitor conflict" is what drives the overall pattern between conflict of interest and an increased probability of voting in favor. Why might individuals display a bias for approving devices from which they do not stand to gain. One explanation that follows from my earlier work on device approvals could be that individuals with a relationship with the medical device industry have an incentive to "pave the way" for the approval of new products within a product category, so that subsequent entrants can benefit from lower regulatory barriers to entering those markets subsequently. I present some preliminary regression evidence that is consistent with this explanation, although the statistical significance of the results is limited by working with a very small sample. It will be of great interest to return to this analysis in the future with additional data.

Next I introduce a simple model of peer effects and ask whether the composition of

an individual's peer group on a panel in addition to her own conflict of interest status is predictive of her voting behavior. My preliminary analyses suggest that financially conflicted individuals are likely to have a bias toward in favor voting and the size of that bias is increasing in the fraction of in favor votes among a) other panelists and b) separately, among other conflicted panelists only. Moreover, I find evidence that the fraction of peers' in favor votes or the fraction of conflicted co-panelists' in favor votes is associated with a higher probability of even an unconflicted individual voting favorably for a new medical device. This result is consistent with a notion of "herding behavior" in panel voting.

Although these are only reduced form estimates, they are generally supportive of a meaningful relationship between panel composition and individual voting behaviors, a framework that has not yet been used to understand the relationship between financial conflict of interest and panel voting.

Still, much work remains to be done. As discussed at several points, I am interested in other sources of influence beyond financial conflict of interest. In conversations with former and current medical device panelists, I have collected anecdotes to suggest that several other characteristics of a panelist's peer group may influence how she votes. These are likely to include whether or not she shares a past or present institutional affiliation with any of her panel peers, whether she shares another (e.g. co-authoring) relationship with any of her panel peers, and whether or not any of her peers would be considered academic experts. An important next step will be collecting data on these characteristics in order to a) explore these other sources of influence and b) compare their magnitudes to those associated with bias due to financial conflict of interest in modeling individual voting behavior.

Additionally, I am interested in simulating how conflicts of interest (and other factors) affect panel outcomes in the modified framework. For example, what if each individual I observe in the data were to sit on a panel composed of different peers? Using an empirical framework that accounts for panel composition effects, I will also be able to assess how often financial conflicts of interest affect majority panel votes when there are spillover effects onto other voters.

I also plan to collect several more years worth of data. An important limitation of the current data set is that once I focus analyses on smaller sub-groups (e.g. only medical device meetings and only permanent panelists), the sample sizes become too small to draw conclusions with appropriate statistical power. Ideally I would like to gather multiple decades worth of data for future work. I look forward to continuing this research and receiving feedback on both the preliminary results and proposed framework and next steps.

Chapter 3

Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending[1]

3.1 Introduction

Regional variations in rates of medical treatments are large in the United States and other countries (Skinner et al., 2012). For example, in the U.S. Medicare population over age 65, price-adjusted per-patient Medicare expenditures ranged from under $7,000 to nearly $14,000, with most of the variation unexplained by regional differences in patient illness or poverty.

What drives such variation in treatment and spending? One possibility is patient demand. Many studies of variations have been conducted in environments where all patients have a similar and fairly generous insurance policy,[2] so price differences are unlikely to be large and income differences are unlikely to be very important. Still, heterogeneity in patient

[1]Co-authored with David Cutler, Jonathan Skinner, and David Wennberg

[2]This is generally true in the U.S. Medicare program. The presence of supplemental insurance coverage differs across the country, but most studies do not find that these differences affect utilization by more than a small degree (McClellan and Skinner, 2006).

preferences for care may play a role. In very acute situations, some patients may prefer to try all possible measures, while others may prefer palliation and an out-of-hospital death. If patients with similar preferences are grouped together geographically – for example, if people who value and demand life-prolonging treatments live in areas with world-class interventional physicians – patient preference heterogeneity could lead to regional variation in equilibrium outcomes (Anthony et al., 2010; Mandelblatt et al., 2012).

Another possible source of variation arises from the supply side. "Supplier-induced demand" describes a situation in which a health care provider shifts a patient's demand curve beyond what the patient would want. This would be true in a principle-agent framework (McGuire and Pauly, 1991), if prices are high enough (and income scarce). While physician utilization has been shown to be sensitive to prices (Jacobson et al., 2006, Clemens and Gottlieb, 2012), it would be difficult to explain observed Medicare variations using profit margins alone, since reimbursement rates are set administratively and do not vary greatly across areas.

Variation in desired supply may also result from non-monetary incentives. Physicians could respond to organizational pressure or peer pressure to perform more procedures, even if their current income is no higher as a consequence. Physicians might also have differing beliefs about appropriate treatments, particularly for conditions where there are few professional guidelines (Wennberg et al., 1982). These differences in beliefs may arise because of differences in where physicians received medical training (Epstein and Nicholson, 2009) or their personal experiences with different interventions (Levine-Taub et al., 2011). If this variation is correlated spatially – for example, if more intensive physicians are more likely to hire physicians with similar views – the resulting regional differences in beliefs could explain regional variations in equilibrium spending.

It has proven difficult to estimate separately the impact of physician beliefs, patient preferences, and other factors as they affect equilibrium healthcare outcomes, largely because of challenges in identifying factors that affect only supply or demand (Dranove and Wehner, 1994). We address this problem using "strategic surveys," as in Ameriks et al. (2011), in

which we use detailed survey vignettes to elicit motivation and clinical beliefs of physicians (suppliers), and attitudes and preferences of patients (demanders) as well as intervention-specific preferences from both groups. These responses are then linked to utilization measures at the regional level, which allows us to estimate directly how supply and demand factors affect regional healthcare utilization.

Patient preferences are measured by a survey of Medicare enrollees age 65 and older asking about whether they would want a variety of aggressive care interventions. We focus on the tradeoff between invasive procedures with potential longevity benefits versus palliative care and comfort at the end of life. Physician beliefs are captured using two surveys: one of cardiologists and the second of primary care physicians. Both groups of physicians were presented with vignettes about four elderly individuals with chronic health conditions, and asked how they would manage each one. Based on their responses, we characterize physicians along two non-exclusive dimensions: those who consistently and unambiguously recommended intensive care beyond interventions consistent with current clinical guidelines ("cowboys"), and those who consistently recommended palliative care for the very severely ill ("comforters").

We first use these surveys to examine the importance of patient and physician preferences in explaining regional variations in care and find that physician preferences are significantly more important in statistical models. In some models, we can explain over half of the variation in end-of-life spending across areas by knowing only how a relatively small sample of physicians in an area would treat hypothetical patients. In contrast, patient preferences explain little of the cross-area variation.

We then try to understand what factors are associated with physicians' treatment preferences, relating physicians' views about optimal treatment to questions about malpractice concerns, patient financial arrangements (fraction of Medicaid and capitated patients), and perceived organizational pressures (providing treatment for patients who expected but didn't need it, or doing a procedure because the referring physician expected it). We find that only a small fraction of physicians claim to have made recent decisions as a result

of purely financial considerations. We also find that "pressure to accommodate" either patients' demands (by providing treatments that are not needed) or referring physicians' expectations (doing procedures to keep them happy and meet their expectations) have a modest but significant relationship with physician beliefs about appropriate care. While many physicians report making interventions as a result of malpractice concerns, these responses do not help to explain the residual variation in treatment recommendations.

Ultimately, the largest degree of regional variation appears to be due to differences in physician beliefs about the efficacy of particular therapies. Physicians in our data have starkly different views about how to treat the same patients, and these views are not highly correlated with demographics, background, and practice characteristics, and are often not consistent with professional guidelines for appropriate care. As much as 36 percent of end-of-life Medicare expenditures, and 17 percent of overall Medicare expenditures, are explained by physician beliefs that cannot be justified either by patient preferences or by evidence of clinical effectiveness.

3.2 A Model of Variation in Utilization

We develop a simple model of patient demand and physician supply. The demand side of the model is a standard one: the patient's indirect utility function is a function of out-of-pocket prices (p), income (Y), and preferences for care (η); $V = V(p, Y, \eta)$. Solving this for optimal intensity of care, x, yields x^D. As in McGuire (2011), we assume that x^D is the fully informed patient's demand for the quantity of procedures prior to any demand "inducement."

On the supply side, we assume that physicians seek to maximize the perceived health of their patient, $s(x)$, by appropriate choice of inputs x, subject to patient demand (x^D), financial considerations, and organizational factors. Note that the function $s(x)$ captures both patient survival and patient quality of life, for example as measured by quality-adjusted life years (QALYs).

Individual physicians are assumed to be price-takers (after their networks have negoti-

ated prices with insurance companies), but face a wide range of reimbursement rates from private insurance providers, Medicare, and Medicaid. The model is therefore simpler than models in which hospital groups and physicians jointly determine quantity, quality, and price, (Pauly, 1980) or where physicians exercise market power over patients to provide them with "too much" health care (McGuire, 2011). Following Chandra and Skinner (2012), we write the physician's overall utility as:

$$U = \Psi s(x) + \Omega(W + \pi x - R) - \phi(|x - x^D|) - \varphi(|x - x^O|) \tag{3.1}$$

where Ψ is perceived social value of improving health, Ω is the physician's utility function of own income, comprising her fixed payment W (a salary, for example) net of fixed costs R, and including the incremental "profits" from each additional test or procedure performed, π.[3] The sign of π depends on the type of procedure and the payment system a physician faces.

The third term represents the loss in provider utility arising from the deviation between the quantity of services the provider recommends (x) and what the informed patient demands (x^D). This function could reflect classic supplier-induced demand – from the physician's point of view, x^D is too low relative to the physician's optimal x – or it may reflect the extent to which physicians are acting as the agent of the (possibly misinformed) patient, for example when the patient wants a procedure that the physician does not believe is medically appropriate. The fourth term reflects a parallel influence on physician decision making exerted by organizational factors that do not directly affect financial rewards, such as (physician) peer pressure.

The first-order condition for (1) is:

$$\Psi s' = -\Omega'\pi + \phi' + \varphi' \equiv \lambda \tag{3.2}$$

Physicians then provide care up to the point where the choice of x reflects a balance between the perceived marginal value of health, $\Psi s'(x)$, and factors summarized by λ: (a)

[3]We ignore capacity constraints, such as the supply of hospital or ICU beds.

the incremental change in net income π, weighted by the importance of financial resources Ω', (b) the incremental disutility from moving patient demand away from where it was originally, ϕ', and (c) the incremental disutility from how much the physician's own choice of x deviates from her organization's perceived optimal level of intervention, φ'.

In this model,[4] there are two ways to define "supplier-induced demand." The broadest definition is simply the presence of any equilibrium quantity of care beyond the level of the ex ante preferences of an informed patient, i.e. $x > x^D$. This is still relatively benign; the marginal value of this care may still be positive. More relevant is the sign of $s(x) - s(x^D)$; does the additional care enhance or diminish health outcomes? Supplier-induced demand could more narrowly be defined as $s(x) - s(x^D) \leq 0$; patients gain no improvement in health outcomes and may even experience a decline in health or a significant financial loss. Importantly, both of these definitions leave the question of physician knowledge of inducement beyond clinically appropriate levels ambiguous. That is, a physician with strong (but incorrect) beliefs may over-treat her patients, even in the absence of financial or organizational incentives to do so.

To develop an empirical model, we adopt a simple closed-form solution of the utility function for physician i:[5]

$$U_i = \Psi s_i(x_i) + \omega[W_i + \pi_i x_i - R_i] - \frac{\phi}{2}(x_i - x_i^D)^2 - \frac{\varphi}{2}(x_i - x_i^O)^2 \tag{3.3}$$

Note that ω/Ψ reflects the relative tradeoff between the physician's income and the value of improving patient lives, and thus might be viewed as a measure of "professionalism." The first-order condition is therefore:

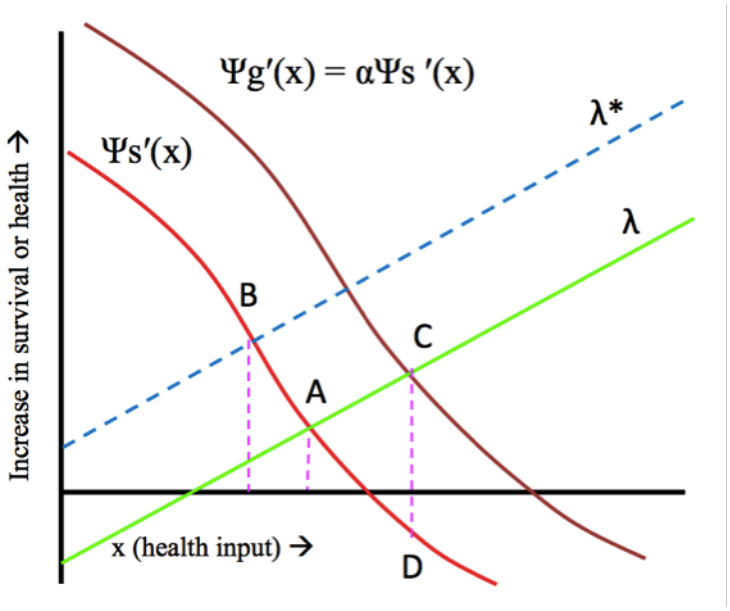
$$\Psi s_i'(x_i) = \lambda \equiv -\omega\pi_i + \phi(x_i - x_i^D) + \varphi(x - x_i^O) \tag{3.4}$$

Figure 3.1 shows $\Psi s_i'(x)$ and λ. Note that λ is linear in x with an intercept equal to

[4] A more general model would account for the patient's ability to leave the physician and seek care from a different physician, as in McGuire (2011).

[5] We are grateful to Pascal St.-Amour for suggesting this approach.

Figure 3.1: *Variations in Equilibrium: Differences in λ and Differences in Actual or Perceived Productivity*



$-(\omega\pi + \phi x_i^D + \varphi x_i^O)$. Note also the key assumption that patients are sorted in order from most appropriate to least appropriate for treatment, thus describing a downward sloping $\Psi s'(x)$ curve. The equilibrium is where $\Psi s'(x) = \lambda$ at point A. A shift in the intercept, which depends on reimbursement rates for procedures π, taste for income ω, regional demand x^D, and organizational or peer effects x^O, would yield a different λ^*, and hence a different utilization rate. But all of these factors affect the intensity of treatments via a movement *along* the marginal benefit curve, $\Psi s'(x)$.

Alternatively, it may be that $s_i'(x)$ differs across physicians – productivity differs, rather than constraints. For example, if $s_i'(x) = \alpha_i s'(x)$, where $s'(x)$ is average physician productivity and α varies across regions, this would be represented as a shift in the marginal benefit curve. Point C in Figure 3.1 corresponds to greater intensity of care than point A and arises naturally when the physician is or just believes she is more productive. For example, heart attack patients experience better outcomes from cardiac interventions in regions with higher rates of revascularization, consistent with a Roy model of occupational sorting (Chandra and Staiger, 2007). Because patients in regions with high intervention rates benefit differentially from these interventions, this scenario does not correspond to the narrow definition of

"supplier-induced demand."

The productivity shifter α may also vary because of "professional uncertainty" – a situation where the physician's perceived α differs from the true α (Wennberg et al., 1982). For example, physicians may be overly optimistic with respect to their ability to perform procedures, leading to expected benefits that exceed actual realized benefits. Baumann et al. (1991) have documented the phenomenon of "macro uncertainty, micro certainty" in which physicians and nurses are sure that their administered treatment benefited a specific patient (micro certainty) even in the absence of a general consensus as to which procedure is more clinically effective (macro uncertainty). Much of the evidence from psychology[6] also argues for overconfidence in one's own ability, leading to a natural bias towards doing more.

To see this in Figure 3.1, suppose the actual benefit is $s'(x)$ but the physician's perceived benefit is $g'(x)$. The equilibrium is point C: the marginal treatment harms the patient, even though the physician believes the opposite. In equilibrium, this supplier behavior would appear consistent with classic supplier-induced demand, but the cause is quite different.

Empirical Specification. To examine these theories empirically, we consider variation in practice at the regional level (for reasons explained below). Taking a first-order Taylor-series approximation of equation (3.4) for region i yields a linear equation that groups equilibrium outcomes into two components, demand factors Z^D and supply factors Z^S:

$$x_i = \bar{x} + Z_i^D + Z_i^S + \varepsilon_i \tag{3.5}$$

The demand-side component is:

$$Z_i^D = \frac{\phi}{M}(x_i^D - x^D) \tag{3.6}$$

where $M = -\Psi s''(\bar{x}) + \phi + \varphi$. This first element of equation (3.6) reflects the higher average demand for health care, multiplied by the extent to which physicians accommodate that demand, ϕ. The supply side component is:

[6]If the patient gets better, the physician gets the credit, but if the patient gets worse, the physician is able to say that she did everything possible (Ransohoff et al., 2002).

$$Z_i^s = \frac{1}{M}[\omega \Delta \pi_i + \pi \Delta \omega_i + \phi(x_i^O - x^O) + \Psi s'(x) \Delta \alpha_i] \tag{3.7}$$

The first term in equation (3.7) reflects how differences in profits in region i vs. the national average $(\Delta\pi)$ affect utilization. The second term reflects the extent to which physicians weigh income more heavily. The third term captures organizational goals in region i relative to national averages $(x_i^O - x^O)$. The final term captures the impact of different physician beliefs about productivity of the treatment $(\Delta\alpha_i)$; this term shifts the marginal productivity curve.[7]

Equation (3.5) can be expanded to capture varying parameter values as well – for example, in some regions physicians may be more responsive to patient demand (a larger ϕ_i). These interactive effects, considered below, reflect the interaction of supply and demand and would magnify the responses here.

3.3 Data and Estimation Strategy

In general, it is difficult to distinguish among demand and supply explanations for treatment variation; even detailed clinical data reveal only a subset of what the physician knows about her patient's health and reveal virtually nothing about non-clinical drivers of patient demand for health care services. Further, patient preferences and physician beliefs about the desirability or appropriateness of different procedures are unknown in ex post clinical data. In studying motives for household saving, Ameriks et al. (2011) implemented "strategic surveys" to identify demand and supply. We follow this approach here, using surveys that ask potential patients about preferences for hypothetical end-of-life choices (that is, x^D before their interaction with the physician), and asking physicians how they would treat a set of hypothetical patients with varying disease severity, as well as questions about their financial and organizational constraints.

[7]Note that these effects are scaled by $\frac{1}{M}$, which depends on $-s''$. If returns to treatment do not decline rapidly, strongly-held physician opinions can lead to highly variable treatment rates (Chandra and Skinner, 2012).

In an ideal world, patient surveys would be matched with surveys from their respective physicians. Because our data do not match physicians with their own patients, we instead match supply and demand at the area level using Hospital Referral Regions (HRRs).[8] In equation (3.5), we therefore define x to be a regional average spending measure. Our primary measure is the natural logarithm of risk- and price-adjusted Medicare expenditures in the last two years of life. We also consider several other measures of utilization such as one-year risk- and price-adjusted expenditures for Medicare enrollees for hip fracture, and overall price-adjusted Medicare expenditures.

Our first estimation, based on Equation 3.5, asks whether area-level supply or demand factors can better explain actual regional expenditures. Our second set of estimates then seek to understand why physicians hold the beliefs they do (Equation 3.7). For the latter, we relate individual physician vignette responses to those physicians' financial and organizational incentives. We interpret the component of vignette responses that cannot be explained by demographic, organizational or financial incentives as reflecting primary physician beliefs (e.g., a shift in perceived marginal treatment curve from $\Psi s'(x)$ to $\Psi g'(x)$). We describe each survey in turn.

Patient Survey. The survey sampling frame was all Medicare beneficiaries in the 20% denominator file who were age 65 or older on July 1, 2003 (Barnato et al., 2009). A random sample of 4,000 individuals was drawn; the response rate was 65%. We limit the final sample to respondents who provided all variables of interest, leaving a total of 1,413 Medicare beneficiary surveys. The final sample of respondents reside in 64 HRRs (an average of over 22 patients per HRR), all of which have sufficient physician observations to be included in the empirical model.

We use responses to 5 survey questions asking patients about their likelihood of wanting unnecessary tests or cardiologist referrals in the case of new chest pain as well as preferences for comfort vs. intensive life-prolonging interventions in an end of life situation. The exact

[8]These HRRs are defined in the *Dartmouth Atlas of Health Care*, which divides the United States into 306 HRRs. Spending measures are based on area of patient residence, not where treatment is actually received.

wording of these vignettes is shown in Panel I of the first Appendix to this chapter. Since the questions patients respond to are hypothetical and typically describe scenarios that have not yet happened, we think of them as x^D, or preferences not affected by physician advice. Importantly, since these patients have not yet faced the tradeoffs described in the survey in the end of life scenario, their views are unlikely to be colored by their physicians' opinions.

Two of the questions relate to unnecessary care, asking people if they would like a test or cardiac referral even if their primary care physician did not think they needed one (Table 3.1).[9] Overall, 73 percent of patients wanted such a test and 56 percent wanted a cardiac referral. However, there is wide variation across regions in averages responses to these question. Figure 3.2 shows density plots of of patient preferences for the main questions in the patient survey for the 64 HRRs considered (weighted by the number of patients per HRR). Simulated distributions based on 1000 bootstrap samples with replacement were used to test the null hypothesis of no geographic correlation. While some of the observed variation is likely due to small sample sizes within regions, we tested for the null of no regional variation by bootstrapping the distribution of area-level averages of all key variables, assuming individuals were randomly assigned to areas. P-values are reported in the last column of Table 3.1.

Three other patient questions, grouped into two binary indicators, measure preferences for end-of-life care. One reflects patients' desire for aggressive care at the end of life: whether they would want to be put on a respirator if it would extend their life for either a week (one question) or a month (another question). The second question asked, if the patient reached a point at which they were feeling bad all of the time, would they want drugs to make them feel better, even if those drugs might shorten their life. In each case, there is statistically significant variation across HRRs (Table 3.1).

Patients' preferences are generally correlated across questions. For example, the correlation coefficient between wanting an unneeded cardiac referral and wanting an unnecessary

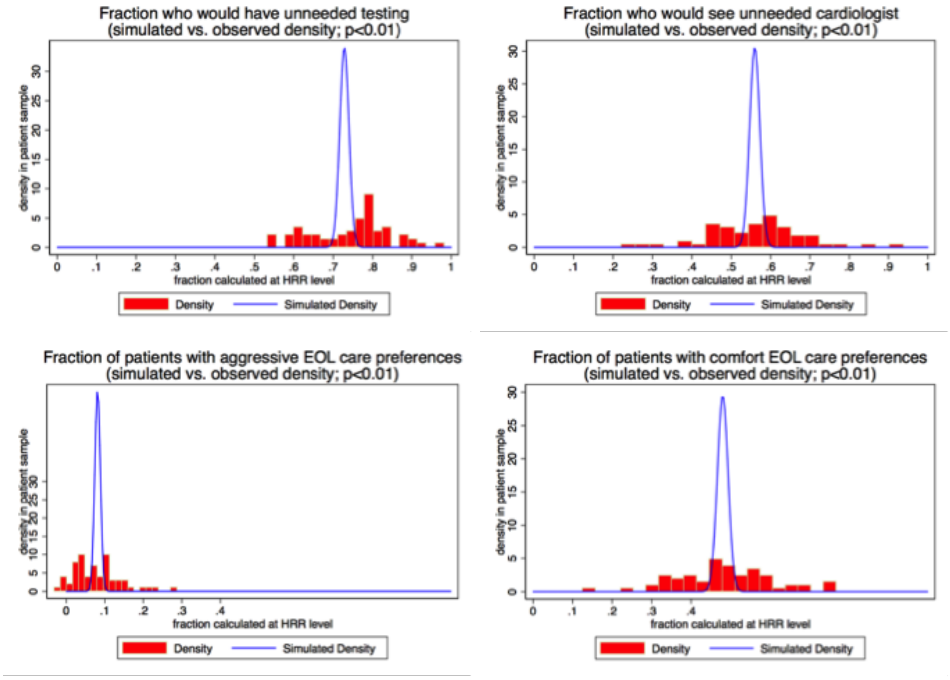
[9]This question captures pure patient demand independent of what the physician wants. Note, however, that patients could still answer they would not seek an additional referral if they were unwilling to disagree with their physician.

Table 3.1: *Primary Variables and Sample Distribution*

Variable	Mean	Individual SD	Area Average SD	p-value
Spending and Utilization				
2-Year End-of-Life Spending	$56,219	-	$10,715	-
6-Month End-of-Life Spending	$14,272	-	$2,660	-
Total Per Patient Spending	$7,837	-	$1,032	-
Hip Fracture Patient Spending	$52,574	-	$4,996	-
Patient Variables				
Have Unneeded Tests	73%	44%	10%	<0.01
See Unneeded Cardiologist	56%	50%	10%	<0.01
Aggressive Patient Preferences Ratio	8%	27%	5%	<0.01
Comfort Patient Preferences Ratio	48%	50%	12%	<0.01
Primary Care Physician Variables				
Cowboy Ratio	19%	39%	19%	<0.01
Comforter Ratio	44%	50%	20%	<0.01
Follow-Up Low	9%	28%	11%	<0.01
Follow-Up High	4%	19%	7%	<0.01
Cardiologist variables				
Cowboy Ratio	27%	45%	19%	<0.01
Comforter Ratio	29%	45%	20%	<0.01
Follow-Up Low	0%	4%	3%	0.09
Follow-Up High	23%	44%	21%	<0.01
Organizational and Financial Variables				
Fraction Capitated Patients	16%	25%	-	-
Fraction Medicaid Patients	10%	13%	-	-
Weekly Patient Days	3.1	1.5	-	-
Physician Age	57.5	9.8	-	-
Board Certified	89%	31%	-	-
Cardiologists per 100k	6.7	1.90	-	-
Responds to Referrer Expectations	10%	30%	-	-
Responds to Colleague Expectations	41%	49%	-	-
Responds to Patient Expectations	59%	49%	-	-
Responds to Malpractice Concerns	43%	49%	-	-
Responds to Practice Financial Incentives	32%	46%	-	-

Note: The table shows means for the sample living or practicing in one of the 64 HRRs with at least 3 cardiologists and 2 primary care physicians. The area average standard deviation is weighted by the number of observations in the HRR. The p-value in the last column is for the null hypothesis of no excess variance across areas. The p-value is taken from a bootstrap of patient or physician responses across areas. For each of 1,000 simulations, we draw patients or providers randomly (with replacement) and calculate the simulated area average and the standard deviation of that area average. The empirical distribution of the standard deviation of the area average is used to form the p-value for the actual area average.

Figure 3.2: *Distributions of Patient Preferences vs. Simulated Distributions (based on 1000 bootstrap samples with replacement)*



test is 0.43 ($p < .01$). But other comparisons point to very modest associations, for example a -0.02 correlation coefficient between wanting palliative care and wanting to be on a respirator at the end of life.

Since survey responses may vary systematically by demographic covariates such as race and ethnicity; we create demographically-adjusted HRR-level measures of patient preferences by adjusting all responses for observed patient characteristics (race, age and sex).[10]

Physician Surveys. A total of 999 cardiologists were randomly selected to receive the survey. Of these, 614 cardiologists responded, for a response rate of 61%. Seventeen physicians did not self-identify as (primarily) cardiologists, and 88 physicians were missing

[10]One early reader suggested that patient preferences for aggressive vs. palliative care and for unneeded tests and/or specialist visits may evolve as patients age. We tested for this by comparing average preferences among individuals for patients that were on average "older" (age > sample mean) or "very old" (age > sample mean + 1 standard deviation) and did not find statistically significant differences between patient preferences in older or very old sub-groups.

crucial information such as practice type, or practiced in HRRs with too few respondents to include in the analysis, leaving us a final sample of 509 cardiologists. These cardiologists practice in 64 HRRs, all of which have 3 or more cardiologists represented in the survey.

The primary care physician (PCP) responses come from a parallel survey of PCPs (family practice, internal medicine, or internal medicine/family practice). A total of 1,333 primary care physicians were randomly selected to receive the survey and the response rate was 73%. A total of 840 PCPs had complete responses to the survey and practiced in HRRs with enough local patient and physician respondents to include in the analysis.

Both sets of physicians were asked about a number of clinical vignettes, discussed in the next section, as well as a variety of characteristics of their practices. Two measures of financial circumstances are reported in Table 3.1 for all physicians: the share of patients for whom they are reimbursed on a capitated basis (on average, 16 percent), and the share of a physician's patients on Medicaid (10 percent), with both factors generally associated with lower marginal reimbursement.

A second set of questions asks about characteristics of the physician and her practice. Twenty-nine percent are in small practices (solo or 2-person), 60 percent are in single or multi-specialty group practices, and 11 percent are in HMOs or hospital-based practices. We also observe a number of characteristics about the physician, including age, gender, whether she is board certified, and the number of weekly patient days practiced.

Third, the survey asks about a physician's actual responsiveness to external incentives over the past year, including how frequently, if ever, in the past 12 months she has intervened for non-clinical reasons. We create a set of binary variables that indicates whether a physician responded to each set of incentives at least "sometimes" (i.e. "sometimes" or "frequently") over the past year. Ten percent of cardiologists reported that they had sometimes or frequently performed a cardiac catheterization because of the expectations of the referring physician and 41 percent of all physicians reported doing so because of a colleague's expectations (Table 3.1).

Like patient surveys, we recognize that physician survey responses may vary system-

atically by demographic covariates such as race and ethnicity. For those exercises that require aggregation of multiple physician surveys, we create demographically-adjusted HRR-level measures of physician beliefs by adjusting all responses for observed physician characteristics (race, age and sex).

Medicare Utilization Data. We match the survey responses with expenditure data by HRR. Our primary measure is Medicare expenditures in the last two years of life for enrollees over age 65 with a number of fatal illnesses.[11] All HRR-level measures are adjusted for age, sex, race, differences in Medicare reimbursement rates and the type of disease (including an indicator for multiple diseases). This measure implicitly adjusts for differences across regions in health status; an individual with renal failure who subsequently dies is likely to be in similar (poor) health regardless of whether she lives in West Virginia or Oregon.[12] End-of-life measures are commonly used to instrument for health care intensity, (e.g., Fisher et al., 2003), are highly correlated with other medical expenditure measures such as one-year expenditures following a heart attack (Skinner et al., 2010), and do not appear sensitive to the inclusion of additional individual-level risk-adjusters (Kelley, et al., 2012). In sensitivity analysis, we consider price-adjusted Medicare expenditures for all fee-for-service enrollees age 65 and above, and a "forward looking" measure of one-year expenditures following hospital admission for a different severe condition, hip fracture. The HRR-level price-adjusted expenditures for the hip fracture cohort are adjusted for age, sex, race, comorbid conditions at admission, and the hierarchical condition categories (HCC) risk-adjustment index for the 6 months prior to admission. We focus on the 64 HRRs in the combined sample with a minimum of 3 cardiologists (average =5.4) and 2 primary care physicians (average = 7.9) surveyed. Among patients, we observe an average of 22 respondents per

[11]These include congestive heart failure, cancer/leukemia, chronic pulmonary disease, coronary artery disease, peripheral vascular disease, severe chronic liver disease, diabetes with end organ damage, chronic renal failure, and dementia.

[12]If more intensive spending saves lives, then in regions with more intensive spending, fewer die, leading to potential biases in the end-of-life measure (Bach et al., 2004). However, given conventional estimates of cost-effectiveness in end-of-life spending, the magnitude of the bias would be small.

HRR.[13]

Clinical Vignettes from the Physician Surveys. Since the clinical vignettes are crucial for our analysis, we describe them in some detail. We note first the obvious: responses to the vignette may not exactly reflect what physicians actually do in practice and because we are unable to link physician responses to those physicians' claims, we cannot test this in the context of this data set. Empirical evidence, however, strongly indicates that clinical vignettes closely predict how physicians actually intervene (Peabody et al., 2004; Mandelblatt et al., 2012; Dresselhaus et al., 2004). Additional tests done on our data confirm that HRR level rates of percutaneous coronary intervention (PCI) in Medicare patients in the year of the survey are correlated with local cardiologists' survey responses, additional evidence that survey vignettes predict actual physician behavior.

Moreover, and importantly for the contribution of this paper, the vignettes have far more detail than the claims data because they yield probabilistic assessments of multiple counterfactual interventions. In claims data, the relative probabilities of counterfactual interventions are unknown because counterfactual interventions are necessarily unobserved. In this respect, among others, the vignette-based survey data we consider are far richer than claims data.

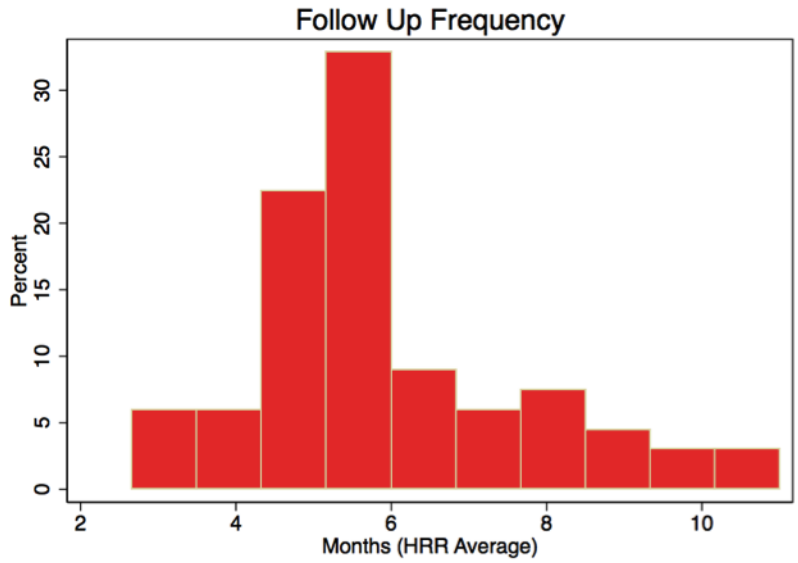
We assume that the physician's responses to the vignettes are "all in" measures (Z^S, as in equation 3.7), reflecting physician beliefs as well as the variety of financial, organizational, and capacity-related constraints physicians face. Alternatively, one could interpret the physician's responses to the vignettes as a pure reflection of beliefs (for example, how one might answer for qualifying boards), and not as representative of the day-to-day realities of their practice. We tested this alternative explanation by including the organizational and financial variables in our estimation equations in addition to the vignette estimates. This

[13]Early readers of this paper wondered how to compare measurement error in the patient responses, which are likely to only capture individual patients' preferences, versus physician responses, which likely capture physicians' experiences with hundreds of their patients. While only partially addresses this concern, we also note that our primary results are robust to focusing only on regions in the top two quartiles of per-HRR patient observations, suggesting that findings are very similar when focusing on those regions with relatively more patients represented.

did not appreciably increase the explanatory power of these equations.[14]

The detailed clinical vignette questions are shown in the first appendix to this chapter (Panel II) and summary statistics are presented in Table 3.1. We begin with the vignette for Patient A, which asks how frequently the physician would schedule routine follow-up visits for patients with stable angina whose symptoms and cardiac risk factors are well controlled on current medical therapy (cardiologists) or patients with hypertension (primary care physicians). The response is unbounded, and expressed in months. Answers ranged from 1 month to 24 months in practice. Figure 3.3 presents a HRR-level histogram of averages from the cardiology survey for all 64 HRRs studied.

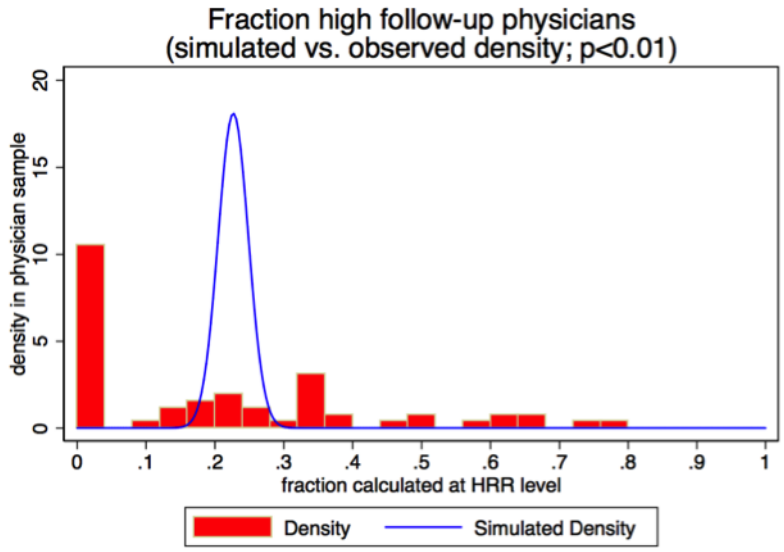
Figure 3.3: *Distribution of Length of Time before Next Visit for Patient with Well-Controlled Angina (Cardiologist HRR-Level Averages)*



[14]One might argue that physicians in regions with, e.g. most of their low-income patients in poor health may "fill in" missing characteristics of the vignettes. This could make such physicians more likely to recommend intensive care, meaning that imperfectly risk-adjusted Medicare expenditures would be spuriously correlated with more intensive vignette recommendations. Alternatively, such physicians may also be less likely to recommend intensive medical or surgical treatments, since outcomes are dependent on coordinated follow-up care that may not be available to patients living in low-income neighborhoods. While we cannot rule out either potential source of bias, we note that in a study of medical students responding to clinical vignettes, individuals' clinical assessments were not associated with patient race or occupation and no association was found between implicit preferences and the assessments (Haider et. al., 2011). Lastly, we note that to the extent that physicians answer questions according to "textbook" answers, the responses we record from doctors could be a lower bound on true variation in physician beliefs.

How do these responses correspond to guidelines for managing chronic stable angina? While diagnosis and management of coronary artery disease (the cause of angina) is the most common clinical issue faced by cardiologists on a day-to-day basis, there are no hard data to support any recommendation. The 2005 American College of Cardiology/American Heart Association [ACC/AHA] guidelines (Hunt et al., 2005) – what most cardiologists would have considered the "Bible" in cardiology at the time the survey was fielded – were very imprecise: they recommended follow-up every 4-12 months. However, even with these broad recommendations, we find that over one fifth (23%) of cardiologists in the sample recommend follow-up visits more frequently than every 4 months. These physicians were geographically clustered in a subset of HRRs ($p < .01$ in a test of the null of no geographic correlation) and the distribution of high follow-up cardiologists across HRRs is shown in Figure 3.4.

Figure 3.4: *Distribution of High Follow-Up Cardiologists and Geographic Correlation (HRR-Level Averages)*



The equivalent follow-up measure for primary care physicians is for a patient with well-controlled hypertension. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (U.S. Department of Health and Human Services, 2004), which would have been the most current guideline

recommendation at the time, suggests follow up every 3-6 months based on expert opinion.

We define a "high follow-up" physician as one who recommends follow-up visits more frequently than clinical guidelines would suggest and a "low follow-up" physician as one who recommends follow-up visits less frequently than clinical guidelines would suggest. By this definition, less than 1 percent of cardiologists and 9 percent of PCPs in our data are classified as "low follow-up" physicians while 23 percent of cardiologists and 9 percent of PCPs are classified as "high follow-up" physicians.

Office visits are not a large component of physicians' incomes (or overall Medicare expenditures). Thus any correlation between the frequency of follow-up visits and overall expenditures would most likely be because frequent office visits are also associated with additional highly remunerated tests and interventions (such as echocardiography, stress imaging studies, and so forth) that further set in motion the "diagnostic-therapeutic cascade," resulting in subsequent diagnostic tests, treatments, and follow-up visits (Lucas, et al., 2008). Thus the next two vignettes focus on patients with heart failure, a much more expensive setting. Heart failure is also natural to ask about because it is common, the disease is chronic, prognosis is poor, and treatment is expensive.

Vignettes for both Patients B and C ask questions about the treatment of Class IV heart failure, the most severe classification and one in which patients have symptoms at rest. In both scenarios the vignette patient is on maximal (presumably optimal) medications, and neither patient is a candidate for revascularization: Patient B has already had a coronary stent placed without symptom change, and Patient C is explicitly noted to not be a candidate for this procedure. The key differences between the two scenarios are patients' ages (75 for patient B, 85 for Patient C), the presence of asymptomatic non-sustained ventricular tachycardia in Patient B, and severe symptoms that resolve partially with increased oxygen in Patient C.

Cardiologists in the survey were asked about various interventions as well as palliative care for each of these patients. For patient B, they were given five choices: three intensive treatments (repeat angiography; implantable cardiac defibrillator [ICD] placement, and

pacemaker insertion), one involving medication (anti arrhythmic therapy), and palliative care. Patient C also has three intensive options (admit to the ICU/CCU, placement of a coronary artery catheter, and pacemaker insertion), two less aggressive options (admit to the hospital (but not the ICU/CCU) for diuresis, and send home on increased oxygen and diuretics) and palliative care. In each case, cardiologists ranked their likelihood of recommending each intervention separately on a 5-interval range from "never" to "always / almost always." Physicians could indicate strong or weak likelihood of recommending multiple options, for example, a physician might "frequently" recommend both palliative care and an intervention.

We start with the obvious: regardless of the religious, political or moral persuasion of the cardiologist, these two men deserve a frank conversation about their prognosis and an ascertainment of their preferences for end-of-life care. One-year mortality for those with Class IV heart failure is nearly 50 percent. If compliant with the guidelines, therefore, every one of the cardiologists should have answered "always/almost always," or at least "most of the time," to initiating or continuing discussions about palliative care.[15]

Studies have shown that patients, physicians and family members are often not "on the same page" when it comes to advanced directive planning (Connors, et al., 1995), and is reflected in the survey data: for Patient B, only 30 percent of cardiologists responded that they would initiate or continue discussions about palliative care "most of the time" or "always/almost always." For Patient C, 43 percent of cardiologists and 50 percent of primary care physicians were likely to recommend this course of action "most of the time" or "always/almost always." In both cases, physicians' recommendations fall far short of clinical guidelines, which would suggest that these discussions are always appropriate for such severely ill patients. We define our second index of physicians to reflect physicians' likelihood of recommending palliative care. We classify the doctor as a "comforter" if the

[15]According to the AHA-ACC directives, "Patient and family education about options for formulating and implementing advance directives and the role of palliative and hospice care services with reevaluation for changing clinical status is recommended for patients with HF [heart failure] at the end of life." (Hunt et al., 2005, p. e206)

physician would discuss palliative care with the patient "always / almost always" for both Patients B and C (among cardiologists) or for patient C (among primary care physicians, who did not have Patient B's vignette in their survey). In our final sample, 29 percent of cardiologists and 44 percent of primary care physicians met this definition of a comforter.

We now turn to more controversial aspects of patient management. The language in the vignettes was carefully constructed to relate to the contemporaneous clinical guidelines. Several key aspects of Patient B rule out both the ICD and pacemaker insertion[16] and indeed the ACC-AHA guidelines explicitly recommend against the use of an ICD for Class IV patients potentially near death (Hunt et al., 2005; p. e206). On the other hand, both treatments are highly reimbursed.

Since patient C is already on maximal medications and is not a candidate for revascularization, the management goal should be to keep him as comfortable as possible. This should be accomplished in the least invasive manner possible (e.g., at home), and if that is not possible in an uncomplicated setting, for example during admission to the hospital for simple diuresis. According to the ACC/AHA guidelines, no additional interventions are appropriate.[17] In fact, even a "simple" but invasive test, the pulmonary artery catheter, has been found to be of no marginal value over good clinical decision making in managing patients with CHF, and could even cause harm (ESCAPE, 2005).

Despite these guideline recommendations, physicians in our data show a surprising degree of enthusiasm for additional interventions. For patient B, nearly one-third of the cardiologists surveyed would recommend a repeat angiography at least as frequently as "some of the time." Similarly, 65 percent of cardiologists recommend an ICD "most of the time," or "always/almost always," while 47 percent recommend a pacemaker with at least these frequencies. For patient C, 18 percent recommend an ICU/CCU admission, 2 percent recommend a pulmonary artery catheter and 15 percent recommend a pacemaker at least

[16]This includes his advanced stage; his severe (Class IV) medication refractory heart failure; and the asymptomatic non-sustained nature of the ventricular tachycardia.

[17]Clinical improvement with a simple intervention (increasing his oxygen) also argues against more intensive interventions.

"most of the time."

Our next measure of Z^S is based on a summary of these intensity recommendations. We start with the three most intensive interventions for both patients. Cardiologists' responses on aggressiveness are highly correlated across patients B and C. Of the 28 percent (N=143) of cardiologists in the sample who would "frequently" or "always/almost always" recommend at least one of the above-listed high-intensity procedures for patient C, 93 percent (N=133) would also frequently or always/almost always recommend at least one high-intensity intervention for patient B. We use this overlap – the highest treatment recommendation overlap in our data – to define a "cowboy" cardiologist as a cardiologist who recommends at least one of the three possible intensive treatments for both patients B and C "most of the time" or "always/almost always." Because Vignette B was not presented to the primary care physicians, we use only their response to Vignette C to categorize them using the same criteria. In total, 27 percent of the cardiologists in our sample are classified as cowboys, as are 19 percent of primary care physicians.

All told, we test four measures of Z^S: high or low frequency of follow-up visits, a dummy variable for being a cowboy, and a dummy variable for being a comforter. How are these measures related? Table 3.2 shows that among both PCPs and cardiologists, chi-squared tests strongly reject the null of no association between follow-up frequencies recommended for vignette patients and a physician's status as a "cowboy" or "comforter." Physicians with a low follow-up frequency are more likely to be comforters and less likely to be cowboys than physicians with a high follow-up frequency. Similarly, cowboy physicians are far less likely to be comforter physicians (even though doctors could be classified as both). Most differences are statistically significant.[18]

[18]Patient and physician responses are only very weakly correlated across regions. The correlations across physician types shown in Table 3.2 are also quite low, with the largest magnitudes on the order of 0.1 and the majority being < 0.1.

Table 3.2: *Distribution of Physicians by Vignette Responses*

Panel A: PCPs						
	Cowboy			**Comforter**		
Follow-Up Frequency	Yes	No		Yes	No	
Low	16	61	8.4%	39	38	8.4%
Medium	98	452	60%	300	250	60%
High	87	200	31%	115	172	31%
	22%	78%		50%	50%	
$p(\chi^2)$:	**<0.01**			$p(\chi^2)$:	**0.02**	
	Comforter					
Cowboy	Yes	No				
Yes	87	114	22%			
No	367	346	78%			
	50%	50%				
$p(\chi^2)$:	0.145					
Panel B: Cardiologists						
	Cowboy			**Comforter**		
Follow-Up Frequency	Yes	No		Yes	No	
Low	17	76	18%	27	66	18%
Medium	85	238	63%	94	229	63%
High	31	69	19%	22	78	19%
	26%	74%		27%	72%	
$p(\chi^2)$:	**<0.01**			$p(\chi^2)$:	**<0.01**	
	Comforter					
Cowboy	Yes	No				
Yes	39	94	26%			
No	104	279	74%			
	28%	72%				
$p(\chi^2)$:	**<0.01**					

This table shows the bivariate relationships between the guideline-defined indicatorsfor recommended Follow-Up Frequency, as well as"Cowboy" and "Comforter" status among both PCPs and Cardiologists in our data. Chi-squared tests evaluate the null that there is no association between pairs of indicators in the table.

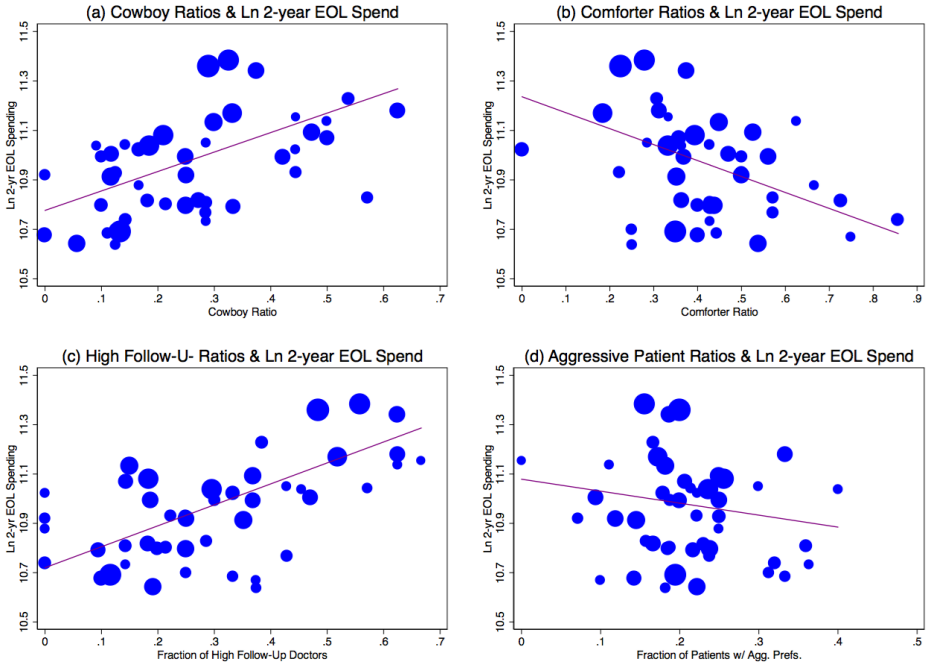
3.4 Model Estimates

We now proceed with our estimates of the models presented above. We first consider Equation (3.5), the relationship between area-level spending and local patient and physician preferences. We then turn to Equation (3.7), modeling the factors leading physicians to be more and less aggressive.

Do Survey Responses Predict Regional Medicare Expenditures?

We start with the basic relationship between area spending, patient preferences and physician preferences for the 64 HRRs with at least 3 cardiologists and 2 primary care physician responses. Figure 3.5 shows scatter plots of area-level end of life spending vs. our measures of supply and demand for care. The measures we include are the fraction of all physicians in the area who are cowboys (panel a), the fraction of physicians who are comforters (panel b), the fraction of physicians who recommend follow-up more frequently than recommended guidelines (panel c), and the share of patients who desire more aggressive care at the end of life (panel d). Each circle represents one HRR, and its size is proportional to the survey sample size in the respective HRR.

Figure 3.5: *Log of Inpatient 2-year End-of-Life Regional Spending vs. Various Independent Variables*



In the case of the three supply-side variables, the results are consistent with the theory: despite the relatively small sample sizes of physicians in each HRR, end of life spending is positively related to the cowboy ratio, negatively related to the comforter ratio, and positively related to high frequency recommendations for follow-up visits. The demand variable, in contrast, is not strongly related to spending: the data points form more of a

Table 3.3: *Regression Estimates of Ln Medicare Expenditures in the Last Two Years*

Combined Sample of PCPs and Cardiologists						
	(1)	(2)	(3)	(4)	(5)	(6)
Cowboy Ratio, All Doctors	0.7535***	0.6056***	0.6096***	0.5928***	0.5972***	
	(0.1626)	(0.1385)	(0.1173)	(0.1446)	(0.1221)	
Comforter Ratio, All Doctors	-0.4068**	-0.3206***	-0.2878**	-0.3089***	-0.2745**	
	(0.1681)	(0.1109)	(0.1103)	(0.1065)	(0.1044)	
Follow-Up Low, All Doctors		-0.4174	-0.3626	-0.4884	-0.4422	
		(0.2755)	(0.2849)	(0.3299)	(0.3215)	
Follow-Up High, All Doctors		0.9712***	0.9721***	0.9680***	0.9670***	
		(0.2053)	(0.1963)	(0.2026)	(0.1910)	
Have Unneeded Tests			0.1177		0.1424	-0.0543
			(0.2062)		(0.2251)	(0.3400)
See Unneeded Cardiologist			0.2728*		0.3035*	0.5397*
			(0.1549)		(0.1679)	(0.2855)
Aggressive Preferences Patient Ratio				-0.2355	-0.2762	-0.5395
				(0.4607)	(0.4409)	(0.7526)
Comfortable Preferences Patient Ratio				-0.1154	-0.2033	-0.1917
				(0.1584)	(0.2015)	(0.2499)
N	64	64	64	64	64	64
R^2	0.3627	0.6092	0.6299	0.6127	0.6377	0.0750

* p<0.10, ** p<0.05, *** p<0.01

2-year End-of-Life Spending is in natural log form and is price, age, sex and race adjusted. Results shown are for the 64 Hospital Referral Regions (HRRs) in which we have at least 3 patients and 3 cardiologists surveyed. All regressions include a constant and control for the fraction of primary care physicians in the sample. Respondent data is adjusted for race, sex and age. Survey sampling weights take into account differences in the number of physician observations per HRR.

cloud than a line.

Table 3.3 explores this result more formally with regression estimates of logged end-of-life expenditures, weighted by the number of physician observations per HRR and including controls for the fraction of PCPs among our survey responders. As the first column shows, the local proportion of cowboys and comforters predicts 36 percent of the observed regional variation in risk-adjusted end-of-life spending. Further, the estimated magnitudes are large: increasing the percentage of cowboys by 10 percentage points is associated with a 7.5 percent increase in end-of-life expenditures, while increasing the fraction of comforters by 10 percent implies a 4.1 percent reduction in expenditures. This relationship between spending and the local fractions of cowboys and comforters also holds when both cardiologists and primary care physicians are analyzed separately, as shown in the Appendix.

Column 2 of Table 3.3 shows that the indicator for high frequency follow-up recommendations is also a meaningful predictor of HRR-level end-of-life spending: conditional on the

fraction of cowboys and comforters, an increase of 10 percentage points of physicians who prefer to see patients more frequently than guidelines recommend is predicted to increase end-of-life spending by 9.5 percent (and while the low frequency follow-up coefficient is large in magnitude (-0.417), it is not statistically significant). Indeed the combination of just these supplier beliefs alone can explain over 60 percent of the observed end-of-life spending variation in the 64 sample HRRs.[19]

The next two columns add measures of patient preferences to the regressions: the share of patients wishing to have unneeded tests, the share wanting to see an unneeded cardiologist, the share preferring aggressive end-of-life care, and the share preferring comfortable end-of-life care. None of these variables are statistically significant at the 5% level. Even excluding the physician belief variables entirely, as in column 6, the R^2 from the patient preference variables is just 0.075. Separate regressions for cardiologists and primary care physicians are presented in Appendices C and D and show similar results.[20]

It is also possible that there could be an interaction effect between patient preferences and physician beliefs, for example if aggressive physicians interact with patients with preferences for aggressive care to generate even more utilization (or conversely for comforter physicians and patients who demand palliative care). These hypotheses are considered in Table 3.4. Column 1 of the table repeats Column 5 of Table 3.3 for reference. The subsequent columns add interaction terms. As shown in Column 2, however, there is little consistent evidence for the interactive aggressiveness hypothesis; the interaction between cowboy physicians and patients with aggressive preferences is negative (not positive as theory would suggest), and while the coefficient between comforter physicians and patients is negative (column 3), it is not statistically significant.

Column 4 of Table 3.4 repeats the analyses in column 1, but uses total average per

[19] As Black et al. (2000) note, the OLS estimate is a lower bound and under weak assumptions, the expected value of the OLS parameter estimate is of smaller magnitude than the true parameter. (The R^2 is also a lower bound owing to measurement error.)

[20] Our results do not appear to be driven by geography. The coefficient estimates are similar when the east and west coasts of the US are estimated separately.

Table 3.4: *Regression Estimates of Ln Medicare Expenditures Considering Interaction Terms and Additional Measures of HRR-Level Spending*

Combined Sample of PCPs and Cardiologists (dependent variables listed in column headings; all are in natural logs)					
	(1) 2-yr EOL Spend (As in Table 3.4)	(2) 2-yr EOL Spend	(3) 2-yr EOL Spend	(4) Total Spend (Av. per Beneficiary)	(5) Total Spend (Hip Fract. Cohort)
Cowboy Ratio, All Doctors	0.5972***	0.5938***	0.5835***	0.3306***	0.2793***
	(0.1221)	(0.1119)	(0.1260)	(0.1028)	(0.0806)
Comforter Ratio, All Doctors	-0.2745**	-0.2600**	-0.3175**	-0.0889	-0.0682
	(0.1044)	(0.1002)	(0.1224)	(0.1064)	(0.0749)
Follow-Up Low, All Doctors	-0.4422	-0.4074	-0.4824	-0.5208	-0.1663
	(0.3215)	(0.2749)	(0.3180)	(0.3751)	(0.2322)
Follow-Up High, All Doctors	0.9670***	1.0267***	0.9436***	0.2480	0.2933**
	(0.1910)	(0.1837)	(0.1870)	(0.1777)	(0.1291)
Have Unneeded Tests	0.1424	0.1015	0.1766	-0.0792	-0.0417
	(0.2251)	(0.2274)	(0.2242)	(0.2005)	(0.1814)
See Unneeded Cardiologist	0.3035*	0.2159	0.2746*	0.3353	0.1996
	(0.1679)	(0.1666)	(0.1617)	(0.2434)	(0.1478)
Aggressive Preferences Patient Ratio	-0.2762	0.1880	0.6315	-0.3026	-0.1027
	(0.4409)	(0.5051)	(0.9285)	(0.4703)	(0.3086)
Comfortable Preferences Patient Ratio	-0.2033	-0.6297***	0.1663	-0.2500	-0.0660
	(0.2015)	(0.1975)	(0.3022)	(0.1830)	(0.1524)
Cowboy Ratio*		-2.1268			
Aggressive Preferences Patient Ratio		(2.1367)			
Cowboy Ratio*		1.5977**			
Comfortable Preferences Patient Ratio		(0.7557)			
Comforter Ratio*			-2.2461		
Aggressive Preferences Patient Ratio			(1.8854)		
Comforter Ratio*			-0.9179		
Comfortable Preferences Patient Ratio			(0.6437)		
N	64	64	64	64	64
R^2	0.6377	0.6603	0.6459	0.3482	0.3705

* p<0.10, ** p<0.05, *** p<0.01; 2-year End-of-Life Spending and total spending are are price, age, sex and race adjusted. Hip fracture cohort spending is adjusted for age, sex, race, comorbid conditions at admission, and the hierarchical condition categories risk-adjustment index for the six months prior to admission. Results shown are for the 64 Hospital Referral Regions (HRRs) in which we have at least 3 patients and 3 cardiologists surveyed. All regressions include a constant and control for the fraction of primary care physicians in the sample. Respondent data is adjusted for race, sex and age. Survey sampling weights take into account differences in the number of physician observations per HRR.

beneficiary Medicare expenditures (adjusted for prices, age, sex, and race/ethnicity) as the dependent variable. This expenditure measure likely reflects a greater share of primary care spending relative to specialty care. In the combined sample, the fraction of cowboys in an HRR is a consistently strong predictor of spending across models. Moreover, although R^2 values are smaller in these models, supply-side factors continue to explain more of the variation in spending than demand-side factors. Finally, we consider fully risk-adjusted one-year expenditures for a "forward looking" cohort of hip fracture patients in Column 5 of Table 3.4. The estimated coefficients suggest relationships similar to those in Column 1, but, like the model explaining overall Medicare expenditures, the coefficients are smaller in magnitude and the R^2 is smaller in magnitude as well (0.37 versus 0.64).

Our data imply a strong relationship between physician type and spending, as a simple back-of-the-envelope calculation suggests. We calculate how much Medicare expenditures would change in a counterfactual setting in which there were no cowboys, all physicians were comforters, and all physicians met guidelines for follow-up care. In this counterfactual,

end-of-life expenditures would be predicted to decline by 36 percent, and total Medicare expenditures would be expected to decline by 17 percent. These comparisons point to the importance of physician beliefs in explaining regional (and national) utilization patterns.

What factors predict physician responses to the vignettes?

To this point, we have shown that physician beliefs matter for spending, and that physician beliefs vary across areas more than would be expected given random variation. The obvious question is then: what explains this variation in physician beliefs? In this section, we estimate the model in Equation (3.7) to test for the relative importance of financial and organizational factors in explaining physician recommendations.

Table 3.5 presents coefficients from a linear probability model with HRR-level random effects for three regressions at the physician level. Our dependent variables are binary indictors for whether the physician is a cowboy (Column 1), a comforter (Column 2), or recommends in high frequency follow-up (Column 3). In each model, we include basic physician demographics: age, gender, board certification status, whether the physician is a cardiologist, days per week spent seeing patients, as well as cardiologists per 100,000 Medicare beneficiaries. Notably, some of these characteristics matter for predicting physician types: male physicians in the sample are both somewhat more likely to be cowboys and less likely to be comforters than female doctors and older physicians are more likely to be high follow-up doctors and cowboys: at the mean age of 57.5 years, a 1 standard deviation increase in physician age (9.8 years) is associated with a 4.6% increase in probability of being a cowboy and a 5.5% increase in probability of being a high follow-up doctor.

The demographic factors included reveal that older physicians are more likely to recommend high rates of follow-up and are also more likely to be cowboys, but age is not a significant predictor of comforter status. Male physicians are less likely to be comforters, while board certification – a rough marker for physician quality – is negatively associated with cowboy status and high follow-up frequency. This result is consistent with Doyle et al. (2010), who found that lower quality physicians spent 10-25% more on treating otherwise identical patients.

Table 3.5: *Predictors of Cowboy, Comforter & High Follow-Up Types*

	(1) Cowboy	(2) Comforter	(3) High Follow-Up
General Controls			
Age	0.0047***	0.0005	0.0056***
	(0.0013)	(0.0015)	(0.0012)
Male	0.0532*	-0.0625*	-0.0165
	(0.0315)	(0.0370)	(0.0314)
Weekly Patient Days	-0.0112	0.0145	0.0008
	(0.0076)	(0.0090)	(0.0076)
Board Certified	-0.0727*	0.0184	-0.1400***
	(0.0379)	(0.0445)	(0.0378)
Cardiologists per 100k	0.0203***	-0.0223***	0.0410***
	(0.0076)	(0.0079)	(0.0061)
Cardiologist Dummy	-0.0187	-0.1752***	-0.0695*
	(0.0363)	(0.0426)	(0.0361)
Financial Factors			
Fraction Capitated Patients	0.0980**	-0.0428	0.1073**
	(0.0462)	(0.0540)	(0.0457)
Fraction Medicaid Patients	0.2894***	0.0325	0.3978***
	(0.0931)	(0.1090)	(0.0924)
Organizational Factors			
(Baseline = Solo or 2-person Practice)	-	-	-
Single/Multi Speciality Group Practice	-0.0584**	-0.0169	-0.2019***
	(0.0265)	(0.0310)	(0.0262)
Group/Staff HMO or Hospital-Based Practice	-0.1539***	0.0357	-0.2221***
	(0.0429)	(0.0502)	(0.0426)
Responsiveness Factors			
Responds to Patient Expectations	-0.0272	0.0307	-0.0145
	(0.0313)	(0.0368)	(0.0313)
Responds to Colleague Expectations	0.0147	-0.0007	0.0360
	(0.0247)	(0.0291)	(0.0247)
Responds to Referrer Expectations	0.1084***	0.0248	-0.0516
	(0.0419)	(0.0493)	(0.0420)
Responds to Malpractice Concerns	-0.0051	0.0222	-0.0105
	(0.0247)	(0.0290)	(0.0247)
N	1349	1349	1349
R^2 (within)	0.0502	0.0509	0.1075
R^2 (between)	0.0379	0.1049	0.2110
R^2 (overall)	0.0613	0.0596	0.1609

* p<0.10, ** p<0.05, *** p<0.01

All logit regressions include a constant, and HRR-level random effects as well as general physician-level controls. Additional explanatory variables include financial, organizational and responsiveness factors. The question about responding to referring doctor expectations appeared in the Cardiologist survey only, and so reflects the preferences of cardiologists only. The cardiology dummy variable therefore reflects both the pure effect of being a practicing cardiologist, and a secondary adjustment arising from the referral question being set to zero for all primary care physicians.

A greater number of cardiologists per 100,000 Medicare beneficiaries is associated with a higher likelihood of a physician being a cowboy or high follow-up doctor and with a lower likelihood of the physician being a comforter. One might be tempted to interpret this as classic "supplier-induced demand" effect, with more cardiologists per capita leading to less income per cardiologist, and hence a greater incentive to treat a given patient more

intensively. Yet the equilibrium supply of cardiologists is likely to depend on a wide variety of factors, suggesting caution in the interpretation.

The substitution effect implies that lower incremental reimbursements associated with Medicaid and capitated patients would lead to fewer interventions and more palliative care. Table 3.5 shows that physicians with a larger fraction of Medicaid and (to a lesser extent) capitated patients are more likely to be cowboys and high-follow-up physicians, rejecting the dominance of the substitution effect. One may appeal again to a dominant income effect to explain these patterns.

Some organizational factors are strongly associated with physician beliefs about appropriate practice. Physicians in solo or 2-person practices are far more likely to be aggressive than physicians in single or multi-specialty group practices or physicians who are part of an HMO or a hospital-based practice. Yet physicians who work in a group or staff model HMOs or hospital-based practice are no more likely to be comforters. Physicians who respond to patient expectations are more likely to be comforters, and those responding to referring physician expectations are more likely to be high follow-up physicians, but neither effect is statistically significant. Whether cardiologists accommodate referring physicians – also a financial factor (since cardiologists will benefit financially from future referrals) as well as an organizational one – is a large and statistically significant predictor of being a cowboy.[21] Finally, malpractice concerns are neither predictive of cowboy nor comforter status, perhaps because procedures performed on high-risk patients (such as Patients B and C) can increase the risk of a malpractice suit.

The explanatory power of these regressions is quite modest – between 6 and 15 percent – suggesting that a considerable degree of the remaining variation is the consequence of physician beliefs regarding the productivity of treatments, rather than behaviors systematically related to financial, organizational, or other factors.

As a final exercise, we include these financial, organizational, and responsiveness variables, aggregated up to the HRR level, in a regression that seeks to explain the variation

[21] Note that this question is asked only of cardiologists.

in log end-of-life spending – an expanded counterpart to Table 3.4. These results are presented in the Appendix. Aside from the per-capita supply of cardiologists – a potentially suspect measure of capacity – none of the additional variables are statistically significant, nor do they add appreciably to the explanatory power of the regression. Physician beliefs, independent of financial or organizational factors, appear to explain a great deal of why physicians are cowboys or comforters and how the frequencies of these typologies, in turn, are related to overall spending.

3.5 Conclusion and Implications

While there is a good deal of regional variation in medical spending and care utilization in the U.S. and elsewhere, there is little agreement about the causes of such variations. Do they arise from variation in patient demand, from variation in physician behavior, or both? In this paper, we found that regional measures of patient demand as measured by responses to a nationwide survey had only modest predictive association with regional end-of-life expenditures. By contrast, regionally aggregated measures of physician beliefs regarding treatment options can explain a substantial degree of observed regional variation in utilization in the U.S. Medicare population. While other results have suggested such a finding (Sirovich et al. (2008), Lucas et al. (2008), Bederman et. al. (2011), and Wennberg et al. (1997)), our paper is the first to directly relate Medicare spending to physician beliefs.

Unfortunately, we are not able to match physicians directly to their own patients, which we acknowledge is a shortcoming of the survey methodology. However, we are able to link the patient and physician surveys at the HRR level and the regional evidence is consistent with the dominant importance of physician beliefs in explaining HRR-level utilization patterns. A back-of-the-envelope calculation using our regression results implies that, were all physicians in the 64 HRRs studied to follow professional guidelines, end-of-life Medicare expenditures in these areas would be expected to be 36 percent lower, and overall Medicare

expenditures 17 percent lower.[22]

We then turned to the factors that lead physicians to have different preferences. We found that the traditional factors in supplier-induced demand models, such as the fraction of patients paid through capitation (or on Medicaid), or physicians' responsiveness to financial factors, play a relatively small role in explaining equilibrium variations in utilization patterns. Organizational factors, such as accommodating colleagues, help to explain only a small amount of observed variation in individual intervention decisions. Instead, differences in physician beliefs about the effectiveness of treatments explain the lion's share of of inter-regional variation in Medicare expenditures.[23]

Our results differ from the existing literature in that they are based on vignettes and thus represent a lower bound to practice variations. Generally, prior studies inferred practice variations as the residual from an area model, leading to estimates being biased either upward (because of unobserved regional factors) or biased downward (because of flawed risk-adjustment, as in Song et al., 2010).

One concern about the interpretation of the vignette responses as "overuse" is that they may reflect the true productivity of physicians. While we cannot rule this out, we note that physicians with greater objective qualifications such as board certification are no more likely to be cowboys. Nor do the updated 2009 heart failure guidelines recommend more aggressive care (Hunt et al., 2009), as a model of inappropriately cautious and slowly evolving recommendations would suggest.

Another hypothesis is that while cowboys may over-treat patients along some dimensions,

[22] As one seminar participant noted, Medicare doesn't reimburse for talks, but talks take a lot of time. Absent financial incentives and given implicit time costs of conversations about end-of-life and palliative care, perhaps we should not be surprised that doctors under-provide this type of service relative to those that are (sometimes quite generously) reimbursed. Another seminar participant noted that medical ethics call for the consultant to speak only to referring doctor and not to the patient; this is another reason we might expect to see fewer palliative care conversations by cardiologists.

[23] This result is consistent with Epstein and Nicholson (2009), who find large variations in Cesarean section surgical rates among obstetricians within the same practice, even after adjusting for where the physicians trained. It is also consistent with Chassin's (1993) "Enthusiasm Hypothesis" – that regional differences in the use of health care services are caused by differences in the prevalence of physicians who are enthusiasts for those services.

they may also avoid the underuse of effective care along other dimensions (e.g., Landrum et al., 2008). Our survey did not ask about whether the physician would recommend appropriate levels of effective care or not. But other evidence does not support this hypothesis: an HRR-level composite AMI quality measure from 2007 Hospital Compare Data, (Dartmouth Atlas, 2013) is negatively associated with the HRR-level fraction of physicians who are cowboys in our data.

Unfortunately, the data we consider in this study cannot shed light on how these differences in physician beliefs arise. Simple heterogeneity in physician beliefs cannot explain regional variation in expenditures, since the observed regional patterns in physician beliefs exhibit far greater inter-region variation than would be expected due to chance alone. Rather, spatial correlation in beliefs is required in order to explain the regional patterns we see. We do find that physicians' propensity to intervene for non-clinical reasons is related to the expectations of physicians with whom they regularly interact, a result consistent with network models. Similarly, Molitor (2011) finds that cardiologists who move to more or less aggressive regions change their practice style to better conform to local norms. However we are still left with questions as to how and why some regions become more aggressive than others.

Our results do not imply that economic incentives are unimportant. Clearly, changes in payment margins have a large impact on behavior, as has been shown in a variety of settings. But the prevalence of geographic variations in European countries, where economic incentives are often nearly entirely blunted, is consistent with the view that physician beliefs play a large role in explaining such variations. A better understanding of both how physician beliefs form, and how they can be shaped, is a key challenge for future research.

Bibliography

Acemoglu, Daron, and Joshua Linn. "Market size in innovation: theory and evidence from the pharmaceutical industry," *The Quarterly Journal of Economics* 119, no. 3 (2004): 1049-1090.

Ackerly, Nyssa, John Eyraud, Cal Franz, Brielle Kissel, Dina Metivier. "Financial Conflict-of-Interest Disclosure and Voting Patterns at FDA Advisory Committee Meetings." Contract No. HHSF 223200810017I: submitted to Nancy Gieser, Office of Policy, Planning and Preparedness, FDA.

Agha, Leila and David Militor. "The Influence of Pioneer Investigators on Technology Adoption: Evidence from New Cancer Drugs" (Working Paper, 2014)

Ameriks, John, Andrew Caplin, Steven Laufer, and Stijn van Nieuwerburgh, 2011. "The Joy of Giving or Assisted Living? Using Strategic Surveys to Separate Public Care Aversion from Bequest Motives," *Journal of Finance*, 66(2) (April), 519-561.

Angell, Marcia. "Industry-sponsored clinical research: a broken system." *JAMA* 300, no. 9 (2008): 1069-1071.

Anthony, Denise L., M. Brooke Herndon, Patricia M. Gallagher, Amber E. Barnato, Julie P.W. Bynum, et al., 2009. "How Much do Patients' Preferences Contribute to Resource Use?" *Health Affairs* 28, May-Jun, 864-73.

Babiarz, Josephine C and Pisano, Douglas J. "Overview of FDA and drug development," in *FDA Regulatory Affairs: A Guide for Prescription Drugs, Medical Devices and Biologics, 2nd ed.* edited by Douglas J. Pisano and David S. Mantus, Informa Healthcare USA, New York, pages 1-32, 2008.

Bach, Peter B., Deborah Schrag, and Colin B. Begg, 2004. "Resurrecting Treatment Histories of Dead Patients: A Study Design That Should Be Laid to Rest," *JAMA* 292(22) (December 8): 2765-2770.

Barnato, Amber E., M. Brooke Herndon, Denise L. Anthony, Patricia Gallagher, Jonathan S. Skinner, et al., 2007. "Are regional variations in end-of-life care intensity explained by patient preferences? A study of the US Medicare population," *Medical Care* 45, 386-393.

Bartel, Ann P., and Lacy Glenn Thomas. "Predation through regulation: the wage and profit effects of the Occupational Safety and Health Administration and the Environmental Protection Agency." *Journal of Law and Economics* 30 (1987): 239.

Baumann, Andrea O., Raisa B. Deber, Gail G. Thompson, 1991, "Overconfidence Among Physicians and Nurses: The 'Micro-Certainty, Macro-Uncertainty' Phenomenon," Social Science and Medicine 32(1): 167-74.

Bederman S. Samuel, Peter C. Coyte, Hans J. Kreder, Nizar N. Mahomed, Warren J. McIsaac, and James G. Wright, 2011. "Who's in the driver's seat? The influence of patient and physician enthusiasm on regional variation in degenerative lumbar spinal surgery: a population-based study," *Spine*; 36(6): 481-489.

Black, Dan A.; Mark C. Berger; Frank A. Scott, 2000. "Bounding Parameter Estimates with Nonclassical Measurement Error," *Journal of the American Statistical Association*, Vol. 95, No. 451, pp. 739-748.

Brennan, Troyen A., David J. Rothman, Linda Blank, David Blumenthal, Susan C. Chimonas, Jordan J. Cohen, Janlori Goldman et al. "Health industry practices that create conflicts of interest: a policy proposal for academic medical centers." *JAMA* 295, no. 4 (2006): 429-433.

Brown SL, Bright RA, Tavris DR. "Medical device epidemiology and surveillance: patient safety is the bottom line." Expert Rev Med Devices 2004;1:1-2.

Budish, Eric, Benjamin N. Roin, and Heidi Williams. "Do fixed patent terms distort innovation? Evidence from cancer clinical trials". National Bureau of Economic Research (NBER) Working Paper 19430, 2013.

Carpenter, Daniel P. "Groups, the Media and Agency Waiting Costs: The Political Economy of FDA Drug Approval," (2002) 46 *American Journal of Political Science* 490-505.

Carpenter, Daniel P. "Protection without capture: Product approval by a politically responsive, learning regulator."*American Political Science Review* 98, no. 4 (2004): 613-31.

Carpenter, Daniel P. *Reputation and Power: Organizational Image and Regulation at the FDA*. Princeton: Princeton University Press, 2010.

Carpenter, Daniel P., Moffitt, Susan. I., Moore, Colin. D., Rynbrandt, Ryan T., Ting, Michael M., Yohai, Ian., & Zucker, Evan James."Early entrant protection in approval regulation: Theory and evidence from FDA drug review". *Journal of Law, Economics, and Organization* (2010), 26(3), 515-545.

Centers for Medicare & Medicaid Services. *National Health Expenditures* (2012). Office of the Actuary, National Health Statistics Group; U.S. Department of Commerce, Bureau of Economic Analysis; and U.S. Bureau of the Census.

Chandra, Amitabh, and Douglas O. Staiger. 2007. "Productivity Spillovers in Healthcare: Evidence from the Treatment of Heart Attacks," *Journal of Political Economy* 115, 103-140.

Chandra, Amitabh, and Jonathan Skinner, 2012, "Productivity growth and expenditure growth in U.S. Health care," *Journal of Economic Literature* 50(3) (September): 645-80.

Chassin, Mark R. "Explaining geographic variations: the enthusiasm hypothesis." *Medical Care* (1993): YS37-YS44.

Chatterji, Aaron K. "Spawned with a silver spoon? Entrepreneurial performance and innovation in the medical device industry." *Strategic Management Journal* 30, no. 2 (2009): 185-206.

Chatterji, Aaron K., Kira R. Fabrizio, Will Mitchell, and Kevin A. Schulman. "Physician-industry cooperation in the medical device industry." *Health Affairs* 27, no. 6 (2008): 1532-1543.

Clemens, Jeffrey, and Joshua Gottlieb, 2012. "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?" Stanford University (SIEPR), http://siepr.stanford.edu/publicationsprofile/2447.

Cockburn, Iain M., and Rebecca M. Henderson. "Absorptive capacity, coauthoring behavior, and the organization of research in drug discovery." *The Journal of Industrial Economics* 46, no. 2 (1998): 157-182.

Connors Alfred F., Jr, Neal V. Dawson, Normal A. Desbiens, et al. 1995: "A Controlled Trial to Improve Care for Seriously III Hospitalized Patients: The Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT)." *JAMA*. 274(20):1591-1598.

Cutler, David, Jonathan Skinner, Ariel Dora Stern, and David Wennberg. "Physician beliefs and patient preferences: a new look at regional variation in health care spending." No. w19320. National Bureau of Economic Research, 2013.

Dartmouth Atlas, 2013. http://www.dartmouthatlas.org/tools/downloads.aspx

Dhruva SS, Bero LA, Redberg RF. "Strength of study evidence examined by the FDA in premarket approval of cardiovascular devices." *JAMA* 2009;302:2679-85. [Erratum, JAMA 2010;303:422.]

DiMasi, J.A. and H.G. Grabowski "R&D Costs for New Biotech Compounds," *Managerial and Decision Economics*, 28: (2007) 469-479.

Doyle, Joseph, Steven Ewer, and Todd Wagner, 2010. "Returns to Physician Human Capital: Evidence from Patients Randomized to Physician Teams" *Journal of Health Economics.* 29(6). December 2010: 866-882.

Dranove, David, and David Meltzer. "Do important drugs reach the market sooner?". *The RAND Journal of Economics* (1994): 402-423.

Dranove, David, and Paul Wehner, 1994. "Physician-induced demand for childbirths," *Journal of Health Economics* 13(1) (March): 61-73.

Dresselhaus, Timothy R., John W. Peabody, Jeff Luck, Dan Bertenthal, 2004. "An Evaluation of Vignettes for Predicting Variation in the Quality of Preventive Care," *Journal of General Internal Medicine*, October; 19(10): 1013-1018.

Epstein, Andrew J., and Sean Nicholson, 2009. "The formation and evolution of physician treatment styles: An application to cesarean sections," *Journal of Health Economics*, 28(6) (December): 1126-1140.

ESCAPE (Evaluation Study of Congestive Heart Failure and Pulmonary Artery Catheterization Effectiveness): The ESCAPE Trial. *JAMA*. 2005; 294(13): 1625-1633.

Fisher, Elliott S., David E. Wennberg, Terese A. Stukel, Daniel J. Gottlieb, Francis Lee Lucas, et al., 2003, "The implications of regional variations in Medicare spending. Part 2: Health outcomes and satisfaction with care," *Ann Intern Med* 138, Feb 18, 288-98.

Food and Drug Administration, Office of the Commissioner. Guidance for the Public, FDA Advisory Committee Members, and FDA Staff on Procedures for Determining Conflict of Interest and Eligibility for Participation in FDA Advisory Committees (2008).

Garber, Alan M. "Modernizing device regulation." *New England Journal of Medicine* 362, no. 13 (2010): 1161-1163.

Generics and Biosimilars Initiative (GaBI) "The future of nanomedicines – nanosimilars" (August, 2013) Available at:
http://www.gabionline.net/Biosimilars/Research/The-future-of-nanomedicines-nanosimilars.

Goldman, Dana, and Darius Lakdawalla. "Intellectual Property, Information Technology, Biomedical Research, and Marketing of Patented Products." In *Handbook of Health Economics Volume 2*, edited by Mark V. Pauly, Thomas G. Mcguire and Pedro P. Barros, 825-872. New York: Elsevier, 2011.

Grabowski, Henry G., John M. Vernon, and Lacy Glenn Thomas. "Estimating the effects of regulation on innovation: an international comparative analysis of the pharmaceutical industry." *Journal of Law and Economics* 21, no. 1 (1978): 133-163.

Gruber, Jonathan, and Maria Owings. 1996. "Physician financial incentives and cesarean section delivery." *Rand Journal of Economics*, 27 (1):99-123.

Guedj, Ilan and David Scharfstein. "Organizational Scope and Investment: Evidence from the Drug Development Strategies and Performance of Biopharmaceutical Firms". NBER Working Paper 10933, 2004.

Haider, Adil H., Janel Sexton, N. Sriram, Lisa A. Cooper, David T. Efron, Sandra Swoboda, Cassandra V. Villegas et al. "Association of unconscious race and social class bias with vignette-based clinical assessments by medical students." *JAMA* 306, no. 9 (2011): 942-951.

Hamburg, Margaret A. "Innovation, regulation, and the FDA." *New England Journal of Medicine* 363, no. 23 (2010): 2228-2232.

Hamburg, Margaret A., and Joshua M. Sharfstein. "The FDA as a public health agency." *New England Journal of Medicine* 360, no. 24 (2009): 2493-2495.

Hellström, Mats, and Alan R Harvey. "Retinal ganglion cell gene therapy and visual system repair". *Current gene therapy* 11.2 (2011): 116-131.

Hunt, Sharon Ann, William T. Abraham, Marshall H. Chin, Arthur M. Feldman, Gary S. Francis, et al., 2005. "ACC/AHA 2005 Guideline Update for the Diagnosis and Management of Chronic Heart Failure in the Adult," *Circulation*. 2005;112:e154-e235; (September 13).

Hunt, Sharon Ann, William T. Abraham, Marshall H. Chin, Arthur M. Feldman, Gary S. Francis, et al., 2009. "2009 Focused Update Incorporated Into the ACC/AHA 2005 Guidelines for the Diagnosis and Management of Chronic Heart Failure in the Adult," *Circulation* 119:e391-e479 (March 26).

Jacobson, Mireille, A. James O'Malley, Craig C. Earle, Juliana Pakes, Peter Gaccione and Joseph P. Newhouse, 2006. "Does Reimbursement Influence Chemotherapy Treatment For Cancer Patients?" *Health Affairs*, 25(2): 437-443.

Kaplan, Aaron V., Donald S. Baim, John J. Smith, David A. Feigal, Michael Simons, David Jefferys, Thomas J. Fogarty, Richard E. Kuntz, and Martin B. Leon. "Medical Device Development From Prototype to Regulatory Approval." *Circulation* 109, no. 25 (2004): 3068-3072.

Kelly, Amy S., Susan L. Ettner, Sean Morrison, Qingling Du, Neil S. Wenger, and Catherine A. Sarkislan, 2011. "Determinants of Medical Expenditures in the Last 6 Months of Life," *Annals of Internal Medicine* 154:235-242.

Kesselheim, Aaron S., Christopher T. Robertson, Jessica A. Myers, Susannah L. Rose, Victoria Gillet, Kathryn M. Ross, Robert J. Glynn, Steven Joffe, and Jerry Avorn. "A randomized study of how physicians interpret research funding disclosures." *New England Journal of Medicine* 367, no. 12 (2012): 1119-1127.

Kramer, Daniel B., Shuai Xu, and Aaron S. Kesselheim. "Regulation of medical devices in the United States and European Union." *New England journal of medicine* 366, no. 9 (2012): 848-855.

Lancet, The. "Approval of biosimilars in the USA – dead ringers?" The Lancet. 25 February 2012 (Vol. 379, Issue 9817, Page 686).

Landrum, Mary Beth, Ellen R. Meara, Amitabh Chandra, Edward Guadagnoli and Nancy L. Keating, 2008. "Is Spending More Always Wasteful? The Appropriateness of Care and Outcomes among Colorectal Cancer Patients," *Health Affairs* 27(1):159-168 (January).

Lerner, Josh, and Julie Wulf. "Innovation and incentives: Evidence from corporate R&D." *The Review of Economics and Statistics* 89.4 (2007): 634-644.

Levine-Taub, Anna A., Anton Kolotilin, Robert S. Gibbons, Ernst R. Berndt, "The Diversity of Concentrated Prescribing Behavior: An Application to Antipsychotics," NBER Working Paper No. 16823 (February 2011).

Loewenstein, George, Sunita Sah, and Daylian M. Cain. "The unintended consequences of conflict of interest disclosure." *JAMA* 307, no. 7 (2012): 669-670.

Lucas, Francis Lee, Andrea E. Siewers, David J. Malenka, and David E. Wennberg, "Diagnostic-therapeutic cascade revisited. Coronary angiography, coronary artery bypass surgery, and percutaneous coronary intervention in the modern era." *Circulation*. 2008;118:2797-2802.

Lurie, Peter, Cristina M. Almeida, Nicholas Stine, Alexander R. Stine, and Sidney M. Wolfe. "Financial conflict of interest disclosure and voting patterns at Food and Drug Administration Drug Advisory Committee meetings." *JAMA* 295, no. 16 (2006): 1921-1928.

Maisel, William H. "Medical device regulation: an introduction for the practicing physician." *Annals of Internal Medicine* 140, no. 4 (2004): 296-302.

Maisel, William H., and Warren K. Laskey. "Drug-eluting stents". *Circulation* 115.17 (2007): e426-e427.

Mandelblatt Jeanne S., Leigh Anne Faul, Gheorghe Luta, Solomon B. Makgoeng, Claudine Isaacs, et al., 2012. "Patient and physician decision styles and breast cancer chemotherapy use in older women: Cancer and Leukemia Group B protocol 369901" *Journal of Clinical Oncology* July 20;30(21): 2609-14.

McClellan, Mark, and Jonathan Skinner, 2006, "The Incidence of Medicare," *Journal of Public Economics* 90, 2006/1, 257-276.

McGuire, Thomas G., 2000, "Physician Agency," in Mark V. Pauly, Thomas G. McGuire and Pedro P. Parros, eds.: em The Handbook of Health Economics (North-Holland).

McGuire, Thomas G., 2011, "Physician agency and payment for primary medical care," in Sherry Glied, and Peter C. Smith, eds.: The Oxford handbook of health economics (Oxford University Press).

McGuire, Thomas G., Pauly, M., 1991. "Physician response to fee changes with multiple payers." *Journal of Health Economics*, 10 (4), 385-410.

Meier, Barry "Costs Surge for Medical Devices, but Benefits Are Opaque." *The New York Times*, November 5, 2009.

Moffitt, Susan L. (2010) "Promoting Agency Reputation Through Public Advice: Advisory Committee Use in the FDA." *Journal of Politics* 72: 880-93.

Moliter, David, "The Evolution of Physician Practice Styles: Evidence from Cardiologist Migration," MIT Working Paper, 2011. http://economics.mit.edu/files/7301.

Monahan, Charles and Josephine C. Babiarz "The New Drug Application," in *FDA Regulatory Affairs: A Guide for Prescription Drugs, Medical Devices and Biologics, 2nd ed.* edited by Douglas J. Pisano and David S. Mantus, Informa Healthcare USA, New York, pages 69-108, 2008.

Munsey, Rodney R. "Trends and events in FDA regulation of medical devices over the last fifty years." Food & Drug LJ 50 (1995): 163.

Olson, Mary. "Substitution in Regulatory Agencies: FDA Enforcement Alternatives," 12 *Journal of Law Economics and Organization* (1995) 376-407.

Pauly, Mark, 1980. *Doctors and their workshops: Economic models of physician behavior* (University of Chicago Press, Chicago).

Peabody John W., Jeff Luck, Peter Glassman, Sharad Jain, Joyce Hansen, Maureen Spell, and Martin Lee, 2004. "Measuring the quality of physician practice by using clinical vignettes: a prospective validation study," *Annals of Internal Medicine* Nov 16;141(10):771-80.

Ransohoff, David F., Mary McNaughton Collins, and Floyd J. Fowler Jr., 2002. "Why is prostate cancer screening so common when the evidence is so uncertain? A system without negative feedback," *The American Journal of Medicine* 113(8): 663-67.

Redberg, Rita F., and Sanket S. Dhruva. "Medical Device Recalls: Get It Right the First Time: Comment on "Medical Device Recalls and the FDA Approval Process"." *Archives of internal medicine* 171, no. 11 (2011): 1011.

Resnic FS, Normand SL. "Postmarketing surveillance of medical devices–filling in the gaps." *N Engl J Med*. 2012 Mar 8;366(10):875-7. Epub 2012 Feb 14.

Sall, Barry "FDA Medical Device Regulation," in *FDA Regulatory Affairs: A Guide for Prescription Drugs, Medical Devices and Biologics, 2nd ed.* edited by Douglas J. Pisano and David S. Mantus, Informa Healthcare USA, New York, pages 69-108, 2008.

Sirovich, Brenda, Patricia M. Gallagher, David E. Wennberg, and Elliott S. Fisher, 2008. "Discretionary Decision Making by Primary Care Physicians and the Cost of U.S. Health Care," *Health Affairs* 27, May-Jun, 813-23.

Skinner, Jonathan, 2012. "Causes and Consequences of Geographic Variation in Health Care" in Thomas G. McGuire, Mark Pauly, and Pedro P. Barros (eds.) *Handbook of Health Economics* Vol. 2, North Holland.

Smith, Jeffrey F., Seth A. Townsend, Navjot Singh, and Philip Ma. "FDA advisory committee meeting outcomes." *Nature Reviews Drug Discovery* 11(Nr.7), (2012): 513-514.

Steinbrook, Robert. "Financial conflicts of interest and the Food and Drug Administration's Advisory Committees." *N Engl J Med* 353, no. 2 (2005): 116-8.

Stern, Ariel Dora. "Innovation under Regulatory Uncertainty: Evidence from Medical Technology." Working paper (March, 2014).

Studdert, David M., Michelle M. Mello, and Troyen A. Brennan. "Financial conflicts of interest in physicians' relationships with the pharmaceutical industry–self-regulation in the shadow of federal prosecution." *N Engl J Med* 351, no. 18 (2004): 1891-1900.

Thomas, Lacy Glenn. "Regulation and firm size: FDA impacts on innovation." *The RAND Journal of Economics* (1990): 497-517.

U.S. Department of Health and Human Services, 2004. *Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure* (JNC7). NIH Publication No. 04-5230.

Wennberg, John E., Benjamin A. Barnes and Michael Zubkoff, 1982. "Professional Uncertainty and the Problem of Supplier-Induced Demand," Social Science and Medicine 16(7): 811-824.

Wennberg, David E., John D. Dickens, Jr., Lois Biener, Floyd J. Fowler, Jr., David N. Soule, and Robert B. Keller, 1997. "Do Physicians Do What they Say? The Inclination to Test and Its Association with Coronary Angiography Rates." *Journal of General Internal Medicine* 12(3):172-176.

Wizemann T, ed. "Public health effectiveness of the FDA 510(k) clearance process: balancing patient safety and innovation: workshop report." Washington, DC: National Academies Press, 2010.

Xu, Shuai, Jerry Avorn, and Aaron S. Kesselheim. "Origins of Medical Innovation The Case of Coronary Artery Stents." *Circulation: Cardiovascular Quality and Outcomes* 5, no. 6 (2012): 743-749.

Yang, Q. Z., J. P. Sun, and C. F. Zhu. "Medical device innovation methods and case studies". Vol. 7. *SIMTech Technical Reports*, 2006.

Zenios, Stefanos A., Josh Makower, and Paul G. Yock. *Biodesign: the process of innovating medical technologies*. 2010. Cambridge: Cambridge University Press.

Zinner, Darren E., Dragana Bolcic-Jankovic, Brian Clarridge, David Blumenthal, and Eric G. Campbell. "Participation of academic scientists in relationships with industry." *Health Affairs* 28, no. 6 (2009): 1814-1825.

Zuckerman, Diana M., Paul Brown, and Steven E. Nissen. "Medical device recalls and the FDA approval process." *Archives of internal medicine* 171, no. 11 (2011): 1006.

Appendix A

Appendix to Chapter 1

A.1 Firm Experiences in the PMA Process

Case 1: A Protracted Review Process for a New Device

The company CardioMEMS is the developer of the Champion Heart Failure Monitoring System device. This device is a permanently implantable pressure measurement system that sits in the pulmonary artery of heart failure patients and monitors pressure and heart rate, transmitting data wirelessly. It is intend to assist in the ambulatory management of heart failure and reduce associated hospital stays (Loh et. al., 2013).

The device was evaluated in the CHAMPION Trial in which 550 patients were randomized into treatment or control groups. In the treatment group, physicians were provided access to patients' pulmonary artery pressure and all physicians were instructed in the adjustment of heart failure medications. According to CardioMEMS,

> *"The CHAMPION trial achieved all pre-specified primary efficacy and safety endpoints. Specifically, the rate of adjudicated hospitalizations for heart failure was significantly lower in the Treatment Arm (0.32) compared to the Control Arm (0.44) (28% reduction, p=0.0002), and the device exhibited an excellent safety and performance profile. All pre-specified secondary endpoints were also achieved."*

The results of this trial were submitted in the company's Premarket Application to the FDA on December 14th, 2010. The reviewing panel raised concerns about potential

bias in the efficacy analysis as well as concerns about the efficacy of the device in some subpopulations and the device was not approved following the first Cardiovascular Devices Panel meeting at which it was considered in December, 2011 (Husten, 2013).

In 2013, CardioMEMS continues to pursue FDA approval, having completed an additional follow-up study. However, the 2013 reviewers have reported that they still find it "difficult to draw conclusions based on unrandomized and unblinded followup data of a segment of the original trial population" (Husten, 2013). The fate of the Champion device remains undecided nearly three years after its original PMA filing.

Case 2: Requests for Additional Information Following PMA Submission and Completion of Pivotal Trials

EnteroMedics is a medical device company that develops neuroscience based technologies to treat obesity and metabolic disease. Its VBLOC therapy device is intended to help obese patients lose weight more comfortably by intermittently blocking the vagus nerve, which resides just above the intersection of the stomach and esophagus. This is accomplished by two small laparoscopically implanted electrodes that are put in contact with the vagus nerve.

EnteroMedics completed a randomized, double-blind, sham-controlled, multicenter pivotal clinical trial of the effectiveness of the VBLOC device in 239 patients at 10 sites (The control group received a non-functional device during the trial period). In February of 2013, EnteroMedics announced a statistically significant and clinically meaningful effect of the device on weight loss and "an excellent safety profile" of the device in trials. The results suggested excess weight loss of approximately 25 percent among treated patients, with over half of patients achieving at least 20 percent excess weight loss. Based on the results of the pivotal trial, a Premarket Application was submitted to the FDA.

In late September of 2013, EnteroMedics reported that it had "received a formal response...from the Food and Drug Administration (FDA) with regard to its Premarket Approval Application (PMA) for approval of the Maestro Rechargeable System as a treatment for obesity." According to a press release by EnteroMedics, "the response contains follow-up questions related to the application, pertaining primarily to device testing and clinical data, including training programs for users and a post approval study." (EnteroMedics, September 24, 2013) EnteroMedics said that it would respond to the FDA's follow-up questions within the weeks immediately following the communication. The Premarket Application for the VBLOC device is still under review at the FDA and EnteroMedics hopes for an approval in 2014.

Case 3: Emerging Classes of Medical Technology and Procedural Uncertainty

There are several classes of emerging medical technologies that do not yet have formal regulatory approval pathways in place for entering U.S. Markets. Two examples (biosimilars and cellular and gene therapies) are presented below. These can be thought of as *extreme* cases of procedural uncertainty – that is, the complete absence of rules for regulating these new technologies has meant that they are not yet available to patients in the United States.

I. Biosimilars

Biologics are a group of large, complex and heterogeneous proteins derived from living organisms, which are often the primary component of vaccines and cancer therapies. Because they are more complex and derived from living cells, biologic drugs are regulated separately from chemical drugs. Bio*similars* or follow-on biopharmaceuticals differ from chemical drug generics in terms of their physical characteristics as well as in how they are regulated. Generic versions of chemically manufactured small molecule drugs are based on bioequivalence – that is, containing the same quantity of active substance(s) as the reference product. These generic drugs can be used in the same dose to treat the same disease with equal expected efficacy. Biosimilars, on the other hand, are much larger molecules and follow-on products are based on *similarity* to the reference product, such as biologically manufactured recombinant proteins (Manheim et. al. 2006; Rovira et. al., 2011).

At present, the FDA's Center for Biologics Evaluation and Research (CBER) is considering how to regulate follow-on biological products and the United States has no established industry for biosimilars. Europe, in contrast, has had biosimilars since 2006 following the establishment of a formal regulatory pathway for their approval. On February 9, 2012, the FDA issued three draft guidance documents on biosimilar product development and the FDA is currently accepting public comments these documents. There remains a fair amount of debate as to what FDA will require of biosimilars – in particular with respect to requirements to prove interchangeability (GaBI, 2012). In a February 2012 editorial, The Lancet urged the FDA "to integrate the data, experience, and lessons learned by the European Medicines Agency, which has approved a dozen biosimilars since 2006." At

present, the absence of regulatory processes for approving biosimilars has meant that patients in the United States have no access to these products.

II. Cellular and Gene Therapies

Several new cellular and gene therapies also provide examples of extreme procedural uncertainty. As is the case for biologics, the applications of cellular and gene therapies are regulated by the FDA's Center for Biologics Evaluation and Research (CBER). Specific products and applications, in turn, are typically regulated following the publication of, and in accordance with, CBER guidance[1] documents. As a corollary, the absence of CBER guidance on a specific therapeutic application typically means that it is unavailable to U.S. patients.

An example is that of retinal ganglion cell gene therapy for visual system repair. In this application, the cells in the retina are genetically modified using viral vectors in order to benefit patients with certain inherited degenerative conditions that compromise visual function (Hellström and Harvey, 2011). Several recent clinical trials have demonstrated that genetic modification can be of meaningful therapeutic benefit to patients and there is now evidence for the long-term expression of genes delivered through the vector, suggesting extended therapeutic effects of the therapy, following a single treatment/dose. However, clinical trials to-date have been heterogeneous in their use of viral vs. non-viral gene therapy vectors and even within viral vector therapies, multiple vectors have been studied in clinical trials (Hellström and Harvey, 2011). In the absence of FDA guidance on the regulation of such therapies and despite evidence of their effectiveness, no rental repair gene therapies are currently approved by the FDA for use outside of clinical trials.

[1] All cellular and gene therapy guidance documents are available at http://www.fda.gov/BiologicsBloodVaccines/guidanceComplianceRegulatoryInformation/Guidances/CellularandGeneTherapy/default.htm

A.2 Approval Regulation Given a Farsighted Regulator

The first model tested in Section 5 is an extension of Carpenter et. al.'s (2010), model of the FDA drug approval process. In this model, drugs are indexed by i, diseases by j, and firms by k. I generalize this model to apply to multiple categories of medical technology products (e.g. drugs, devices, and others) and a common regulator, the FDA. New products can then be characterized by two parameters:

1. γ_{ij} (where $0 < \gamma_{ij} \leq 1$) is the *curing probability* of the product. Assume γ_{ij} is fixed and known with certainty throughout agency's decision problem.

2. μ_i is the *danger* of the drug or the expected number of people who will be harmed or killed by it over an interval of time, which can be normalized to 1 such that μ_i can be thought of as the rate of harming consumers. The greater is μ_i, the more its approval will harm the regulator's reputation.

   Note: for simplicity, it is helpful to assume that $\text{cov}(\mu_i, \gamma_{ij}) = 0$ – that is, danger and curing power are independently distributed.

The agency observes information (e.g. clinical trials) in which a product either harms or does not harm the consumer. Harm evolves according to a Wiener process $X_{it} = X(t)$ a linear function of underlying danger (μ) plus a random component:

$$X(t) = \mu t + \sigma z(t)$$

where μ and $\sigma > 0$ are constants and where $z(t)$ is a standard normal variable with mean 0 and variance t. Then the agency applies Bayes' Rule to the stochastic history of X(t) to learn about μ. In this model, assume that σ is the same across products, but that μ (normally distributed) differs across them and has a mean, m and variance s. Then, Carpenter et. al. (2010) note that for any product review of time t and accumulated harm $X(t) = x$, $[x, t]$ constitutes a sufficient statistic for the agency's problem.

Bayseian estimates of μ are then:

$$PosteriorMean \equiv E_{xt}(\mu) = \hat{\mu} = \frac{m/s + x/\sigma^2}{1/s + t/\sigma^2}$$

And

$$PosteriorVariance \equiv S(t) = \frac{1}{1/s + t/\sigma^2}$$

Where the posterior variance can be thought of as the FDA's uncertainty about μ, the harm the product may induce (Carpenter et. al., 2010).

Approval Payoff

Scholars of the political economy of FDA drug approvals have found that the FDA may be more responsive to the demands of lobby groups representing (potential) drug consumers, such as cancer or AIDS organizations (Olson, 1995; Carpenter 2002; Carpenter et. al., 2010) and that individual firms may also exert pressure on the FDA. Once can think of a general model of payoff for the regulator as follows:

$$A_{ijk} = g(\gamma_N, N_J, \rho_K, \theta_J, \chi)$$

where:

- γ_N is the curing probability, as noted above
- $N_J - 1$ is the number of products in the same product category that have already applied for FDA approval
- ρ_K is the "political clout" of the submitting firm, K
- θ_J is a disease-specific factor that may represent the disease's prevalence and/or the strength of its political lobby[2]
- χ is a set of relevant specialty area and time-varying effects that may affect the payoff associated with product approval

Agency Decision-Making

[2]Note that in contrast to Carpenter et. al. (2009), I am not interested in identifying the disease-specific effects *per se*. Rather, knowing that they may exist for some subset of illnesses, I control for them when estimating other model coefficients.

As in Carpenter et. al. (2010), I can write an agency objective function, in which the Agency wants to maximize its approval payoff given information about $\hat{\mu}$:

$$\max Ee^{\delta(T)}\{A - E_{\mu,t}\int_t^{\infty} e^{-\delta(y-t)}\mu*(y,\omega)dy\}$$

where δ is the discount factor, T is approval time, $\mu*$ is the agency's estimate of danger at the optimal stopping time for clinical trials and other data collection, ω represents an elementary event in probability space Ω and y is a variable of integration.

Early Entrant Advantages

In a model like the one used above, early entrant protection should be observed within a product category. All else equal, this is a result of a regulator making approval decisions in the present while expecting a discounted pipeline of future innovations. For example, given two products $i = N$ and $i = N + 1$ with the same expected levels of danger ($\mu_N = \mu_{N+1}$) and curing probability, then the Nth product should have a shorter expected approval time. This result is, of course, in the absence of regulatory uncertainty, which is introduced and discussed in Section 3.

A.3 Product Code Examples

Examples of unique cardiovascular products:

TITLE 21--FOOD AND DRUGS
CHAPTER I--FOOD AND DRUG ADMINISTRATION
DEPARTMENT OF HEALTH AND HUMAN SERVICES
SUBCHAPTER H--MEDICAL DEVICES
PART 870 CARDIOVASCULAR DEVICES

Subpart A--General Provisions
§ 870.1 - Scope.
§ 870.3 - Effective dates of requirement for premarket approval.
§ 870.9 - Limitations of exemptions from section 510(k) of the Federal Food, Drug, and Cosmetic Act (the act).

Subpart B--Cardiovascular Diagnostic Devices
§ 870.1025 - Arrhythmia detector and alarm (including ST-segment measurement and alarm).
§ 870.1100 - Blood pressure alarm.
§ 870.1110 - Blood pressure computer.
§ 870.1120 - Blood pressure cuff.
§ 870.1130 - Noninvasive blood pressure measurement system.
§ 870.1140 - Venous blood pressure manometer.
§ 870.1200 - Diagnostic intravascular catheter.
§ 870.1210 - Continuous flush catheter.
§ 870.1220 - Electrode recording catheter or electrode recording probe.
§ 870.1230 - Fiberoptic oximeter catheter.
§ 870.1240 - Flow-directed catheter.
§ 870.1250 - Percutaneous catheter.
§ 870.1270 - Intracavitary phonocatheter system.
§ 870.1280 - Steerable catheter.
§ 870.1290 - Steerable catheter control system.
§ 870.1300 - Catheter cannula.
§ 870.1310 - Vessel dilator for percutaneous catheterization.
§ 870.1330 - Catheter guide wire.
§ 870.1340 - Catheter introducer.
§ 870.1350 - Catheter balloon repair kit.
§ 870.1360 - Trace microsphere.
§ 870.1370 - Catheter tip occluder.
§ 870.1380 - Catheter stylet.

Example product code definition for an implantable pacemaker pulse generator:

[CITE: 21CFR870.3610]

TITLE 21--FOOD AND DRUGS
CHAPTER I--FOOD AND DRUG ADMINISTRATION
DEPARTMENT OF HEALTH AND HUMAN SERVICES
SUBCHAPTER H--MEDICAL DEVICES

PART 870 -- CARDIOVASCULAR DEVICES

Subpart D--Cardiovascular Prosthetic Devices

Sec. 870.3610 Implantable pacemaker pulse generator.

(a) *Identification.* An implantable pacemaker pulse generator is a device that has a power supply and electronic circuits that produce a periodic electrical pulse to stimulate the heart. This device is used as a substitute for the heart's intrinsic pacing system to correct both intermittent and continuous cardiac rhythm disorders. This device may include triggered, inhibited, and asynchronous modes and is implanted in the human body.

(b) *Classification.* Class III (premarket approval).

(c) *Date PMA or notice of completion of PDP is required.* A PMA or notice of completion of a PDP is required to be filed with the Food and Drug Administration on or before September 20, 2012, for any implantable pacemaker pulse generator device that was in commercial distribution before May 28, 1976, or that has, on or before September 20, 2012, been found to be substantially equivalent to any implantable pacemaker pulse generator device that was in commercial distribution before May 28, 1976. Any other implantable pacemaker pulse generator device shall have an approved PMA or declared completed PDP in effect before being placed in commercial distribution.

[45 FR 7907, Feb. 5, 1980, as amended at 52 FR 17736, May 11, 1987; 77 FR 37576, June 22, 2012]

Appendix B

Appendix to Chapter 2

B.1 Determining Conflict of Interest and Eligibility for Advisory Committee Participation[1]

1: Is the subject matter of the meeting a "particular matter?"

2: Will the particular matter have a direct and predictable effect on the financial interest(s) of any organization?

3: Identify potentially affected products/organizations and request that the employee complete the financial disclosure form

4: Does the employee, or persons/organizations whose interests are imputed to him, have a financial interest in one or more of the potentially affected products and/or organizations?

5: Will the particular matter have a direct and predictable effect on the financial interest of the employee and/or persons/organizations whose interests are imputed to him?

6: After applying applicable regulatory exemptions, does the employee or persons/organizations whose interests are imputed to him have a disqualifying financial interest?

7: Are There disqualifying financial interests for which a waiver would not be considered?

8: Is the combined value of the employee's personal disqualifying financial interests and

[1]Source: FDA (2008)

those of his spouse and minor children $50,000 or Less[2]?

9: Is the individual's participation necessary to afford the advisory committee essential expertise?

10a: If the individual is a special government employee, does the need for the individual's services outweigh the potential for a conflict of interest created by the interest involved?
10b: If the individual is a regular government employee, is the financial interest not so substantial as to be deemed likely to affect the integrity of the services provided by that individual?

11: Waiver may be recommended if consistent with waiver cap[3].

---

[2]If the combined value of these disqualifying financial interests is greater than $50,000, the member would not ordinarily be considered for a waiver and would not participate in the advisory committee meeting.

[3]Provided that the applicable waiver cap would not be exceeded, staff may recommend that a waiver for the individual be granted. FDA has discretion to issue limited waivers under 18 U.S.C. 208 and under section 712(c)(2)(C) of the Act; e.g., by limiting participation to non-voting. If staff decides to recommend that a waiver be granted, they should determine which type of waiver(s) (including any recommended limitations) is appropriate to recommend to FDA officials who will review and decide whether to approve the waiver.

B.2 Example Meeting Roster

Panel Roster
Circulatory System Devices Panel Meeting
Edwards SAPIEN Transcatheter Heart Valve P110021
June 13, 2012

Name	Affiliation	Role
Warren Laskey, MD	Univ New Mexico School of Medicine Albuquerque, NM	Temporary Panel Chair
David J. Slotwiner, MD	Long Island Jewish Medical Center New Hyde Park, NY	Voting Member
David C. Naftel, PhD	University of Alabama at Birmingham Birmingham, AL	Voting Member
E. Magnus. Ohman MB, F.R.C.P.I., F.A.C.C.	Duke University Medical Center Durham, NC	Voting Member
Valluvan Jeevanandam MD	University of Chicago Chicago, IL	Voting Member
John C. Somberg, MD	Rush University Medical Center Lake Bluff, IL	Voting Member
Richard A. Lange M.D.	University of Texas San Antonio, TX	Temporary Voting Member
Jeffrey S. Borer, M.D.	State Univ. of New York, Downstate Medical Ctr. New York, NY	Temporary Voting Member
Gregory J. Dehmer, M.D.	Scott & White Healthcare, Texas A&M University Temple, TX	Temporary Voting Member
George W. Vetrovec, M.D.	Medical College of Virginia, Richmond, VA	Temporary Voting Member
David C. Good, M.D	Penn State Milton S. Hershey Medical Center, Hershey, PA	Temporary Voting Member
David E. Kandzari, M.D.	Piedmont Heart Institute Atlanta, GA	Temporary Voting Member
Brett C. Sheridan, M.D.	University of North Carolina at Chapel Hill Chapel Hill, NC	Temporary Voting Member
Marc R. Katz, M.D., M.P.	St Mary's Hospital Richmond, VA	Temporary Voting Member
Elizabeth B. Patrick-Lake, M.F.S.	PFO Research Foundation	Patient Representative
Burke T. Barrett, B.A., B.S., M.B.A.	Vice President of Regulatory & Clinical Affairs, CardioFocus, Inc.	Industry Representative
Kristine R. Mattivi, Ms, Pt	Colorado Foundation for Medical Care Englewood, CO	Consumer Representative
Jamie Waterhouse	Food and Drug Administration Silver Spring, MD	Designated Federal Officer

B.3 Example Meeting Agenda

DEPARTMENT OF HEALTH & HUMAN SERVICES Public Health Service

AGENDA

GENERAL AND PLASTIC SURGERY DEVICES PANEL
of the
MEDICAL DEVICES ADVISORY COMMITTEE

Hilton Washington DC North/Gaithersburg
620 Perry Parkway, Gaithersburg, MD

June 21, 2012

Panel Chairperson
Joseph LoCicero III, M.D.

Designated Federal Officer
Natasha G. Facey

08:00	**a.m.**	**Call to Order** **Panel Introductions** **Conflict of Interest** **PMA# P110014 Dune Medical Margin Probe System, by Dune Medical Devices, Inc.**
08:10	**a.m.**	**Sponsor Presentation**
09:40	**a.m.**	**Q&A Sponsor**
10:00	**a.m.**	**Break**
10:15	**a.m.**	**FDA Presentation**
11:45	**p.m.**	**Q&A FDA Presentation**
12:05	**p.m.**	**Lunch**
01:05	**p.m.**	**Open Public Hearing***
02:05	**p.m.**	**Panel Deliberations**
03:05	**p.m.**	**Break**
03:20	**p.m.**	**FDA Questions**
05:30	**p.m.**	**FDA and Sponsor Summations**
05:40	**p.m.**	**Panel Vote**
06:00	**p.m.**	**Adjournment**

*** Open Public Hearing –** Interested persons may present data, information, or views, orally or in writing, on the issue pending before the panel. Scheduled speakers who have requested time to address the panel will speak at this time. After they have spoken, the Chair may ask them to remain if the panel wishes to question them. Then the Chair will recognize unscheduled speakers as time allows. Only the panel may question speakers during the open public hearing.

Last updated 6/18/2012 1

Appendix C

Appendix to Chapter 3

C.1 Clinical Vignettes and Response Options for Patients, Cardiologists and Primary Care Physicians

Panel I: Patient Questions

SCENARIO 1- Questions relating to less-severe cardiac care preferences: *Suppose you noticed a mild but definite chest pain when walking up stairs....Suppose you went to your regular doctor for that chest pain and your doctor did not think you needed any special tests but you could have some tests if you wanted.*
a) If the tests did not have any health risks, do you think you would probably have the tests or probably not have them?
a - have tests
b - not have tests
b) Suppose your doctor told you he or she did not think you needed to see a heart specialist, but you could see one if you wanted. Do you think you would probably ask to see a specialist, or probably not see a specialist?
a - see specialist
b - not see specialist

SCENARIO 2 - Questions relating to end of life care preferences: *The next set of questions are about care a patient may receive during the last months of life. Remember, you can skip any question you don't want to answer. Suppose that you had a very serious illness. Imagine that no one knew exactly how long you would live, but your doctors said you almost certainly would live less than 1 year.*
a) If you reached the point at which you were feeling bad all the time, would you want drugs that would make you feel better, even if they might shorten your life?
a - yes: drugs
b - no
b1) If you needed a respirator to stay alive, and it would extend your life for a week, would you want to be put on a respirator?
b2) If it would extend your life for a month, would you want to be put on a respirator?
a - yes: respirator
b - no

Answers other than "yes" or "no" (e.g., "not concerned" or "I dont know") are treated as missing data. Item non-response was less than 1% among eligible respondents.

Panel II: Physician Questions

In the next set of questions, you will be presented with brief clinical descriptions for three different patients. For each, you will be asked a series of questions regarding how you would be likely to treat that patient were he or she in your care.

PATIENT A - CARDIOLOGIST - *For this question, think about a patient with stable angina whose symptoms and cardiac risk factors are now well controlled on current medical therapy. In general, how frequently do you schedule routine follow-up visits for a patient like this?*

*Answer recorded in number of months

PATIENT A - PCPs: *In general, how frequently do you schedule routine follow-up visits for a patient with well-controlled hypertension?*

*Answer recorded in number of months

PATIENT B: *A 75 year old man with severe (Class IV) congestive heart failure from ischemic heart disease, is on maximal medications and has effective disease management counseling. His symptoms did not improve after recent angioplasty and stent placement and CABG is not an option. He is uncomfortable at rest. He is noted to have frequent, asymptomatic nonsustained VT on cardiac monitoring. He has adequate health insurance to cover tests and medications. At this point, for a patient presenting like this, how often would you arrange for each of the following?*

CARDIOLOGIST SURVEY

a - Repeat angiography
b - Initiate antiarryghmic therapy
c - Recommend an Implantable Cardiac Defibrilator (ICD)
d - Recommend biventricular pacemaker for cardiac resynchronization
e - Initiate or continue discussions about palliative care

POSSIBLE RESPONSES

1 Always/Almost always
2 Most of the time
3 Some of the time
4 Rarely
5 Never

PATIENT C: *An 85 year old male patient has severe (Class IV) congestive heart failure from ischemic heart disease, is on maximal medications, and is not a candidate for coronary revascularization. He is on 2 liters per minute of supplemental oxygen at home. He presents to your office with worsening shortness of breath and difficulty sleeping due to orthopnea. Office chest xray confirms severe congestive heart failure. Oxygen saturation was 85% and increased to 94% on 4 liters and the patient is more comfortable. He has adequate health insurance to cover tests and medications. At this point, for a patient presenting like this, how often would you arrange for each of the following?*

PCP and CARDIOLOGIST SURVEY

a - Allow the patient to return home on increased oxygen and increased diuretics
b - Admit to the hospital for aggressive diuresis (not to the ICU/CCU)
c - Admit to the ICU/CCU for intensive therapy and monitoring
d - Place a pulmonary artery catheter for hemodynamic optimization
e - Recommend biventricular pacemaker for cardiac resynchronization
f - Initiate or continue discussions about palliative care

POSSIBLE RESPONSES (both surveys)

1 Always/Almost always
2 Most of the time
3 Some of the time
4 Rarely
5 Never

C.2 Regression Estimates of Ln Medicare Expenditures in the Last Two Years (Cardiologists Only)

	Cardiologists					
	(1)	(2)	(3)	(4)	(5)	(6)
Cowboy Ratio, Cardiologists	0.1825*	0.1831**	0.2460***	0.1726**	0.2391***	
	(0.1027)	(0.0864)	(0.0883)	(0.0857)	(0.0868)	
Comforter Ratio, Cardiologists	-0.1261	-0.0400	-0.0016	-0.0449	-0.0111	
	(0.1100)	(0.0848)	(0.0903)	(0.0852)	(0.0862)	
Followup Low, Cardiologists		-0.6662***	-0.5460***	-0.7836***	-0.6951***	
		(0.1062)	(0.1373)	(0.1648)	(0.1691)	
Followup High, Cardiologists		0.5323***	0.5265***	0.5333***	0.5292***	
		(0.1077)	(0.1027)	(0.1062)	(0.1017)	
Have Unneeded Tests			0.2587		0.2705	0.2343
			(0.1925)		(0.2066)	(0.2302)
See Unneeded Cardiologist			0.2674		0.2894	0.2411
			(0.1834)		(0.1791)	(0.2083)
Aggressive Preferences Patient Ratio				-0.2385	-0.2539	-0.2870
				(0.3013)	(0.3044)	(0.4397)
Comfortable Preferences Patient Ratio				-0.0628	-0.1267	0.0120
				(0.1488)	(0.1482)	(0.1559)
N	64	64	64	64	64	64
R^2	0.0535	0.4073	0.4446	0.4119	0.4530	0.0406

* p<0.10, ** p<0.05, *** p<0.01

2-year End-of-Life Spending is in natural log form and is price, age, sex and race adjusted. Results shown are for the 64 Hospital Referral Regions (HRRs) in which we have at least 3 patients and 3 cardiologists surveyed. All regressions include a constant and control for the fraction of primary care physicians in the sample. Respondent data is adjusted for race, sex and age. Survey sampling weights take into account differences in the number of physician observations per HRR.

C.3 Regression Estimates of Ln Medicare Expenditures in the Last Two Years (PCPs Only)

	PCPs					
	(1)	(2)	(3)	(4)	(5)	(6)
Cowboy Ratio, PCPs	0.6689***	0.5476***	0.4773***	0.5383***	0.4728***	
	(0.1687)	(0.1416)	(0.1333)	(0.1251)	(0.1223)	
Comforter Ratio, PCPs	-0.2489*	-0.2436**	-0.2104*	-0.1987**	-0.1724*	
	(0.1380)	(0.1137)	(0.1157)	(0.0944)	(0.0972)	
Followup Low, PCPs		-0.4729*	-0.4639*	-0.5905**	-0.5682*	
		(0.2754)	(0.2706)	(0.2938)	(0.2930)	
Followup High, PCPs		0.9091*	0.9918*	0.8640*	0.9333*	
		(0.5359)	(0.5386)	(0.5135)	(0.5064)	
Have Unneeded Tests			-0.2231		-0.1341	-0.2371
			(0.3258)		(0.3037)	(0.3941)
See Unneeded Cardiologist			0.4045*		0.4135**	0.7422**
			(0.2154)		(0.2046)	(0.3350)
Aggressive Preferences Patient Ratio				-0.8012	-0.7712	-0.6638
				(0.6915)	(0.6460)	(0.9768)
Comfortable Preferences Patient Ratio				-0.2719	-0.3058	-0.3864
				(0.2521)	(0.2739)	(0.3348)
N	64	64	64	64	64	64
R^2	0.3430	0.4613	0.4888	0.4852	0.5126	0.1290

* p<0.10, ** p<0.05, *** p<0.01

2-year End-of-Life Spending is in natural log form and is price, age, sex and race adjusted. Results shown are for the 64 Hospital Referral Regions (HRRs) in which we have at least 3 patients and 3 cardiologists surveyed. All regressions include a constant and control for the fraction of primary care physicians in the sample. Respondent data is adjusted for race, sex and age. Survey sampling weights take into account differences in the number of physician observations per HRR.

C.4 Expanded Regression Estimates of Ln Medicare Expenditures in the Last Two Years

Combined Sample: Cardiologists and PCPs					
	(1)	(2)	(3)	(4)	(5)
Cardiologists per 100k				0.0390**	0.0499***
				(0.0165)	(0.0156)
Cowboy Ratio, All Doctors	0.6080***	0.5212***	0.5930***	0.5115***	0.3942***
	(0.1345)	(0.1232)	(0.1385)	(0.1252)	(0.1340)
Comforter Ratio, All Doctors	-0.3098***	-0.2876**	-0.3018***	-0.2289*	-0.1998*
	(0.1093)	(0.1144)	(0.1134)	(0.1277)	(0.1022)
Follow-Up Low, All Doctors	-0.3481	-0.1154	-0.3931	-0.1235	0.0410
	(0.2246)	(0.2165)	(0.2642)	(0.2010)	(0.2364)
Follow-Up High, All Doctors	0.9409***	0.7724***	1.0192***	0.7609***	0.5836*
	(0.1945)	(0.2239)	(0.2312)	(0.2169)	(0.2951)
(mean) Fraction Capitated		0.1622			0.2325*
		(0.1313)			(0.1245)
(mean) Fraction Medicaid		-0.5005*			-0.3495
		(0.2976)			(0.2288)
Base = (mean) Solo or 2-person Practice		-			-
(mean) Single/Multi Speciality Group Practice		-0.2432			-0.2381
		(0.1739)			(0.1580)
(mean) Group/Staff HMO or Hospital-Based Practice		-0.1735			-0.4342*
		(0.2104)			(0.2221)
(mean) Responds to Patient Expectations			0.0785		-0.0723
			(0.1415)		(0.1074)
(mean) Responds to Colleague Expectations			-0.1456		-0.0044
			(0.1208)		(0.0967)
(mean) Responds to Referrer Expectations			-0.0772		-0.1260
			(0.1690)		(0.1311)
(mean) Responds to Malpractice Concerns			0.1298		0.2344*
			(0.1830)		(0.1295)
N	64	64	64	64	64
R^2	0.6008	0.6442	0.6112	0.6641	0.7310

* p<0.10, ** p<0.05, *** p<0.01

2-year End-of-Life Spending is in natural log form and is price, age, sex and race adjusted. Results shown are for the 64 Hospital Referral Regions (HRRs) in which we have at least 3 patients and 3 cardiologists surveyed. All regressions include a constant and control for the fraction of primary care physicians in the sample. Respondent data is adjusted for race, sex and age. Survey sampling weights take into account differences in the number of physician observations per HRR.

This figure provides additional visual evidence of the relationship between cowboy status and recommended follow-up frequency for the HRRs with the greatest number of respondents; a point that is further out on the scale corresponds to a larger fraction of physicians.

Figure C.1: *Radar Plots of Select High Follow-up Frequency and Cowboy Prevalence by HRR*