Digital Innovation in a Regulated Industry: Evidence from Software-Driven Medical Devices

(Preliminary and incomplete: please do not circulate)

Cirrus Foroughi and Ariel Dora Stern*

December 18, 2017

Abstract

Does technological opportunity enable the rise of new entrants or reinforce the position of incumbents? Research on this classic topic rarely considers the unique features of regulated industries. We offer a novel approach to this question in the context of regulated medical technology, where the introduction of software is of growing importance and has created fresh opportunities for new product development. Pioneering a new application of supervised document classification, we consider over 37,000 new medical devices that came to market in the United States from 2002-2016 in order to identify predictors of product digitization in this industry. We consider the relative importance of several key factors such as geographic and within-firm capabilities, as well as the role of financial resources. We find that location in regions of concentrated expertise and prior firm commercialization experience reinforce one another in predicting digital innovation. While venture capital appears to play a role in supporting innovative entrants, a closer look at the data suggests that the relationship is entirely driven by selection on other variables that predict digitization. We conclude that this regulated setting favors firms coming from a position of incumbent geographic and/or experiential advantage.

Keywords: Innovation; Digitization; Medical Technology; Software

*Harvard Business School, Boston, Massachusetts, 02163 (astern@hbs.edu). We are grateful to Carliss Baldwin, Moritz Fischer, Shane Greenstein, Joachim Henkel, Kyle Myers, Ivan Png, Scott Stern, Eric von Hippel, and participants at the AIEA-NBER Conference on Innovation and Entrepreneurship, the Weatherhead School of Management, Bocconi, Carnegie Mellon, UCSD, UCSB, the University of Munich/the Max Planck Institute for Innovation and Competition, the American-European Health Economics Study Group, the Harvard Science-Based Business Initiative, and the NBER Productivity Lunch for helpful suggestions. Lynne Chou O'Keefe provided generous insight on the investor's perspective on digital health and Melissa Ouellet and Paul Jonak provided excellent research assistance.

"The heart failure patient will have his state of hydration and vital signs monitored through sensors embedded in a watch, a wristband, or a stick-on device. The diabetic or the kidney-failure patient who needs periodic blood monitoring will be able to prick a finger at home (assuming that test still requires a drop of blood; many of today's blood tests will be replaced by sophisticated skin sensors). The specimen will be processed in seconds through a smartphone attachment, and the result will be automatically entered into the electronic record"

-Robert Wachter, on the future of health care in "The Digital Doctor"

1 Introduction

In recent years, major industries ranging from manufacturing and inventory management to music and health care have undergone a "digital transformation," in which key aspects of both day-to-day business and new frontiers of product development have migrated to a primarily digital (i.e. software-driven) context. In health care, medical technologies – the devices and equipment used in treating and caring for patients – have become increasingly digitized, as software and networking capabilities have become integrated into a growing number and share of new products. Today, medical devices can interface with proprietary software for tasks ranging from simple blood pressure monitoring to the processing and analysis of digital radiology images. Digital medical technology is now commonplace and its use is inescapable for health care professionals: a recent report found that U.S. hospitals use an average of 10 to 15 "connected" digital devices per bed (Newman, 2017).

Like other industries undergoing meaningful changes, this scenario raises questions about how new technological opportunities affect both new entrants and incumbents. Unlike some other settings, however, the medical device industry is strictly regulated: the U.S. Food and Drug Administration (FDA) is the sole regulatory authority with the power to grant marketing approval for medical devices in the United States. Taken together, the complex approval process and its associated costs, combined with other resource-intensive features

of medical device development, suggest that smaller and less experienced firms are likely to face higher costs of new product commercialization than seasoned incumbents. We explore this phenomenon in a novel dataset and find evidence to suggest that the existence of entry regulation favors experienced and advantageously located firms and is likely to perpetuate their commercial leadership.

Health care spending now comprises nearly 18% of the U.S. economy (CMS, 2016), creating substantial opportunity for innovators to grow digital businesses in the industry. In recent years, there has been dramatic growth in funding for "digital health" (Tecco and Zweig, 2017) with notable private[1] and public[2] initiatives emerging to fund research and investment in the space. Digital health is broadly defined to include companies and products at the intersection of healthcare and technology.[3] The digital health space includes health care IT and information systems, as well as a host of companies that build and sell technologies like wireless sensors, software-enabled diagnostic and imaging devices, and artificial intelligence and machine learning software programs with health care applications.

This study is set in the context of the increasingly software-driven medical device industry, an important subset of the digital health ecosystem. Surprisingly, the growth of software in medical devices has not yet been characterized across products or firms, nor have its implications for innovation in this sector been rigorously studied. We therefore begin by documenting a set of trends relating to the growth of software in medical devices. We then explore the factors that predict which types of firms are more likely to pursue software-driven innovations, and under what circumstances.

The term "digitization" technically refers to the action or process of moving from analog to digital information or data. Common examples include the digital use and storage of images, video, text, and other information. The "digital transformation" of industries

[1]For example, Rock Health describes itself as "the first venture fund dedicated to digital health" (https://rockhealth.com/about/)

[2]For example, the Massachusetts Digital Healthcare Initiative was launched by Governor Charlie Baker in January, 2016 as "a comprehensive public-private partnership that will advise the administration on the future of the Commonwealth's digital healthcare industry." (http://www.mass.gov/governor/press-office/press-releases/fy2017/governor-establishes-mass-digital-healthcare-council.html)

[3]https://rockhealth.com/what-digital-health-is-and-isnt/

broadly refers to the adoption or increase in use of digital or computer technology by an organization, industry, country, or other actor(s). Generally speaking, the phrase "digital transformation" is used to describe the substantial changes that come about in a business setting due to the growth of software and networking technology.[4] This study takes advantage of the *digitization* of a host of regulatory documents to study the *digital transformation* of the medical device industry in the early 21st century.

Previous studies have highlighted the importance of software and digitization in determining both firm innovation (Arora, Branstetter, and Drev, 2013; Branstetter, Drev, and Kwon, 2015) and performance (Brynjolfsson and McElheran, 2016a; Brynjolfsson and McElheran, 2016b). Our study builds in many ways on the literature linking software and networking capabilities to innovative activity; however, unlike these previous studies, our primary measure of innovation goes beyond counts of (pre-market) patenting activity to assess the precise and complete set of new products ultimately brought to market. Because new product commercialization in the medical device industry typically occurs well after patenting, this study characterizes software-driven innovation at the very end of the innovation pipeline, focusing on the final phase of development in the R&D process. Furthermore, an important distinction from much of the prior literature is that we empirically model digital innovation as a *dependent* variable, whereas others have typically treated the use of software as an independent variable. In doing so, we identify the set of characteristics and circumstances that predict which firms are most likely to bring software-driven medical technologies to market.

The first contribution of this paper is to quantify and describe the growth of software in new medical technologies. Using detailed, comprehensive data over the years 2002-2016, we present descriptive statistics across medical specialty areas, different types of innovator firms, and over time. The second contribution of the paper is to shed light on determinants of digital innovation in a regulated setting. We consider if and how geography, firm capabilities, and financial resources matter for digital innovation. We find that firms operating

[4]Colloquially, however, the term "digitization" (and in Europe, the related term "digit*al*ization") will often be used to describe the phenomenon of "digital transformation" more broadly.

in regions of concentrated expertise are more likely to engage in digital device development. Further, we find strong evidence for the importance of existing firm capabilities (here, prior software-specific product class commercialization experience), suggesting that prior relevant experience, both regulatory and technological, predicts subsequent innovation activity. While venture capital (VC) funding appears to play a role in supporting innovative entrants, a closer look at the data suggests that the relationship is entirely driven by selection on other variables that predict digitization. We conclude that this regulated setting favors firms coming from a position of incumbent geographic and/or experiential advantage.

We proceed as follows: Section 2 provides background on FDA medical device regulation as it relates to the process of bringing new products to market along with a review of related literature. Section 3 presents a conceptual framework for thinking about the integration of software into new products with entry regulation and lays out testable hypotheses. Section 4 describes the data employed in the empirical analyses. Section 5 presents estimation results and Section 6 concludes.

2 Background

The FDA is the sole regulatory authority with the power to grant marketing approval for medical devices in the United States. An agency within the U.S. Department of Health and Human Services, the FDA regulates over two trillion dollars' worth of products annually, including all medical technologies (Babiarz and Pisano, 2008). The FDA is organized into centers, each of which focuses on one type of product. Medical devices, including radiation-emitting products such as X-ray and ultrasound machines, are regulated by the Center for Devices and Radiological Health (CDRH).[5] Within the CDRH, the Office of Device Evaluation reviews new products, which since 1976, have been governed by the Medical Device Amendments (MDA) to the Federal Food, Drug, and Cosmetic Act of 1938.

Devices are wide-ranging in their complexity and the risk that they pose to patients.

[5]Other centers are responsible for other product categories. For example, drugs are regulated by the Center for Drug Evaluation and Research (CDER) and biologics are regulated by the Center for Biologics Evaluation and Research (CBER).

They include everything from low-risk products such as stethoscopes and tongue depressors, to moderate-risk products, such as hearing aids and blood pressure monitors, to complex, high-risk products such as cardiac pacemakers and replacement heart valves. While devices of the lowest risk are subject only to so-called "general controls" of labeling and compliance with the FDA's good manufacturing practices,[6] moderate-risk and high-risk devices must submit applications to the FDA for regulatory clearance and regulatory approval, respectively.[7] The administrative data from these regulatory processes along with each new product's formal description are made publicly available as of the time a device comes to market. These documents constitute the main source of new product data used in the analyses that follow.

As described in detail below, a growing number and share of devices now contain software. Yet despite recognition of the increasingly digital nature of medical devices,[8] the FDA does not formally or directly track the use of software and other networking technologies in its administrative databases. Because the use of software in medical devices is not tracked in official regulatory data, a quantitative assessment of digital innovation in the medical device sector is impossible based on review and analysis of administrative data alone. Therefore, this paper's first contribution is to provide a detailed description and quantification of the growth of software in new medical devices over recent years.

2.1 Moderate Risk Devices and the 510(k) Process

Moderate risk devices are approved through a process called "premarket notification," which is often referred to as the "510(k) process" – a reference to the section of the MDA that established this regulatory pathway. The 510(k) is defined as:

> *"a premarket submission made to FDA to demonstrate that the device to be marketed is at least as safe and effective, that is, substantially equivalent, to a legally marketed device...that is not subject to [the more extensive process of Premarket Approval,] PMA [,which is required for devices of the highest risk]."*

[6] https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpcd/315.cfm
[7] https://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/Overview/
[8] See, for example, FDA, 1999; FDA, 2005a; FDA, 2014; FDA, 2016

Firms submitting a 510(k) application must compare their device "to one or more similar legally marketed devices and make and support their substantial equivalency claims." [9] That is, moderate risk devices approved through this process build upon previously regulated technology and claim one or more previously marketed devices as the regulatory starting point for their new product's path to market. Additional detail on the 510(k) process can be found in Appendix A.

One important component of the 510(k) application the "510(k) Summary," a text document describing the device and published at the time of clearance. The summary includes:

> *"A description of the device such as might be found in the labeling or promotional material for the device, including an explanation of how the device functions, the scientific concepts that form the basis for the device, and the significant physical and performance characteristics of the device, such as device design, material used, and physical properties."*

It is these summaries (and their equivalents from high-risk devices) that are used to construct the text database, described below.[10] Section E of Title 21 of the Code of Federal Regulations (21CFR807.92) outlines the requirements for the "summary" in detail (see Appendix C). Further, the summary includes "a statement of the intended use of the device" and a description of the device's technological characteristics, making it a clear, standard resource for product information. An example 510(k) Summary can be seen in Exhibit 1.

[9] In this setting, the device that is the basis (comparator) for a claim of "substantial equivalence" is commonly known as the "predicate." The FDA further notes that "although devices recently cleared under 510(k) are often selected as the predicate to which equivalence is claimed, any legally marketed device may be used as a predicate." (http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/PremarketSubmissions/PremarketNotification510k/)

[10] In lieu of a "510(k) Summary" a device manufacturer may, in rare cases, submit a "510(k) Statement," a certification that the 510(k) holder will provide a copy of the 510(k) submission, with certain exclusions, to any person within 30 days of a written request. These documents also include product information about the device. In the analysis below, the vast majority of devices include a 510(k) Summary, however when Summaries are not available, 510(k) Statements are instead used in the text analysis.

2.2 High Risk Devices and the PMA Process

High-risk (Class III) devices are cleared through a process called Premarket Approval (PMA), which typically requires data from clinical trials in order to establish a device's safety and effectiveness with reasonable certainty.[11] Evidence from pivotal trials – the major studies used to support FDA/CDRH approval – is presented to the FDA as part of the PMA package (Kramer et. al., 2012) along with several other components.[12] Appendix B contains detail of the PMA process.

Like the 510(k) process, the PMA process includes a product-specific summary document, which is made publicly available at the time the device is approved by the FDA. (These summaries are used along with their moderate-risk device equivalents in the analysis below.) Much like 510(k) summaries, PMA summary documents contains information on indications for use, a detailed device description ("how the device functions, the basic scientific concepts that form the basis for the device, and the significant physical and performance characteristics of the device") and synopeses of existing studies of the device. This includes nonclinical laboratory studies as well as clinical investigations involving human subjects along with safety and effectiveness data, information about device failures and replacements, and results of statistical analyses of the clinical investigations.[13] An example PMA Summary can be seen in Exhibit 2.

2.3 Software in Medical Devices

The integration of software into medical devices is a relatively recent phenomenon. It was not until 1999 that the FDA first released guidance detailing expectations and standards for software embedded in new medical technologies (FDA, 1999). Today, medical devices that not only incorporate software, but also functionally *rely* on it, are commonplace. From imaging devices for radiology, to software-enabled insulin pumps, to implantable heart failure

[11] See: http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/PremarketSubmissions/PremarketNotification510k/

[12] Detail of the PMA review process can be found at: http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/PremarketSubmissions/PremarketApprovalPMA/ucm047991.htm

[13] https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?FR=814.20

monitors capable of wireless transmission, thousands of patients and their physicians have come to depend on software-enabled medical devices.

In fact, software and networking capabilities have become so commonplace that some have pointed out that they pose a growing security risk. A 2015 article in *Wired Magazine* profiled Dr. David Klonoff "the doctor on a quest to save our medical devices from hackers" (Zharkov, 2015) and recent incidents, such as the hacking of a commonly-used pacemaker have highlighted the security risks inherent in the ongoing digital transformation of the medical device industry (Newman, 2017). While cybersecurity is not a focus of this study, the concerns highlighted above represent another reason why understanding the drivers of innovation and the backgrounds of innovators in software-driven medical technologies is of growing relevance.[14]

Unfortunately, because the inclusion of software is not directly tracked by the FDA in its regulatory databases, its prevalence and recent growth trajectory remain unknown. The first portion of this paper is therefore dedicated to using information embedded in the text of medical device summaries to quantify the use of software in regulated medical devices and document its growth over time. Using text mining, supervised document classification, and other simple natural language processing tools for medical texts, we analyze 15 years of medical device product summaries in order to establish recent trends in the use of software in new products and consider differences across medical specialty areas.

2.4 Software in the Health Care Industry

While we are not aware of any studies of the digitization of medical devices, a small but growing body of literature in management and economics explores topics at the intersection of digitization and health care. Most prominently, a number of papers have analyzed the

[14]In recent years, the FDA has provided regulatory clarity on cybersecurity for networked medical devices (FDA, 2005a; FDA, 2014; FDA, 2016) as well as guidance on software-contained in medical devices (FDA, 2005b) and software innovation in the medical technology industry has continued apace. In early 2017, the FDA approved the first machine learning application to be used in a clinical setting, a decision that was heralded as "a big step forward for AI and machine learning in health care and industry as a whole" (Marr, 2017). Future regulatory documents are expected to include additional details of how software-driven products will be regulated moving forward (Comstock, 2016).

use and adoption of electronic medical records (EMRs), one of the primary ways in which software has impacted health care delivery in the past decade (e.g. Dranove et. al., 2014; Agha, 2014; Adler-Milstein et. al., 2014; Lee et. al. 2013). The studies have documented the ongoing adoption of EMRs at some cost to facilities and consumers along with heterogenous (and typically limited) impacts on patient outcomes.

This study is also indirectly related to a small literature on the use and adoption of software and digital data elsewhere in health care delivery. For example, Athey and Stern (2002) find that basic digitalization of emergency services (911) increased the short term survival rate of patients in cardiac distress, even after controlling for non-IT procedures and protocols that could otherwise bias the results. Other researchers have considered subtler regulatory factors such as data privacy laws (Miller and Tucker, 2016) in order to understand how new technologies are adopted and used by patients. In the context of telehealth, Dorsey and Topol (2016) describe the major trends shifting telehealth, with a particular focus on home and mobile device applications. They predict that in addition to increasing the scope of treatable conditions, digitization of health care will lead to lower costs and ultimately greater access. Yet beyond these studies, research at the intersection of digitization and health care is scant and the impacts of software and digitization on health care innovation have not been rigorously examined.

2.5 Software and Innovation in Management and Economics

A number of studies have explored the implications of the introduction of software in other settings. Branstetter, Drev, and Kwon (2015) discuss how historical data on the trajectory of innovative activity has pointed towards a "software-bias," but also note that many studies have been based on small samples of potentially unrepresentative firms. Improving on prior studies, the authors undertake a more comprehensive study, documenting the importance of software for patent productivity (patents per R&D dollar spent) and the market value of R&D investments in four industries: 1) autos and auto parts, 2) aerospace and defense, 3) medical devices, and 4) pharmaceuticals.

We build on this and other related studies by considering the growth of software and digitization in the medical device industry broadly, and by describing in detail the firm heterogeneity that arises from this process. We document the evolution of software-driven medical technology over time and consider differences across firms in terms of financial resources as well as incumbents vs. new entrants into a product space. By considering a large sample that represents the vast majority of innovative activity over the most recent 15 years, we are also confident that our results are representative of the industry.

Arora, Branstetter, and Drev (2013) offer one of the first treatments of the idea of heterogeneous effects due to software-based innovation. They find that US firms' inherent advantage in deploying new software allowed them to leapfrog Japanese IT firms as the frontier of IT innovation shifted towards increasing utilization of software between 1983 and 2004. Within the United States alone, the authors find that the relative strength of firms increased in the IT sub-sectors where software innovation intensity was the greatest. Brynjolfsson and McElheran (2016a & 2016b) further explore differential adoption of digital technology (in this case, data-driven decision-making, or "DDD") and find that adoption of DDD is more common among larger, older manufacturing plants of multi-unit firms. Interestingly, smaller firms seem to adopt digital technology later, and among these firms, the ones who do adopt gain persistent and growing performance benefits. Our results are evocative of this study in that we find notable differences between software-driven innovation in new entrants vs. established, firms with a bias towards introduction of software by those firms with the most commercialization experience.

With its focus on the medical device industry, this study also presents a new setting for studying the impact of software adoption in an industry that has historically consisted exclusively of hardware products. Iansiti and Lakhani (2014) have described the rise of the "industrial internet" in the past decade, led by firms such as General Electric, and point to the implications of this digitization trend for other industries. Indeed, there is plenty of evidence of the growth in software and connectivity in medical technology: a recent study found that in the U.S. alone, over 36,000 health care devices were discoverable on Shodan,

a network that has been likened to "a sort of search engine for connected devices" (Trend Micro, 2016). The growth of connected devices in recent years and concerns about their software security (Newman, 2017 and as noted above) render an understanding of software-driven innovation in this setting all the more important.

This study considers the role of firm capabilities that have been known to impact innovative activity in other sectors and contexts. These factors include firm experience – both inside and outside of a specific R&D area – as well as geography/geographic clusters, and firm financial resources. Firm experience and incumbency have been shown to play a positive role in innovative activity in contexts ranging from biotech (Henderson & Cockburn, 1996) and pharmaceuticals (Scott Morton, 1999; Nerkar & Roberts, 2004) to computer and IT hardware (King & Tucci, 2002; Bayus & Agarwal, 2007). In various settings and competitive environments, other research has shown that a firm's experience in an industry is important for predicting when and how it enters new markets.

Explanations for the enduring role of incumbent firms are numerous, but include organizational experience in specific types of markets (Scott Morton, 1999), productivity spillovers in R&D activities (Henderson & Cockburn, 1996), complementarities among technological and product-market experience (Bayus & Agarwal, 2007), and experience with the process of new market entry *itself* (King & Tucci, 2002). Using detailed commercialization histories, we revisit this set of questions in the medical technology setting.

Other studies have highlighted the role of geography in innovative activity. Forman, Goldfarb, and Greenstein (2016) study the competing effects of colocation and coagglomeration of invention, or, respectively, whether invention agglomerates in the same area as the primary using industry, or if inventions themselves lead to spillovers and agglomeration. They find that the coagglomeration theory is most likely, showing evidence of geographic clustering of patents within the San Francisco Bay Area in information and communication technologies as well as more generally. Earlier research from Jaffe and Trajtenberg (1993) suggests similar dynamics – namely, that local knowledge spillovers lead to geographic clustering of patent citations. In the context of biotechnology, Mariani (2004) highlights the

importance of the role of knowledge spillovers and agglomeration economies in research-intensive sectors.

Finally, firm resources are thought to explain innovation in some settings. Cohen (2010) reviews the literature on this topic and concludes that in many (but not all) settings, a firm's cash flow is associated with higher levels of R&D spending, noting that at least for smaller firms, the causality is thought to run from the former to the latter (Hao and Jaffe, 1993). Thus, we consider firms' recent revenues as well as access to capital (in particular, public markets and/or VC funding) as specific types of resources that may predict digital innovation. The role of each of these potential factors is explored, in turn, in the analysis that follows.

3 Conceptual Framework

Below, we outline a simple conceptual framework for considering how firms make decisions to pursue new product development projects, given heterogeneous past experiences and resources. In particular, we emphasize that the *existence of entry regulation* and the costs and institutional know-how required to overcome the hurdle of regulatory approval generate differences in the relative costs of commercialization activities across different types of firms.

A typical feature of digital products is low (or zero) marginal cost of provision to additional customers (Goldfarb, et al., 2015). We build on this intuition, noting that in the case of a multi-purpose technology such as software – e.g. for digital data transmission, imaging, or data display – the marginal cost of applying the technology to subsequent products within a firm's portfolio will fall as the firm acquires experience.

In this setup, previously acquired capabilities related to the commercialization of digital products lower the subsequent costs of digital product development, thereby shaping differences in the cost of commercializing future products. These differences in costs can be either mitigated or exacerbated by the availability of financial resources, which are known to shape R&D investments at the firm level (CITES CITES CITES, and many others).

3.1 Framework for firm decision-making

A simple framework for considering firm investments can be seen in the following stylized 2-period model: consider a firm, f, facing a decision in period 1 ($t = 1$) regarding commercialization of a product in class, c. Commercializing each product involves costs, C_{fct}, which include manufacturing/production costs, M_{fct} and investment costs, I_{fct} – e.g. for product design and R&D. That is, $C_{fct} = M_{fct} + I_{fct}$.

Commercialization of a product results in expected revenues in period 2, r_{fct+1}. Firms will invest in commercializing new products when $C_{fct} < r_{fct+1}$, that is, whenever net expected profits from a given product are positive.[15]

$$r_{fct} - C_{fct} = \pi_{fct} > 0$$

Investment costs required to commercialize a new product, I_{fct}, in turn, will be shaped by both the firm's financing costs, K_{ft}, and the firm's existing capabilities, E_{fct}, in period 1, such that $I_{fct} = f(K_{ft}, E_{fct})$, where the first derivative of I_{fct} with respect to both K_{ft} and E_{fct} is negative. Therefore, we can re-write total cost as a function of manufacturing costs, financing costs, and capabilities in period 1:

$$C_{fct} = f(M_{fct}, K_{ft}, E_{fct})$$

Assumption 1: We assume that all firms can access the same local labor and raw materials markets so that the remaining variation in the cost of manufacturing (after controlling for location) is only related to differences in financing costs and capabilities, both of which vary across firms and over time. Then we can acknowledge that differences in product commercialization costs, C_{fct}, vary only as a function of financing costs and firm capabilities: $C_{fct} = f^*(K_{ft}, E_{fct})$.

[15]This stylized 2-period model does not incorporate the fact that it may take more than one period for an investment to realize positive profits. This could be included in a more complex model, however we note that since the average product lifecycle is just 1.5-2 years (Wizemann, 2010), it is not unrealistic to assume that products should achieve profitability on a very short time horizon in order to justify commercialization.

In this setup, two further assumptions, consistent with both the theoretical and empirical literature, are required in order to observe variation in firm investment and commercialization decisions over time.

Assumption 2a: We assume that E_{fct} accounts for both local and within-firm capabilities, and that these capabilities can be both general (i.e. any experience with software-driven products) or specific (e.g. experience with software-driven radiology devices). Local capabilities include the local labor market, as above, in addition to the more specific capabilities related to software device commercialization that may exist in a region. The importance of regional expertise and geography in predicting innovative activity has been well studied (Delgado, Porter, & Stern, 2014; Forman, Goldfarb, & Greenstein, 2016; Mariani, 2004; Jaffe, Trajtenberg & Henderson, 1993, and many others) and so has the role of within-firm capabilities (Henderson & Cockburn, 1996; Scott Morton, 1999; King & Tucci, 2002; Bayus & Agarwal, 2007; Nerkar & Roberts, 2004, and many others).

We can write E_{fct} as a function h where $E_{fc} = h(\gamma_{fc}, \alpha_{fc})$ such that $\frac{\delta E}{\delta \gamma}$ and $\frac{\delta E}{\delta \alpha} > 0$. That is, a firm's total capabilities are increasing in both local geographic capabilities, γ and within-firm capabilities, α.

Assumption 2b: We assume that there is a hierarchy in applicability of the above capabilities such that class-specific software experience is the most relevant followed by general software experience. This hierarchy applies to both local (geographic) capabilities and within-firm capabilities. Hence, we can write the equation for E_{fc} with more granularity such that $E_{fc} = h(\gamma_f, \gamma_{fc}, \alpha_f, \alpha_{fc})$, where we consider both general and class-specific components to local and firm capabilities. The hierarchy in applicability results in a hierarchy of importance, where more specific experiences should drive down costs further than general experiences. Thus, $\frac{\delta E_{fc}}{\delta \gamma_f} < \frac{\delta E_{fc}}{\delta \gamma_{fc}}$ and $\frac{\delta E_{fc}}{\delta \alpha_f} < \frac{\delta E_{fc}}{\delta \alpha_{fc}}$.

Assumption 3: We assume that K_{ft} incorporates a number of factors such as the cost of capital the firm faces at the time of its investment decision and other sources of financial resources such as past revenues and recent venture capital investments. This assumption is consistent with literature linking firm performance and innovation to access

to finance and financial constraints (Hao & Jaffe, 1993; Cohen & Klepper, 1996; Cohen, 2010; Stern, 2017).

As such, we can write K_{ft} as a decreasing function of a) being publicly listed (i.e. having access to public capital markets), ϕ_{ft}; b) being VC-funded, v_{ft}, and c) recent revenues, $r_{f(t-T)}$, where T≥1.[16] We can define K_{ft} as a function g where $K_{ft} = g(\phi_{ft}, v_{ft}, r_{f(t-T)})$ and $\frac{\delta K}{\delta \phi}$, $\frac{\delta K}{\delta v}$, and $\frac{\delta K}{\delta r} < 0$.

In plain terms, the costs of commercializing a new product will vary with firm capabilities as well as financial inputs to R&D activities.[17] Then, C_{fct} will be decreasing in E_{fct} and increasing in K_{ft}. As a direct corollary, expected profits in period 2 (of firm f commercializing a given device in period 1), π_{fct+1}, will be increasing in E_{fct} and decreasing in K_{ft}. All cross-partial derivatives of C and π can therefore be signed, leading to a set of testable hypotheses.

The first set of hypotheses deal with the implications of differences in the cost of new product development vis-a-vis relevant firm experience, from broadest to most specific:

- **Hypothesis 1**: $\frac{\delta C_{fct}}{\delta E_{fc}} \frac{\delta E_{fc}}{\delta \gamma fc} < \frac{\delta C_{fct}}{\delta E_{fc}} \frac{\delta E_{fc}}{\delta \gamma f} < 0$. That is, local capabilities decrease commercialization costs and do so in a way that is increasing in the specificity/relevance of those capabilities.

- **Hypothesis 2**: $\frac{\delta C_{fct}}{\delta E_{fc}} \frac{\delta E_{fc}}{\delta \alpha fc} < \frac{\delta C_{fct}}{\delta E_{fc}} \frac{\delta E_{fc}}{\delta \alpha f} < 0$. That is, within-firm capabilities will decrease commercialization costs and do so in a way that is increasing in the specificity/relevance of those capabilities.

The second set of hypotheses address the implications of financial resources for patterns of commercialization, as smaller and more capitally constrained firms will have reduced incentives to pursue new software product commercializations:

[16] More formally, we can imagine a discount factor specifying how the impact of past revenues diminish over time.

[17] Medical device innovation occurs in a setting where there are known delays for the commercialization of novel products, leading to resource-constrained firms being less likely to enter new markets relative to less financially constrained firms (Stern, 2017).

- **Hypothesis 3a**: $\frac{\delta C_{fct}}{\delta K_{fc}} \frac{\delta K_{fc}}{\delta \phi_{ft}} < 0$. The cost of new product development will be lower for publicly listed companies, leading these firms to be more likely to engage in software-driven innovation.

- **Hypothesis 3b**: $\frac{\delta C_{fct}}{\delta K_{fc}} \frac{\delta K_{fc}}{\delta v_{ft}} < 0$. The cost of new product development will be lower for firms with venture capital funding, leading these firms to be more likely to engage in software-driven innovation.

- **Hypothesis 3c**: $\frac{\delta C_{fct}}{\delta K_{fc}} \frac{\delta K_{fc}}{\delta r_{ft}} < 0$. The cost of new product development will be lower when recent revenues have been high(er), leading those firms with higher lagged revenues from all products to be more likely to engage in software-driven innovation.

We test each of these hypotheses in the analyses described below.

4 Data, Classification, and Summary Statistics

4.1 Summary

This project draws on four main sources of data, which are described in detail below. We begin with regulatory data on the universe of moderate-risk and high-risk medical devices cleared by the FDA over 15 recent calendar years (all new products from 2002-2016, inclusive). For each new device, we collect and analyze the text of the accompanying product statement or summary. Using an automated script and two different types of supervised document classification, we identify and characterize digital (i.e. software-driven) devices. First, we document the incidence and frequency of several keywords related to software and networking capabilities in products and track these keywords over time. Subsequently we use the National Library of Medicine's Medical Text Indexer (MTI) – a set of document classification algorithms that take free text and provide subject indexing recommendations based on the Medical Subject Headings ("MeSH®" vocabulary) established by the National Institutes of Health (NIH) – to validate the keyword-driven classification exercise. Using the commercializing firm's identity (included in each device's clearance and approval documents)

and combining this information with historical data about location and firm-level financial data, we characterize commercializing firms at the time each medical device in our dataset came to market.

4.2 Administrative Data on New Medical Devices

The first dataset for this project comes from combined regulatory clearance documents associated with all new moderate-risk and high-risk medical devices that have come to market in the United States since 1996. Moderate risk devices, such as hearing aids, blood pressure monitors and echocardiograph devices, are the largest category of devices cleared for marketing by the FDA. This clearance happens through a process called "510(k)," described briefly above and (in detail) in the appendix. The 510(k) is the final step of the research and development process, after which a cleared product can be legally marketed in the United States. The FDA has historically received approximately 4,000 applications for new 510(k) devices annually, compared to fewer than 100 PMA Applications (Maisel, 2004). PMA track devices are typically high-risk, as a result of being implantable and/or life-sustaining. These include products like pacemakers and stents.

The FDA's 510(k) clearance database[18] and PMA approval database[19] include the full set of device names, product codes (specific 3-letter classifications that categorize devices according to site of use and purpose), and submission and FDA decision dates for all products historically cleared/approved for marketing by the FDA. The top eight medical specialty areas (classes) account for over 75% of all new product approvals and are the focus of this study (Table 1). Each of these classes experienced over 2,000 unique new device approvals between January 1, 2002 and December 31, 2016.[20] Due to availability of rich text data described in the next sub-section, this represents our period of analysis. Over this time, a total of 35,794 new moderate-risk and high-risk devices came to market in the United States.

[18]http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/DeviceApprovalsandClearances/510kClearances/ucm089428.htm

[19]https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpma/pma.cfm

[20]These eight classes are dentified using the full set of FDA clearance records available and therefore represent the universe of newly-approved, FDA-regulated devices.

As noted above, these devices are assigned to medical specialty-related classes, with multiple product codes in each class and (typically) multiple unique devices within each product code. Figure 1 presents a simple example of the hierarchy of the classification system.

4.3 Rich Text Data

The second source of data is a novel database of unstructured text files comprised of the device summaries (product descriptions) that accompany a new product's regulatory clearance by the FDA. At the time of 510(k) clearance or PMA approval, a standardized "summary" or "statement" is published for each device. Among other requirements, the summary must contain:

> *"a description of the device such as might be found in the labeling or promotional material for the device, including an explanation of how the device functions, the scientific concepts that form the basis for the device, and the significant physical and performance characteristics of the device, such as device design, material used, and physical properties; [807.92(a)(4)]."*[21]

Exhibits 1 and 2 present examples of 510(k) and PMA summaries, respectively. In under 10% of cases in our sample years, a premarket "statement" was published in lieu of a summary[22] and when this was the case, we used the text from the product statement instead. While significantly less detailed, product statements can and do contain relevant information about the content of products (e.g. mentions of the use of software) and therefore provide additional information that is relevant for this study. We note, however, that the use of statements may lead to under-counting of software products. Therefore, in robustness tests, we repeat all results to show that they hold when considering the sample of product summary documents only.

[21] https://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/PremarketSubmissions/PremarketNotification510k/ucm142651.htm

[22] https://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/DeviceApprovalsandClearances/510kClearances/ucm089452.htm

Because device summaries and statements are published as online PDFs following a standardized URL-format, we use an automated script to batch download all documents. These documents began to be digitized in May 2001 and we begin our study sample in 2002, the first full calendar year with digitized documents available. Using Abbyy FineReader optical character recognition (OCR) software, we then convert these documents into machine-readable text files. In total, 98% of product summaries/statements could be converted to a machine-readable format for a total of 35,794 device-text pairs.[23] We have no systematic concerns regarding selection or time trends in missing text data: in the years we consider, the machine readability of online PDFs was not statistically different across medical specialties overall, for any given year, or over time. For all years, at least 97% of all digital documents were machine-readable following OCR processing. Appendix Table 1 presents the total number of machine readable summaries included in our sample by calendar year.

The use of text data in empirical analysis is nascent, but includes a growing set of techniques that can be applied to research. Gentzkow, Kelly, and Taddy (2017) describe several methods of parsing and analyzing text data and highlight the fact that "the information encoded in text is a rich complement to the more structured kinds of data traditionally used in research." Indeed, text has previously been used in studies ranging from sentiment analysis of policy uncertainty (e.g. Baker, Bloom, and Davis, 2016) to labor economics (e.g. Deming and Kahn, 2017) and in the analysis of patent data (e.g. Moser, Ohmstedt, and Rhode, 2017). Here, we demonstrate its utility in understanding the content and functionality of new medical technologies.

We process the text data in two ways, each of which leads to a similar classification of digital medical devices. Our first approach is a form of supervised document classification in which we identify the incidence and frequency of keywords related to software and networking capabilities in each device description. These terms were selected in advance using two online glossaries of computer related terms[24] (a list of the 36 most frequently used keywords – each of which were found in over 100 unique product descriptions – can be found in Appendix

[23]These include 35,495 510(k) summaries and 299 PMA summaries.

[24]composite list from http://www.math.utah.edu/∼wisnia/glossary.html and https://pc.net/glossary

Table 2). Unsurprisingly, "software" and several related keywords have increased in their frequency of use over time. Because "software" is the most common among our search terms and is highly correlated with others, we rely on summaries containing the keyword "software" as our primary indicator to identify digital products.

These data represent our first application of simple document classification to characterize devices. This classification method has the advantage of being simple and highly transparent, but the disadvantage of being somewhat *ad hoc*; however, we believe this method will have a high rate of success in identifying products of interest. In particular, since the product descriptions included in FDA clearance documents are standardized and parsimonious, there is no reason, and indeed no option, to include extraneous words related to technologies that are not involved in the device itself. To put it simply, keywords such as "software" or "computer" will not appear in the product description if they do not relate to meaningful aspects of the device's functionality (See Exhibits 1 and 2). Nevertheless, we performed several manual inspections to confirm that incidents of keywords found were indeed references to the technology in the device. For example, we drew a random sample of 120 devices (8 per calendar year) that had been flagged for including general "software" capabilities and manually inspected each of these devices' summaries. In this sample, 100% of devices flagged as including "software" were found to be correctly coded (i.e. a 0% rate of type I error in this random sub-sample).

We validate our *ad hoc* supervised document classification using the National Library of Medicine's MTI algorithm. As noted above, the MTI takes free text as an input to provide subject indexing recommendations based on the MeSH vocabulary established by the NIH. Since our primary measure of digitization is the incorporation of software into new products, we classify devices using the MTI and generate an indicator for whether the algorithm assigned the MeSH code for *software* to the product. This MeSH code broadly covers "sequential operating programs and data which instruct the functioning of a digital computer."[25]

[25]In the MeSH Tree, "Software" takes the Tree Number L01.224.900. We identify all products that are classified as being anywhere on the "Software" branch of the MeSH Tree.

The MTI algorithm has the advantage of being externally validated by the NIH and through years of experience in use by the National Library of Medicine, but two clear disadvantages. First, we believe that the bar will be higher for flagging products as containing software, since the MTI will require a *discussion* of software programs in the text, beyond simply invoking the keyword "software." For this reason, our expectation is that the MTI is likely to identify the more software intensive subset of products in our sample. Second, the MTI is non-transparent in how it assigns concepts to text.[26]

Comparing the MTI output to our own keyword-based document classification method, we find high degree of overlap: 100% of the devices flagged by the MTI as describing software are also identified by the keyword method as being about the subject of software. However, as expected, not all summaries using the keyword "software" are identified by the MTI. The rightmost column of Appendix Table 1 presents a cross-tabulation of our *ad hoc* keyword-based document classification vs. the MTI's classification. Notably, the actual keyword "software" has the highest degree of overlap with the MTI-based definition. Because we care primarily about digitization in the case of the incorporation of *any* software, we focus on the keyword-based definition, however in all regression analyses, we test the alternative (MTI-based) definition. These results are presented in the appendices and the choice of definition does not appear to change the sign or statistical significance of the main results below, however magnitudes are attenuated roughly proportionally to the decrease in the number of software devices that are included in the MTI-defined sample.

Figure 2 presents the growth of new digital devices over our period of observation at the firm-product[27]-year (F-P-Y) level. Figure 3a shows the growth in digitized product codes – i.e. unique *types* of devices – over time, while Figure 3b shows growth in the number of *firms* pursuing digital innovation. Through these figures, we see that the growth in digital devices over our period of observation was a result of the entry of both new products and new firms. Figure 4a shows that the number of digital product codes grew by over 400% over

[26]The MTI algorithm itself is not observable/open source; rather, we batch process the text of product descriptions through the algorithm and record the subject headings that it returns as output.

[27]FDA product code

our period of observation, while non-software product codes grew by only about 150% (albeit off of a higher base). Figure 4b breaks down the growth of digital devices across medical specialty classes, revealing interesting heterogeneities. Although all classes show growth, the early incidence of digitization varies dramatically.

4.4 Firm Financial Data

Next, each device is linked by its commercializing entity to detailed firm financial data. In this process, we first link commercializing entities to a panel of firm acquisitions created using data from *EvaluateMedTech*[28] in order to properly handle subsidiary ownership and introduce the notion of *child* (acquired) and *parent* (acquirer) firms, important distinctions that we use to obtain a better understanding of the effects of capital constraints on firm innovation. These child and parent firms are then separately linked to revenue and venture capital data.[29] Firm names were also cleaned using a consistent set of rules to account for suffixes and abbreviations.[30]

Data on venture capital funding is obtained from a combination of *EvaluateMedTech* and *Preqin*, with precedence given to the latter.[31] We are able to observe deal dates and funding amounts for each linked firm, which we use in creating both lagged binary indicators for whether a firm was ever venture-funded or venture-funded prior to product commercialization, as well as running totals for dollar values of venture funding.

Data on firm financial performance is obtained from a combination of *EvaluateMedTech* and *Capital IQ*, with precedence given to the *former*, as it has broad coverage of the medical device industry.[32] These data include revenues, for which we create lagged analogues, and

[28] A market intelligence database that tracks public and private firms in the medical device industry

[29] All linking in this section is done using the software program *matchIT*, which performs fuzzy matching of company names (or addresses) between (or within) datasets and grades the text match quality by score. We used this software because it is highly flexible, fully parameterized, and deals effectively with foreign names.

[30] This method is similar in nature to work done for the NBER Patent Data Project by Bessen

[31] Preqin is the widely considered the best publicly available dataset for venture funding and is used in a variety of contemporary literature, as in Korteweg and Nagel (2016) and Harris, Jenkinson, and Kaplan (2013).

[32] We validate EvaluateMedTech data using *Capital IQ*, long considered a primary source for detailed firm financials. See, for example, Sheen (2014); Acharya and Xu (2016); Booth and Salehizadeh (2011).

a binary indicator for whether a firm is public at the time of product commercialization, as well as ever. Although our sample includes both public and private firms, in many countries – and in particular in the United States – private firms are not required by law to disclose detailed financial information. Nonetheless, some data (e.g. estimated revenues) are available in industry-specific datasets like *EvaluateMedTech*.

Figures 5 and 6 show the share of digital devices that were commercialized by venture capital-funded firms (among all non-public firms) and public firms, respectively.

4.5 Classifying Innovation and Accounting for Firm Geography and Experience

In the final steps of data assembly, we collapse our data set to the F-P-Y level. This only impacts firms that commercialized multiple devices within a specific product code in a given calendar year (e.g. a firm that brought two carotid artery stents to market in 2010 would only be counted once in the product code for carotid artery stents in that year. This coding is consistent with observing a firm doing a certain *type* of product innovation in a given year). In the case of digital innovation, we take the maximum of our indicator for digtial innovation at the F-P-Y level. For example, if a firm brought two blood pressure monitors to market in 2010 and only one of those devices contained software, the firm would be characterized as having done digtial innovation in blood pressure monitors in the year 2010. Collapsing the data to F-P-Y level creates a standard unit of innovation by which we can compare firms over time. This yields a data set of 27,310 observations of firm-product-year level innovative activity over our study period. [33]

With respect to geography, we characterize firms as to whether or not they are in a

[33]In order to collapse the full set of new product data properly, we cleaned commercializing firm names using the same set of rules as used in the firm financial data construction. In addition, we hand-checked all firm names to capture instances in which similar, non-subsidiary firms were recorded differently due to either misspellings or omission/addition of words. For instance, BD was hand replaced with Beckton Dickinson and *Medartis Medizinprodukte und Forschung* ("Medartis Medical Products and Research") was replaced simply with *Medartis* Although time-consuming, this allows us to properly aggregate firm experience with the regulatory approval process and within specific product types and medical device classes (as described below).

"software cluster" in three (increasingly specific) ways. First, we consider local labor market expertise. Using annual data from the U.S. Bureau of Labor Statistics (BLS), we compile data on each state's share of software engineers in the labor force in order to consider whether there is a relationship between the characteristics of the skilled IT workforce in a given state and the likelihood of digital innovation emerging from that state. Figure 7 presents a set of example states. While there is some variation over time within states, the primary source of variation in this measure is across states.

Next we consider two definitions of state clusters as defined by where device commercialization has taken place in recent years. We define state software clusters as the top 10 states for software device commercialization, based on a 5-year moving average. We then consider a class-specific version of this definition, in which we define the top 10 states for software device commercialization within each medical specialty class and define an indicator variable for whether or not a device was originating from a cluster, based on each of the definitions. Both sets of definitions are based on a 5-year moving average of products brought to market preceding the year of observation. As a result of this step, our analysis sample is limited to the years 2006-2016 (inclusive), to facilitate a 5-year look-back on regional product expertise. Figures 8a and 8b show the share of digital devices in top-10 states vs. non-top-10 states using these two definitions.

Finally, we characterize firms' digital device experience on two dimensions. First we calculate the count of total digital devices the firm has commercialized up to, but not including, the year observation. Second, we calculate the count of total digital devices within each *class* that the firm has commercialized up to, but not including, the year of observation (Table 3). Table 2 presents summary statistics of all firm financial variables, while Table 3 presents summary statistics of all firm experience variables.

5 Estimation and Results

In the estimation exercises that follow, we are interested in testing the hypotheses outlined above. First, we explore evidence for Hypotheses 1 and 2 by modeling the relationship

between capabilities (external and internal, respectively) and the probability of digital innovation. Next, we explore evidence for Hypotheses 3a-3c by modeling the relationship between firm financial resources and digital innovation.

5.1 Overall estimates

Recall the trends in the digital innovation in medical technology and observed variation across medical specialty areas seen in Figures 2-4. Notably, there is significant heterogeneity across medical specialty areas by volume (Figure 2) and share (Figure 4b) of of products digitized. These descriptive findings suggest the importance of using class fixed effects in empirical models.

At the F-P-Y level, we therefore model the likelihood of digital innovation, D, as:

$$D_{fpct} = f(\beta \mathbf{X})$$

Where **X**s include:

- Indicators of firm capabilities that are related to geographic expertise as well as within-firm commercialization experience. Both geographic and within-firm capabilities can be defined with varying levels of specificity ranging from general past experience to class-specific past experience with product digitization.

- Indicators of firm resources (including whether a product emerged from a publicly listed firm, a VC-funded firm, and measures of recent revenues)

- Controls for clearance year (in order to capture time trends in software inclusion over time) and medical specialty class (in order to account for persistent differences in the relative ease or applicability of software in a given area of medicine and medical technology).

In the regression models that follow, all specifications include year and class fixed effects, with standard errors clustered at the product code level in acknowledgement of

potential differences across product types (in e.g. innovation behavior or regulatory burden). All tables report marginal effects from logit models, facilitating clearer interpretation of statistical relationships. A full set of corresponding linear probability models are included in Appendix X and lead to the same conclusions as those presented below.

Table 4 presents a full set of controls. As expected, there are statistically significant differences across classes and over time. Column 1 shows the full analysis sample, while Column 2 shows the sub-sample for which firm revenue data is available. Notably, the coefficients on sample controls are very similar across the two samples. Columns 3 and 4 present coefficients from the same samples, but use a time trend rather than year fixed effects in regression models. The pseudo-r-squared values are trivially higher in the models using year fixed effects rather than a time trend, so we include both class and year fixed effects in all subsequent regressions (however results are stable regardless of the convention chosen).

5.2 Geographic and within-firm capabilities

Table 5 presents results predicting digital innovation at the F-P-Y level, specifically evaluating Hypotheses 1 and 2. Controlling for medical specialty class and year, we find that although the state's share of software engineers is not a strong predictor of software inclusion in new devices, other measures of geographic expertise are quite strong predictors of digitization. Dummy variable indicators for both being in either a digital innovation cluster (Column 2) or being in a class-specific digital innovation cluster (Column 3) are both strongly associated with higher probabilities of software inclusion in new products and individually increase the probability of digital innovation when included together in predictive models (Column 4). These results suggest that both general and specific geographic expertise are important predictors of digital products, providing support for Hypothesis 1.

Columns 5-7 of Table 5 consider the role of accumulated firm experience in commercializing digital products. Column 5 shows a strong statistical relationship between past digital device experience and current likelihood of digital innovation, while Column 6 shows that this relationship is even stronger when considering class-specific commercializa-

tion experience. Considering both experience variables together reveals that the effect is entirely driven by class-specific experience (indeed, general experience outside of a class is negatively predictive of product digitization). These results suggest that with respect to within-firm capabilities, only accumulated experience that is relevant to a specific product area is positively predictive of digital innovation, providing partial (class-specific), but not overall (general experience driven) support for Hypothesis 2.

Column 8 of Table 5 includes all measures of geographic and within-firm expertise in one regression model and presents coefficients that are very similar to those in Columns 4 and 7, supporting the conclusions drawn thus far.

5.3 Firm financial resources

Table 6 presents results from regressions designed to test Hypotheses 3a-3c. We first consider whether public firms (Column 1) and VC-funded private firms (Column 2) are more likely to engage in digital innovation and also consider these indicators jointly (Column 3). We find that while both coefficients have positive coefficients, only VC funding is a statistically significant predictor of digital innovation (here a binary indicator, but in alternative specifications, the total *amount* of VC funding is also a statistically significant predictor). Column 4 presents the same regression as Column 3, but focuses on the limited sample of firms for which revenue data are available, resulting in similar results in the restricted sample. Columns 5 and 6 add the natural log of firm revenue to regression models, with Column 6 representing the most complete model. The natural log of lagged revenue is a statistically significant predictor of digital innovation in Column 6, but not in Column 5 and is small in magnitude, relative to the coefficient on VC funding.

With respect to Hypotheses 3a-3c, all coefficients in Table 6 are of the predicted sign, but only Hypothesis 3b is broadly supported by the data, with limited support for Hypothesis 3c. In the next section, we consider the relationships between the various factors that are predictive of digital innovation in order to understand how they interact in practice.

5.4 Further regression analysis

Table 7 presents a set of combined models in which Hypotheses 1-3 are evaluated simultaneously. Differences between these results and those seen in Tables 5 and 6 will therefore indicate the size and direction of any omitted variable bias that is unintentionally included in assessing hypotheses individually. The primary difference between the "all-in" models and the regressions to test individual hypotheses is that in the restricted sample specifications, the dummy variable for publicly-listed firms becomes a positive, statistically significant predictor of digital innovation. However, the coefficient switches sign when the same results are run on the full sample (rather than the restricted sample of firms for which revenue data are available). This is unsurprising, since the restricted sample selects strongly for inclusion of publicly listed firms, but excludes many private firms. While these results introduce some noise into interpretation, they are generally supportive of the conclusions drawn thus far.

Table 8 considers the primary factors that predict digital innovation and how they interact in our sample. In particular, we interact our measures of capabilities (geographic and within-firm) with an indicator for receiving VC funding. All results point to *selection* in the flow of VC funds, rather than any separate, additive relationship between VC funding and digitization. Specifically, we find that VC selects on exactly the same set of capabilities that predict digital in other models. Even with interaction terms, these capabilities – general and class-specific geographic experience as well as class-specific, within-firm commercialization experience – continue to be positive, statistically significant predictors of digital innovation alone, however VC does not. These results simultaneously bolster support for the importance of these capabilities in driving digital innovation, while diminishing the evidence that financial resources such as VC funding can separately support digital innovation in their absence.

5.5 Future directions

Future versions of this paper will explore differences between novel vs. follow-on innovation in digital medical technologies as well as more vs. less complex software innovations.

6 Discussion and Conclusions

We document several trends in the digitization of medical technology and its implications for firms in the industry. We characterize the growth of digital in medical devices over time and across medical specialty classes. This descriptive data is novel in that, to our knowledge, no other study has yet quantified the growth of software in medical technology. Along the way, we develop and validate a method for using supervised document classification to analyze the contents of new product descriptions. We use multiple methods to collect indicators of the use of software and other digital capabilities embedded in new medical devices.

We then turn to understanding the types of firms that are most likely to lead digital innovation in medical technology and find a number of pieces of evidence that suggest significant incumbent advantages. Specifically, we observe a strong relationship between location in a geographic cluster and prior commercialization of digital devices in predicting digital innovation. These relationships exist across various levels of specialization in geographic expertise, but are driven by class-specific within-firm experiences.

We also consider a number of financial resources that may predict digital innovation, but find little evidence to suggest that such resources are a driving force in such innovative activity. While venture capital (VC) funding appears to play a role in supporting innovative entrants, a closer look at the data suggests that the relationship is entirely driven by selection on other variables that predict digitization. Taken together, our results include several pieces of evidence in support of within-region and within-firm positive spillovers from past digital innovation into future digital innovation. We conclude that this regulated setting favors firms coming from a position of incumbent geographic and/or experiential advantage.

While future versions of this paper will include additional empirical exercises, our results suggest that stimulating digital innovation in health care will necessitate a nuanced understanding of when and how different types of firms enter new markets. Further, to the extent that new regulatory policies emerge to either streamline or complicate the use of software in new medical technologies, these policies are likely to affect small vs. large firms and incumbents vs. (potential) new entrants differently.

References

Acemoglu, Daron, and Joshua Linn. "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry."Quarterly Journal of Economics3 (2004): 1049-1090.

Acharya, Viral, and Zhaoxia Xu. "Financial Dependence and Innovation: The Case of Public versus Private Firms." Journal of Financial Economics (2016).

Adler-Milstein, Julia, et al. "More than half of US hospitals have at least a basic EHR, but stage 2 criteria remain challenging for most."Health Affairs(2014): 10-1377.

Adler-Milstein, Julia, Joseph Kvedar, and David W. Bates. "Telehealth among US hospitals: several factors, including state reimbursement and licensure policies, influence adoption."Health Affairs33.2 (2014): 207-215.

Agha, Leila. "The effects of health information technology on the costs and quality of medical care."Journal of health economics34 (2014): 19-30.

Arora, Ashish, Lee G. Branstetter, and Matej Drev. "Going Soft: How the Rise of Software-Based Innovation Led to the Decline of Japan's IT Industry and the Resurgence of Silicon Valley."Review of Economics and Statistics95, no. 3 (2013): 757-775.

Averch, Harvey, and Leland L. Johnson. "Behavior of the firm under regulatory constraint."The American Economic Review52.5 (1962): 1052-1069.

Babiarz, Josephine C and Douglas J. Pisano. "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.

Baker, S.R., Bloom, N. and Davis, S.J., 2016. "Measuring economic policy uncertainty." *The Quarterly Journal of Economics*, 131(4), pp.1593-1636.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. "Americans do IT better: US multinationals and the productivity miracle." The American Economic Review 102.1 (2012): 167-201.

Booth, Bruce L., and Bijan Salehizadeh. "In Defense of Life Sciences Venture Investing." Nature Biotechnology 29.7 (2011): 579-83.

Branstetter, Lee G., Matej Drev, and Namho Kwon."Get With the Program: Software-Driven Innovation in Traditional Manufacturing." No. w21752. National Bureau of Eco-

nomic Research, 2015.

Bresnahan, Timothy F., and Peter C. Reiss. "Entry and competition in concentrated markets."Journal of Political Economy(1991): 977-1009.

Brynjolfsson, Erik, and Kristina McElheran. "Digitization and Innovation The Rapid Adoption of Data-Driven Decision-Making."The American Economic Review106, no. 5 (2016): 133-139.

Centers for Medicare and Medicaid Services (CMS). *National Health Expenditure Accounts Highlights*. December 6, 2016.

Comstock, Jonah. "What the FDA did in digital health in 2016." MobiHealthNews, December 30, 2016.

Danzon, Patricia M., and Li-Wei Chao. "Does Regulation Drive Out Competition in Pharmaceutical Markets?." *The Journal of Law and Economics*43, no. 2 (2000): 311-358.

Delgado, Mercedes, Michael E. Porter, and Scott Stern. "Clusters, convergence, and economic performance." Research Policy 43, no. 10 (2014): 1785-1799.

Dorsey, E. Ray, and Eric J. Topol. "State of Telehealth." *New England Journal of Medicine*375, no. 2 (2016): 154-161.

Dranove, David, et al. "The trillion dollar conundrum: Complementarities and health information technology." *American Economic Journal: Economic Policy* 6.4 (2014): 239-270.

FDA, Center for Devices and Radiological Health, Guidance for Industry, FDA Reviewers and Compliance on Off-The-Shelf Software Use in Medical Devices. September 9, 1999 (supersedes document, Guidance on Off-the-Shelf Software Use in Medical Devices, June 4, 1997).

FDA, Center for Devices and Radiological Health, Guidance for Industry Cybersecurity for Networked Medical Devices Containing Off-the-Shelf (OTS) Software. January 15, 2005(A).

FDA, Center for Devices and Radiological Health, Guidance for Industry and FDA Staff on the Content of Premarket Submissions for Software Contained in Medical Devices. May 11, 2005(B).

FDA, Center for Devices and Radiological Health, Guidance for and FDA Staff on the Content of Premarket Submissions for Management of Cybersecurity in Medical Devices. October 2, 2014.

FDA, Center for Devices and Radiological Health, Guidance for Industry and Food and Drug Administration Staff on Postmarket Management of Cybersecurity in Medical Devices. December 28, 2016.

Forman, Chris, Avi Goldfarb, and Shane Greenstein. "Agglomeration of Invention in the Bay Area: Not Just ICT." *The American Economic Review* 106, no. 5 (2016): 146-151.

Deming, David, Lisa B. Kahn. "Firm heterogeneity in skill demands." No. 23328. National Bureau of Economic Research, 2017.

Gentzkow, Matthew, Bryan T. Kelly, and Matt Taddy. "Text as data." No. w23276. National Bureau of Economic Research, 2017.

Goldfarb, Avi, Shane M. Greenstein, and Catherine E. Tucker. "Introduction to 'Economic Analysis of the Digital Economy'." Economic Analysis of the Digital Economy. University of Chicago Press, 2015. 1-17.

Goldfarb, Avi, and Catherine Tucker. "Digital Economics." No. w23684. National Bureau of Economic Research, 2017.

Huckman, Robert, and Maya Uppaluru, "The Untapped Potential of Health Care APIs." *Harvard Business Review*, December 23, 2015.

Iansiti, Marco, and Karim R. Lakhani. "Digital Ubiquity: How Connections, Sensors, and Data Are Revolutionizing Business (Digest Summary)." *Harvard Business Review* 92.11 (2014): 91-99.

Kuehn, Bridget M. "Wearable Biosensors Studied for Clinical Monitoring and Treatment." *JAMA* 316, no. 3 (2016): 255-257.

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*366.9 (2012): 848-855.

Lee, Jinhyung, Yong-Fang Kuo, and James S. Goodwin. "The effect of electronic medical record adoption on outcomes in US hospitals." *BMC health services research*13.1 (2013): 1.

Maisel, William H. "Medical device regulation: an introduction for the practicing physician." Ann Intern Med 2004;140:296-302.

Marr, Bernard. "First FDA Approval for Clinical Cloud-Based Deep Learning in Health-

care." *Forbes*, January 20, 2017.

Miller, Amalia R., and Catherine Tucker. "Privacy protection and technology diffusion: The case of electronic medical records.//*Management Science* 55.7 (2009): 1077-1093.

Miller, Amalia R., and Catherine Tucker. "Privacy Protection, Personalized Medicine and Genetic Testing." Working Paper (2016)

Moser, P., Ohmstedt, J. and Rhode, P.W., 2017. "Patent Citations–An Analysis of Quality Differences and Citing Practices in Hybrid Corn." *Management Science*.

Newman, Lily Hay. "Medical Devices are the Next Security Nightmare." *WIRED Magazine*, March 2, 2017.

Sheen, Albert. "The Real Product Market Impact of Mergers." *The Journal of Finance* 69.6 (2014): 2651-688.

Stern, Ariel Dora. "Innovation under regulatory uncertainty: Evidence from medical technology." *Journal of Public Economics* 145 (2017): 181-200.

Tecco, Halle and Megan Zweig. "Digital Health Funding 2017 Midyear Review: A record breaking first half." RockHealth, 2017"

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

Zharkov, Boris. "The Doctor on a Quest to Save our Medical Devices from Hackers." *WIRED Magazine*, November 24, 2015.

Figures

Exhibit 1: Extract from 510(k) Statement

510(K) SUMMARY JAN 1 1 2008

This summary of 510(k) safety and effectiveness information is being submitted in accordance with the requirements of SMDA 1990 and 21 CFR §807.92.

The assigned 510(k) number is: K073198

1. **Submitter's Identification:**

   Microlife Intellectual Property GmbH, Switzerland

   Espenstrasse 139
   9443 Widnau / Switzerland

   Date Summary Prepared: October 30, 2007

2. **Name of the Device:**

   Microlife Upper Arm Automatic Digital Blood Pressure Monitor, Model WatchBP Home (BP3MX1-1).

3. **Information for the 510(k) Cleared Device (Predicate Device):**

   Microlife Upper Arm Automatic Digital Blood Pressure Monitor, Model BP3AC1-1 PC, K#060686.

4. **Device Description:**

   Microlife Upper Arm Automatic Blood Pressure Monitor, Model WatchBP Home is designed to measure the systolic and diastolic blood pressure and pulse rate of an individual by using a non-invasive technique in which an inflatable cuff is wrapped around the Upper arm. Our method to define systolic and diastolic pressure is similar to the auscultatory method but uses an electronic capacitive pressure sensor rather than a stethoscope and mercury manometer. The sensor converts tiny alterations in cuff pressure to electrical signals, by analyzing those signals to define the systolic and diastolic blood pressure and calculating pulse rate, which is a well - known technique in the market called the "oscillometric method".
   The device has <DIAG> and <USUAL> measurement mode. In additional, the device can be used in connection with your personal computer **(PC)** running the WatchBP 1.0 software. The memory data can be transferred to the PC by connecting the monitor via cable with the PC.

Exhibit 2: Extract from PMA Statement

SUMMARY OF SAFETY AND EFFECTIVENESS DATA (SSED)

I. GENERAL INFORMATION

Device Generic Name:	Continuous Glucose Monitoring (CGM) System
Device Trade Name:	iPro2 Continuous Glucose Monitoring (CGM) System
Device Procode:	MDS
Applicant's Name and Address:	Medtronic MiniMed 18000 Devonshire Street Northridge, CA 91325
Date(s) of Panel Recommendation:	None
Premarket Approval Application (PMA) Number:	P150029
Date of FDA Notice of Approval:	June 17, 2016
Priority Review:	*Not Applicable*

II. INDICATIONS FOR USE

iPro2 CGM System (MMT-7745)
The iPro2 Recorder is to be used with either Enlite sensor or Sof-Sensor and is intended to continuously record interstitial glucose levels in persons with diabetes mellitus. This information is intended to supplement, not replace, blood glucose information obtained using a standard home glucose-monitoring device. The information collected by the iPro2 Recorder may be uploaded to a computer (with Internet access) and reviewed by healthcare professionals. This information may allow identification of patterns of glucose level excursions above or below the desired range, facilitating therapy adjustments which may minimize these excursions.

VI. Software

The current software version for the iPro2 CGM system is v1.1A. Software verification and validation were carried out in accordance with the FDA guidance document *Guidance for the Content of Premarket Submissions for Software Contained in Medical Devices: General Principles of Software Validation: Final Guidance for Industry and FDA Staff (2002).* Software development activities included establishing detailed software requirements, linking requirements with associate verification tests, software code reviews, unit testing, system level testing and defect tracking and dispositioning to ensure the software conforms to patient needs and intended uses. Software was previously reviewed under P980022/S071.

VII. Human Factors Testing

The sponsor referenced human factors testing from previous submissions (P980022 and P120010) and provided new testing to support the proposed system configuration. New testing included the following:

- Evaluation of tasks regarding the removal of the iPro2 recorder from the Enlite sensor and inspection of fluids on the recorder before initiating contact with the iPro2 docking station.
- Evaluation of specific tasks performed in the software.

Figure 1: Device classification (example)

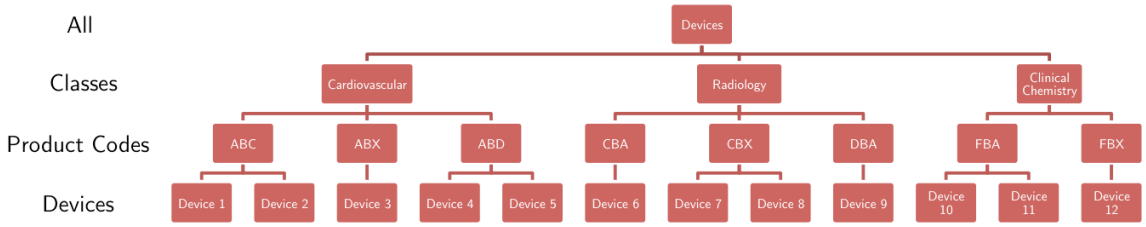


Figure 2: Number of newly approved devices with software (F-P-Y level)

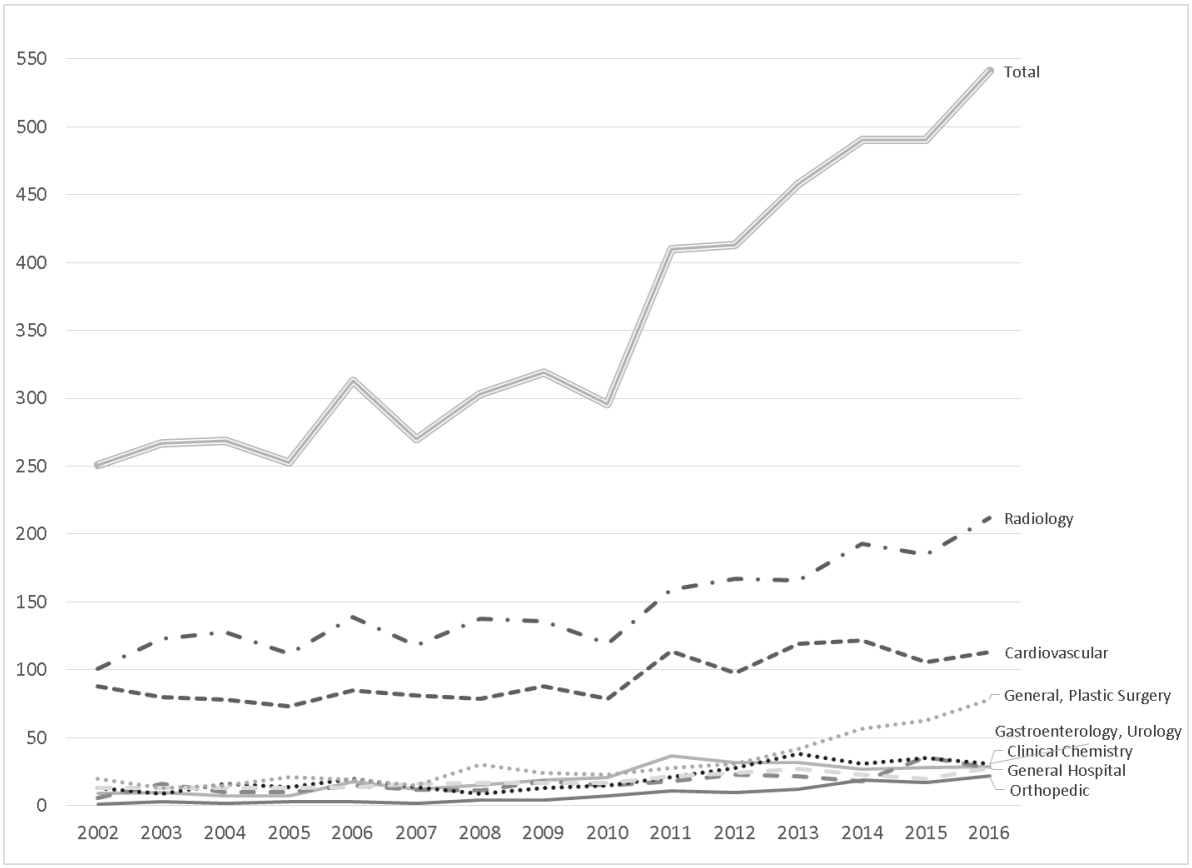


Figure 3a: Cumulative number of device product codes with software

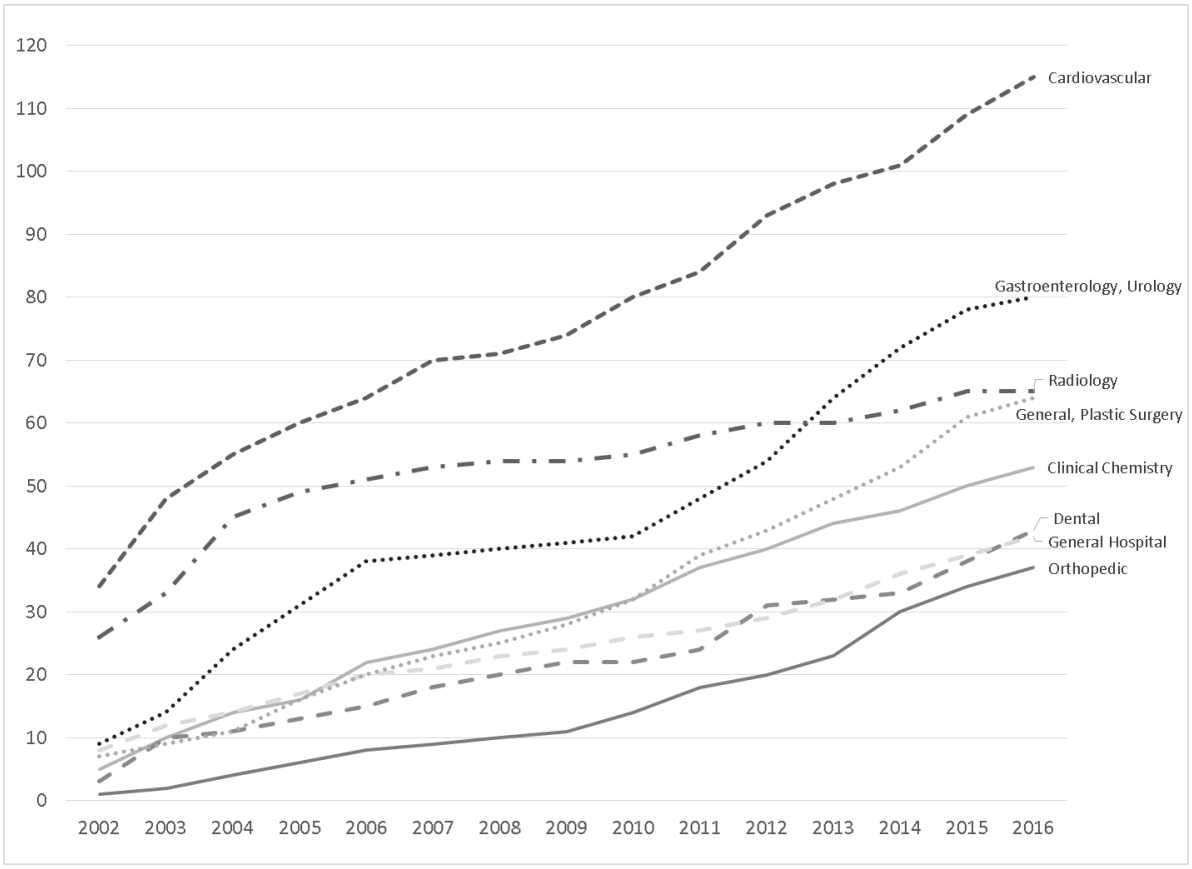


Figure 3b: Cumulative number of firms with software devices (F-P-Y level)

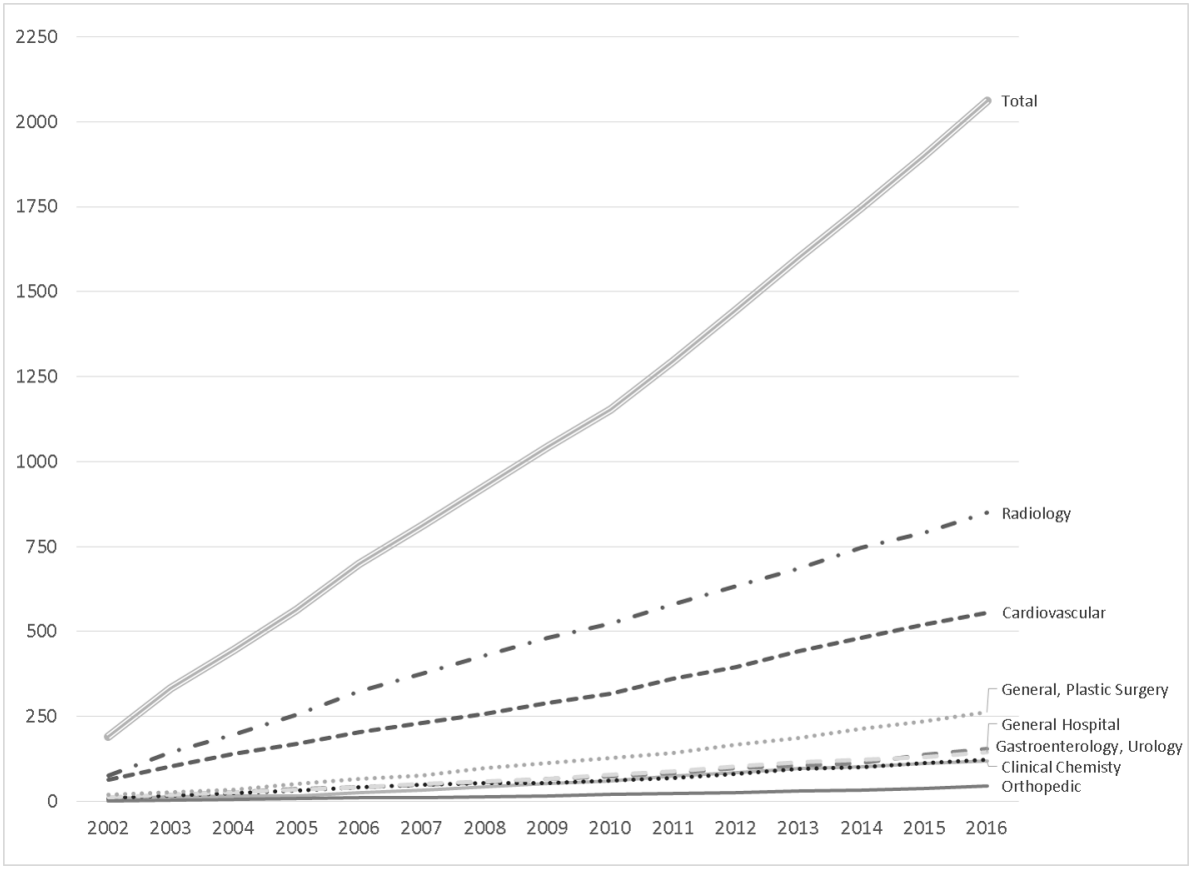


Figure 4a: Cumulative growth of product codes with software (base year = 2002)

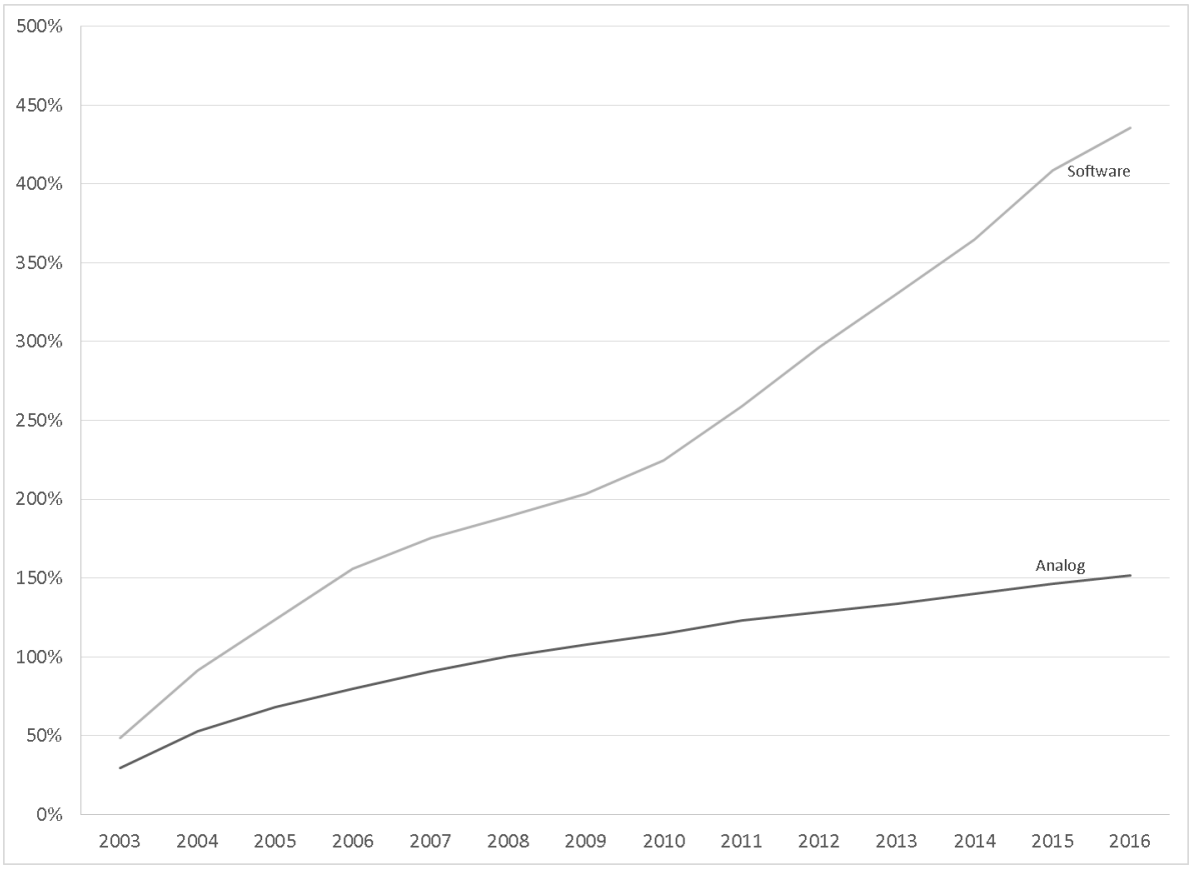


Figure 4b: Share of newly-approved devices with software (F-P-Y level)

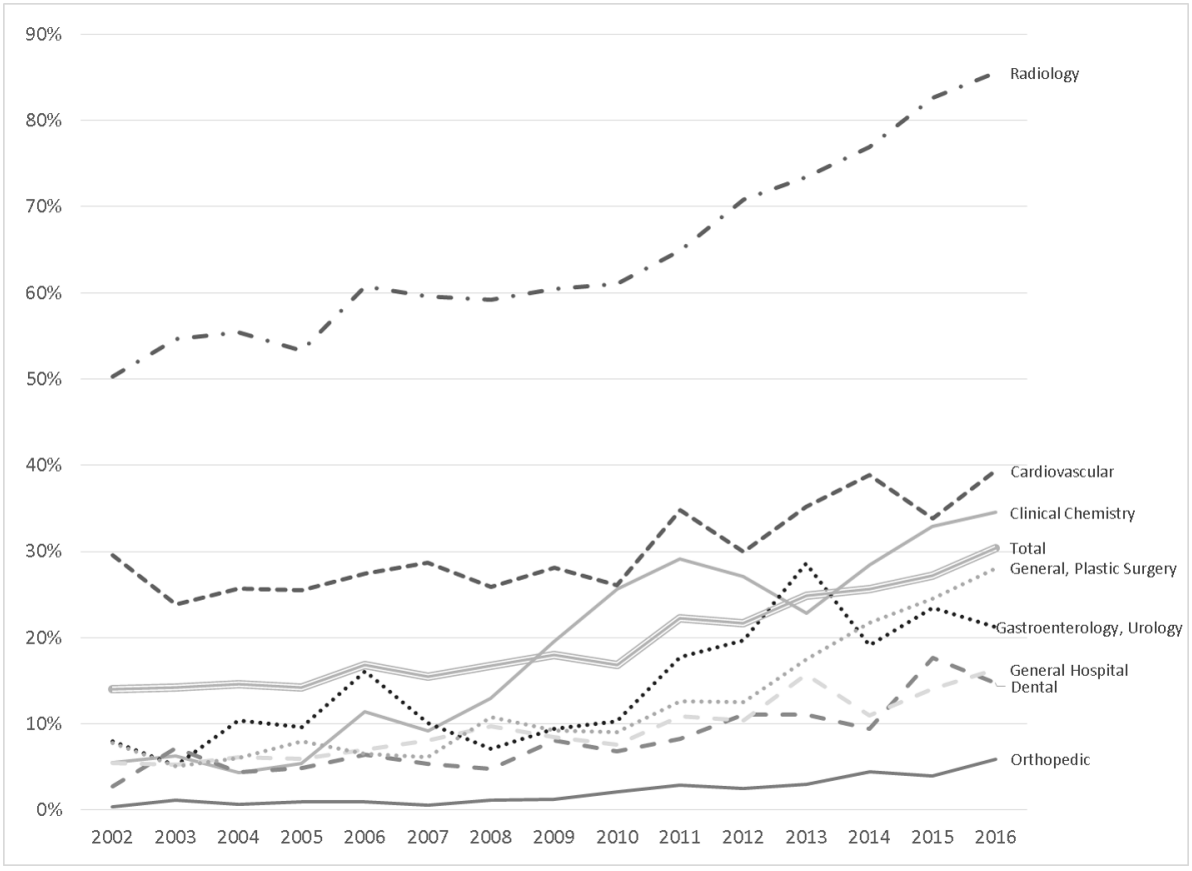


Figure 5: Share of devices with software: VC vs. non-VC-funded private firms

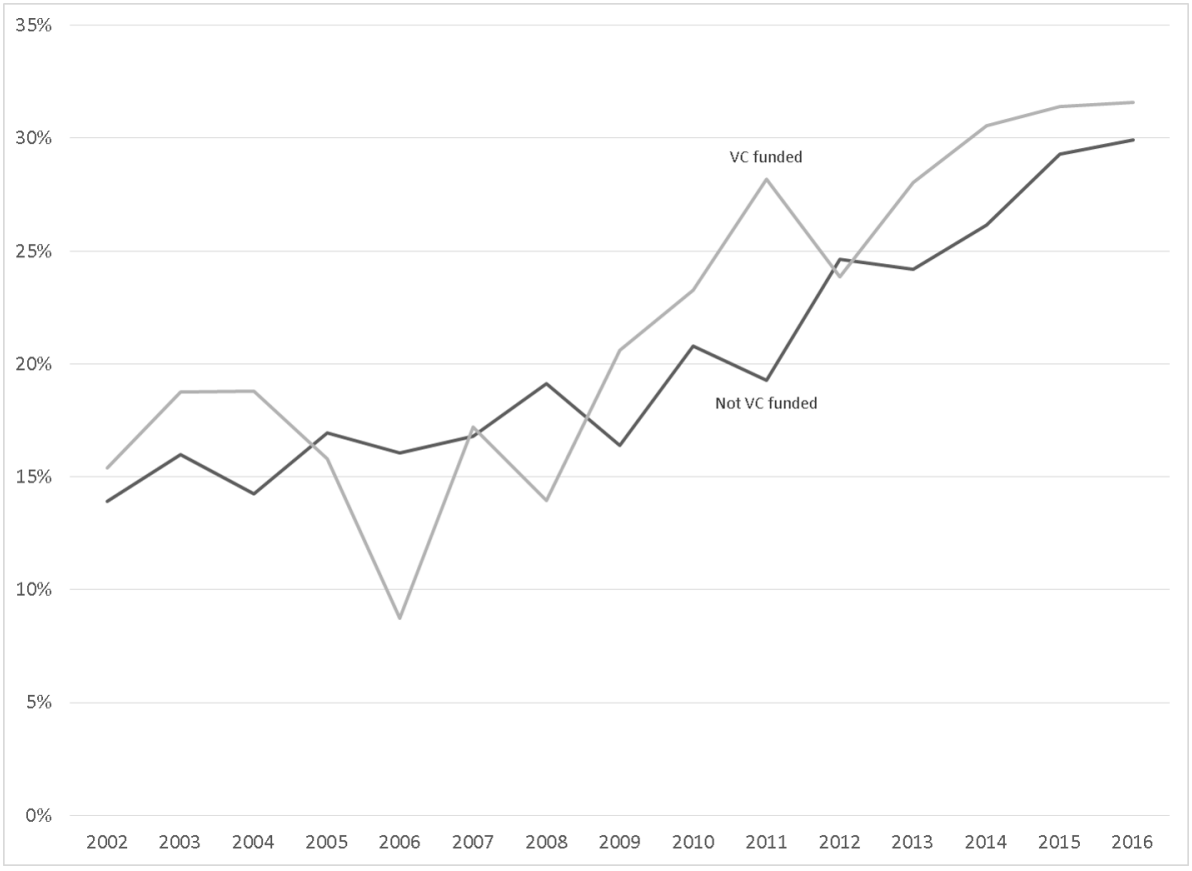


Figure 6: Share of devices with software: publicly-listed vs. private firms

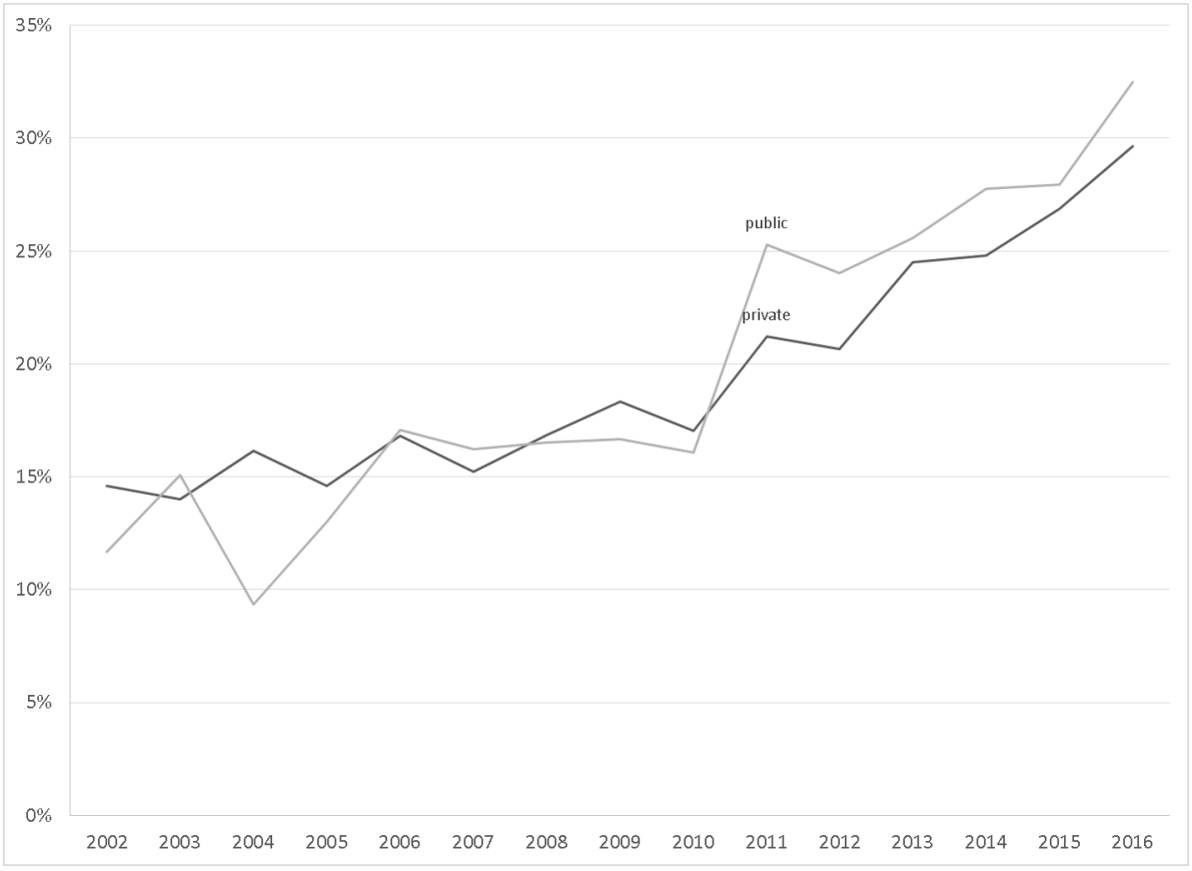
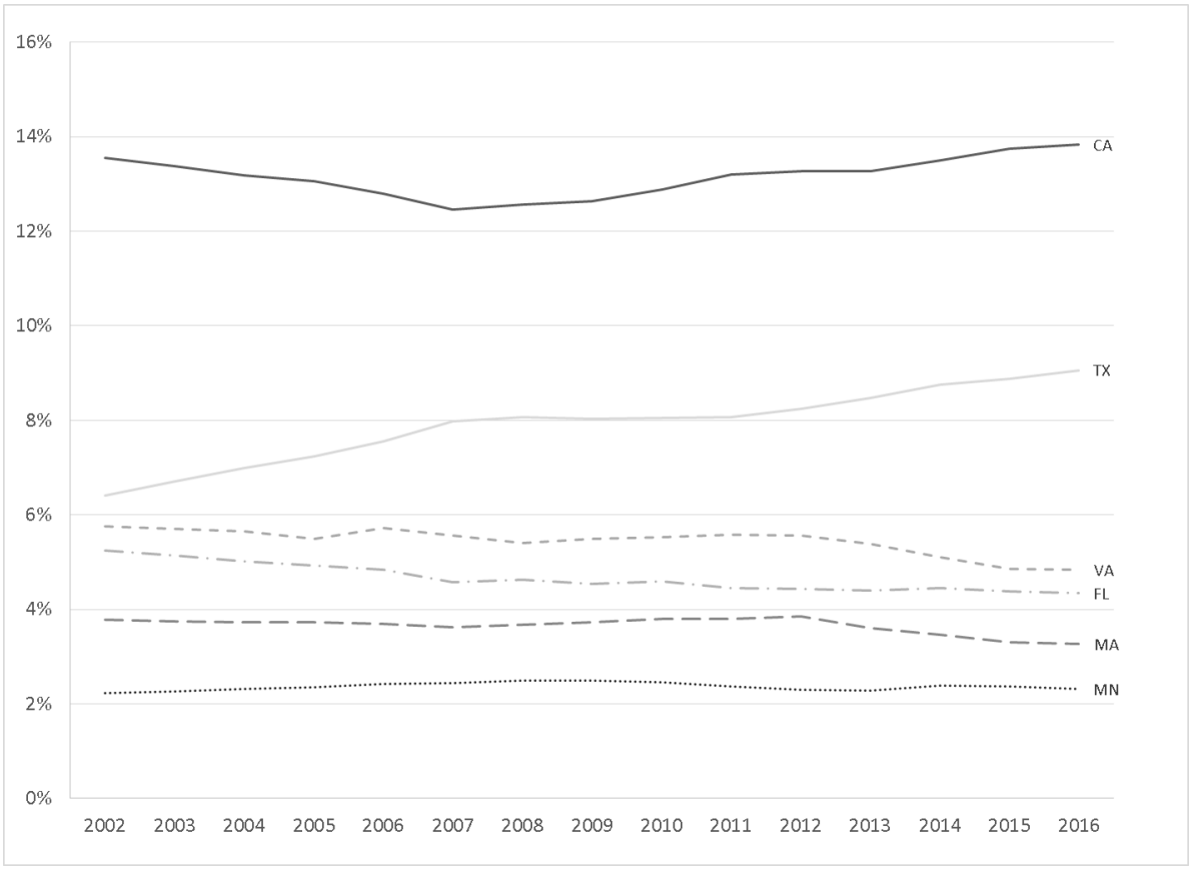


Figure 7: variation in state share of software engineers*



*linear imputation for years 2002-2004 and 2016

Figure 8a: Share of devices with software: top-10 states for *all* software devices vs. rest

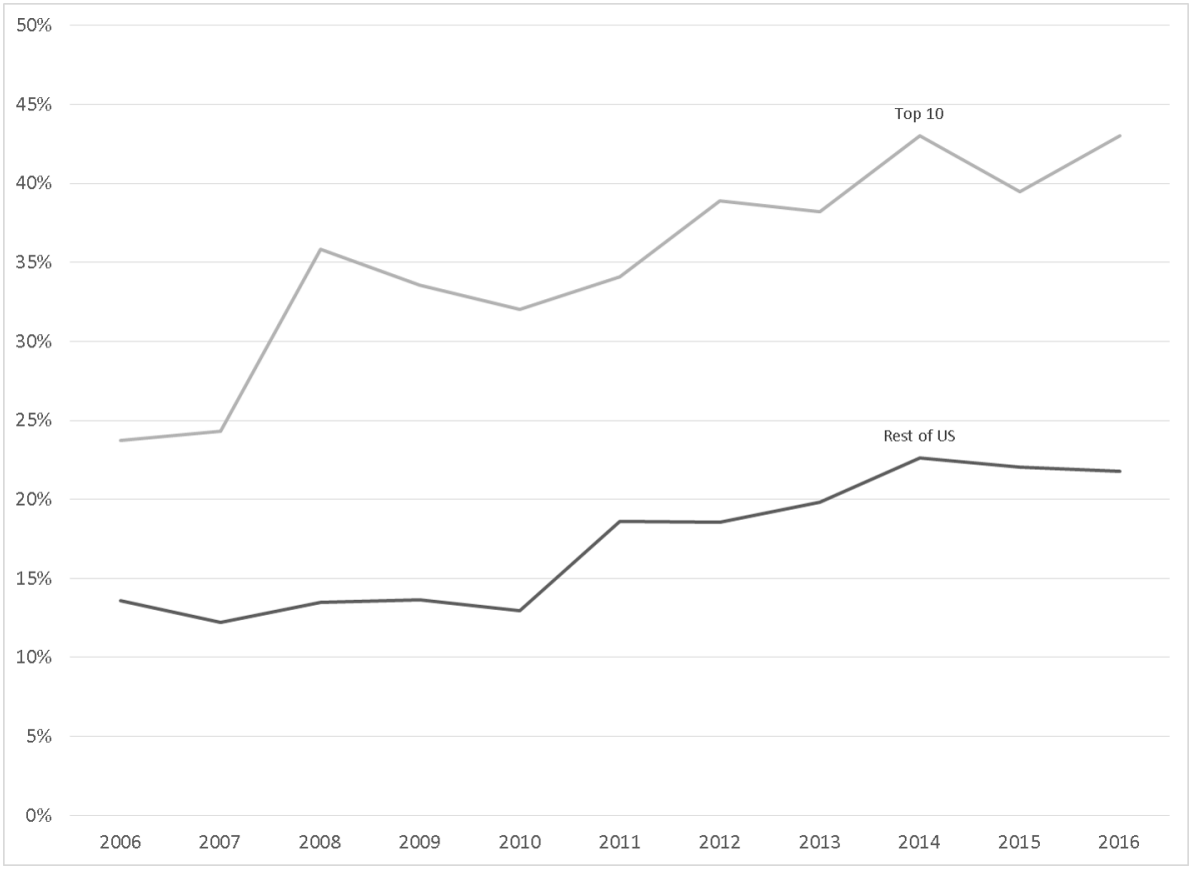
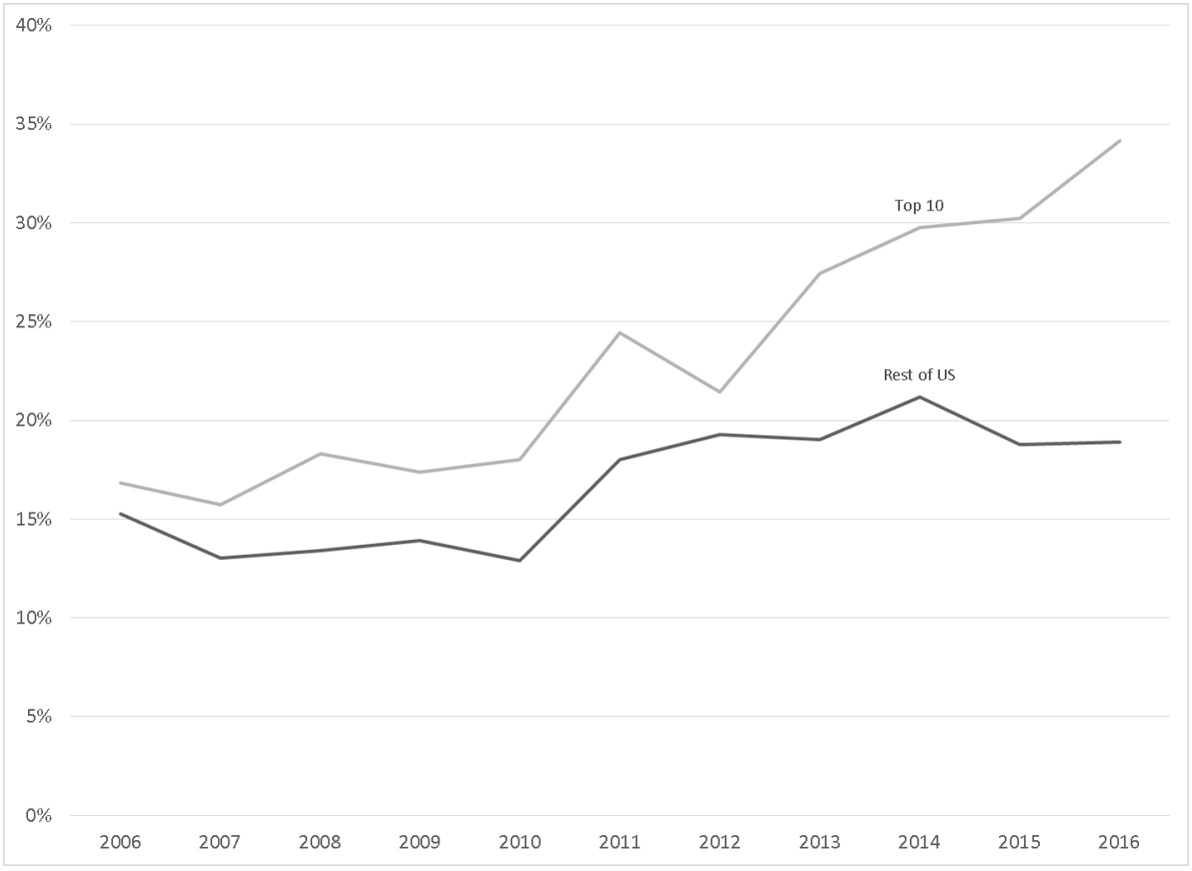


Figure 8b: Share of devices with software: top-10 states for software devices in *class* vs. rest



Tables

Table 1: Summary statistics by medical specialty (class)

	Unique devices at p-y* level		Unique devices at f-p-y* level		Unique devices at f-p-y level**	
	N	%	N	%	N	%
Cardiovascular	6,092	17.0 %	4,643	17.0 %	2,761	17.6 %
Clinical Chemistry	2,353	6.6 %	1,845	6.8 %	956	6.1 %
Dental	3,942	11.0 %	3,207	11.7 %	1,718	10.9 %
Gastroenterology, Urology	2,571	7.2 %	2,156	7.9 %	1,281	8.1 %
General Hospital	3,779	10.6 %	3,037	11.1 %	1,432	9.1 %
General, Plastic Surgery	4,959	13.9 %	3,851	14.1 %	2,285	14.5 %
Orthopedic	7,228	20.2 %	5,194	19.0 %	3,566	22.7 %
Radiology	4,870	13.6 %	3,377	12.4 %	1,732	11.0 %
Total	35,794	100.0 %	27,310	100.0 %	15,731	100.0 %

*f=firm, p=product, y=year

**post 2005, US only

Table 2: Firm financial data

Metric	Sample Mean (± SD)
Revenue, all years $USD (,000s)	16,068 ± 32,477
Revenue, 2006 $USD (,000s)	13,420 ± 30,919
Revenue, 2011 $USD (,000s)	17,308 ± 34,091
Revenue, 2016 $USD (,000s)	7,474 ± 5,489
Prior analog devices	23.77 ± 44.76
Prior software devices	4.23 ± 14.53
Prior analog devices (class)	13.73 ± 26.76
Prior software devices (class)	2.19 ± 8.52
Publicly listed, %	28.95
VC funded (applicant), %	16.17
Based in the United States, %	100.00

Notes: Non-binary variables are given as mean ± SD
n=15,731 except for publicly listed firms, n=4,554
Prior analog/software devices calculated using keyword-based definition
Revenue data was left-truncated at $0 and right-winsorized at the 99.5% level

Table 3: Firm experience data

	Full FYP sample	Analog	Software	T-Statistic
Software	(n=15,731)	(n=12,673)	(n=3,058)	(n=15,731)
Prior software devices	4.23	2.71	10.51	-17.17
Prior analog devices	23.77	25.39	17.08	10.46
Prior software devices (in class)	2.19	1.05	6.93	-19.54
Prior analog devices (in class)	13.73	15.08	8.12	17.22
Software - MTI	(n=15,731)	(n=13,557)	(n=2,174)	(n=15,731)
Prior software devices	3.36	2.22	10.47	-15.48
Prior analog devices	24.64	26.02	16.02	12.02
Prior software devices (in class)	1.76	0.90	7.11	-16.90
Prior analog devices (in class)	14.16	15.13	8.12	16.26

Notes: Software devices defined using keyword-based classification. T-statistic is from a difference-in-means t-test with unequal variances between the analog and software samples. All tests have a corresponding p-value of < 0.000

Table 4: Controls (used in all regression models)

Logit model, software inclusion at FPY level				
	(1)	(2)	(3)	(4)
Clearance year=2007	-0.008	0.006		
	(0.009)	(0.016)		
Clearance year=2008	-0.002	0.010		
	(0.012)	(0.019)		
Clearance year=2009	-0.003	0.002		
	(0.011)	(0.018)		
Clearance year=2010	0.000	-0.012		
	(0.013)	(0.019)		
Clearance year=2011	0.050***	0.057**		
	(0.012)	(0.020)		
Clearance year=2012	0.046***	0.067***		
	(0.013)	(0.019)		
Clearance year=2013	0.069***	0.079***		
	(0.013)	(0.020)		
Clearance year=2014	0.099***	0.121***		
	(0.016)	(0.023)		
Clearance year=2015	0.104***	0.117***		
	(0.017)	(0.022)		
Clearance year=2016	0.115***	0.139***		
	(0.016)	(0.029)		
Clearance year			0.014***	0.015***
			(0.001)	(0.002)
Cardiovascular	0.075	0.002	0.073	0.003
	(0.072)	(0.067)	(0.072)	(0.067)
Dental	-0.127*	-0.086	-0.129*	-0.086
	(0.061)	(0.066)	(0.061)	(0.066)
Gastroenterology, Urology	-0.056	-0.032	-0.058	-0.033
	(0.060)	(0.060)	(0.060)	(0.060)
General Hospital	-0.085	-0.072	-0.087	-0.072
	(0.065)	(0.074)	(0.065)	(0.074)
General, Plastic Surgery	-0.080	-0.079	-0.082	-0.079
	(0.061)	(0.059)	(0.061)	(0.059)
Orthopedic	-0.190***	-0.185***	-0.192***	-0.184***
	(0.055)	(0.052)	(0.055)	(0.052)
Radiology	0.451***	0.446***	0.449***	0.447***
	(0.091)	(0.075)	(0.092)	(0.076)
N	15731	5452	15731	5452
Pseudo R^2	0.223	0.190	0.221	0.187

* p<0.05, ** p<0.01, *** p<0.001 Regression results for years 2006-2016, inclusive. Column 1 includes full U.S. sample. Column 2 represents limited sample of firms with available revenue data. Omitted group = Clinical Chemistry, marginal effects reported.

Table 5: Role of geographic & firm capabilities

Logit model predicting software inclusion								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ln state software employees	0.007			0.005				0.006
	(0.005)			(0.004)				(0.004)
In software device cluster		0.068***		0.032**				0.023*
		(0.012)		(0.010)				(0.010)
In software device cluster for prod. class			0.136***	0.129***				0.115***
			(0.010)	(0.009)				(0.007)
Ln prior software devices (internal)					0.039***		-0.023**	-0.017*
					(0.005)		(0.008)	(0.007)
Ln prior software devices in class (internal)						0.061***	0.086***	0.072***
						(0.007)	(0.012)	(0.011)
N	15731	15731	15731	15731	15731	15731	15731	15731

Notes: all models include full set of time and product class fixed effects, marginal effects reported
Standard errors are clustered at product code level
+ p<0.10 * p<0.05, ** p<0.01, *** p<0.001

Table 6: Role of financial resources

Logit model predicting software inclusion						
	(1)	(2)	(3)	(4)	(5)	(6)
Publicly listed firm	0.009		0.011	0.009	-0.002	0.004
	(0.011)		(0.011)	(0.020)	(0.020)	(0.020)
VC-funded firm		0.029*	0.030*	0.057**		0.066***
		(0.012)	(0.013)	(0.019)		(0.020)
Ln revenues at t-1					0.005	0.007*
					(0.003)	(0.003)
N	15731	15731	15731	5452	5452	5452

Notes: all models include full set of time and product class fixed effects
Standard errors are clustered at product code level, marginal effects reported
+ p<0.10 * p<0.05, ** p<0.01, *** p<0.001

Table 7: Role of financial resources

Logit model predicting software inclusion						
	(1)	(2)	(3)	(4)	(5)	(6)
Ln state software employees	0.006	0.024**		0.017*	0.015*	0.003
	(0.004)	(0.008)		(0.007)	(0.007)	(0.004)
In software device cluster	0.023*	0.020		0.019	0.020	0.022*
	(0.010)	(0.018)		(0.017)	(0.017)	(0.010)
In software device cluster for prod. class	0.115***	0.136***		0.135***	0.134***	0.115***
	(0.007)	(0.013)		(0.012)	(0.012)	(0.007)
Ln prior software devices (internal)	-0.017*	-0.013+		-0.005	-0.004	-0.013+
	(0.007)	(0.008)		(0.009)	(0.009)	(0.008)
Ln prior software devices in class (internal)	0.072***	0.071***		0.070***	0.069***	0.071***
	(0.011)	(0.011)		(0.011)	(0.011)	(0.011)
Publicly listed firm			0.003	0.035*	0.034*	-0.018+
			(0.020)	(0.014)	(0.014)	(0.010)
Ln total VC funding			0.025***		0.019***	0.014***
			(0.006)		(0.005)	(0.003)
Ln revenues at t-1			0.008**	-0.006+	-0.005	
			(0.003)	(0.003)	(0.003)	
VC-funded firm				0.058***		
				(0.016)		
N	15731	5452	5452	5452	5452	15731

Notes: all models include full set of time and product class fixed effects
Standard errors are clustered at product code level, marginal effects reported
+ p<0.10 * p<0.05, ** p<0.01, *** p<0.001

Table 8: Combined models

Logit model predicting software inclusion					
	(1)	(2)	(3)	(4)	(5)
Ln state software employees	0.004	0.005	0.005	0.004	0.005
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
In software device cluster	0.022*	0.021*	0.022*	0.021+	0.022*
	(0.010)	(0.010)	(0.010)	(0.011)	(0.010)
In software device cluster for prod. class	0.115***	0.115***	0.113***	0.115***	0.108***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)
Ln prior software devices (internal)	-0.014+	-0.021*	-0.019*	-0.014+	-0.014+
	(0.008)	(0.009)	(0.008)	(0.008)	(0.008)
Ln prior software devices in class (internal)	0.072***	0.076***	0.070***	0.072***	0.072***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
public	-0.018+	-0.015	-0.016	-0.018+	-0.018+
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Binary venture funding for applicant	0.039***	0.024*	0.003	0.038***	0.019
	(0.010)	(0.011)	(0.011)	(0.011)	(0.014)
VC-funded firm * Ln prior software devices		0.019*			
		(0.008)			
VC-funded firm * Ln prior software devices in class			0.060***		
			(0.013)		
VC-funded firm * In software cluster				0.004	
				(0.020)	
VC-funded firm * In software cluster for prod. class					0.038*
					(0.017)
N	15731	15731	15731	15731	15731
Notes: all models include full set of time and product class fixed effects, marginal effects reported					
Standard errors are clustered at product code level					
+ p<0.10 * p<0.05, ** p<0.01, *** p<0.001					

Appendix A: The 510(k) Process

The information in this appendix is taken directly from the FDA's official description of the 510(k) (premarket notification) process (http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/PremarketSubmissions/PremarketNotification510k/)

Introduction

Each person who wants to market in the U.S., a Class I, II, and III device intended for human use, for which a Premarket Approval (PMA) is not required, must submit a 510(k) to FDA unless the device is exempt from 510(k) requirements of the Federal Food, Drug, and Cosmetic Act (the Act) and does not exceed the limitations of exemptions in .9 of the device classification regulation chapters (e.g., 21 CFR 862.9, 21 CFR 864.9). There is no 510(k) form, however,21 CFR 807Subpart E describes requirements for a 510(k) submission. Before marketing a device, each submitter must receive an order, in the form of a letter, from FDA which finds the device to be substantially equivalent (SE) and states that the device can be marketed in the U.S. This order "clears" the device for commercial distribution.

A 510(k) is a premarket submission made to FDA to demonstrate that the device to be marketed is at least as safe and effective, that is, substantially equivalent, to a legally marketed device (21 CFR 807.92(a)(3)) that is not subject to PMA. Submitters must compare their device to one or more similar legally marketed devices and make and support their substantial equivalency claims. A legally marketed device, as described in 21 CFR 807.92(a)(3), is a device that was legally marketed prior to May 28, 1976 (preamendments device), for which a PMA is not required, or a device which has been reclassified from Class III to Class II or I, or a device which has been found SE through the 510(k) process. The legally marketed device(s) to which equivalence is drawn is commonly known as the "predicate." Although devices recently cleared under 510(k) are often selected as the predicate to which equivalence is claimed, any legally marketed device may be used as a predicate. Legally marketed also means that the predicate cannot be one that is in violation of the Act.

Until the submitter receives an order declaring a device SE, the submitter may not proceed to market the device. Once the device is determined to be SE, it can then be marketed in the U.S. The SE determination is usually made within 90 days and is made based on the information submitted by the submitter.

Please note that FDA does not perform 510(k) pre-clearance facility inspections. The submitter may market the device immediately after 510(k) clearance is granted. The manufacturer should be prepared for an FDA quality system (21 CFR 820) inspection at any time after 510(k) clearance.

What is Substantial Equivalence

A 510(k) requires demonstration of substantial equivalence to another legally U.S. marketed device. Substantial equivalence means that the new device is at least as safe and effective as the predicate. A device is substantially equivalent if, in comparison to a predicate it:

- has the same intended use as the predicate; **and**
- has the same technological characteristics as the predicate;

or

- has the same intended use as the predicate; **and**
- has different technological characteristics and the information submitted to FDA;
  - does not raise new questions of safety and effectiveness; **and**
  - demonstrates that the device is at least as safe and effective as the legally marketed device.

A claim of substantial equivalence does not mean the new and predicate devices must be identical. Substantial equivalence is established with respect to intended use, design, energy used or delivered, materials, chemical composition, manufacturing process, performance, safety, effectiveness, labeling, biocompatibility, standards, and other characteristics, as applicable.

A device may not be marketed in the U.S. until the submitter receives a letter declaring the device substantially equivalent. If FDA determines that a device isnotsubstantially equivalent, the applicant may:

- resubmit another 510(k) with new data,
- request a Class I or II designation through thede novoprocess
- file areclassification petition, or
- submit a premarket approval application (PMA).

Who is Required to Submit a 510(k)

The Act and the 510(k) regulation (21 CFR 807) do not specify who must apply for a 510(k). Instead, they specify which actions, such as introducing a device to the U.S. market, require a 510(k) submission.

The following four categories of parties must submit a 510(k) to the FDA:

1. Domestic manufacturers introducing a device to the U.S. market;

Finished device manufacturers must submit a 510(k) if they manufacture a device according to their own specifications and market it in the U.S. Accessories to finished devices that are sold to the end user are also considered finished devices. However, manufacturers of device components are not required to submit a 510(k) unless such components are promoted for sale to an end user as replacement parts. Contract manufacturers, those firms that manufacture devices under contract according to someone else? s specifications, are not required to submit a 510(k).

2. Specification developers introducing a device to the U.S. market;

A specification developer develops the specifications for a finished device, but has the device manufactured under contract by another firm or entity. The specification developer submits the 510(k), not the contract manufacturer.

3. Repackers or relabelers who make labeling changes or whose operations significantly affect the device.

Repackagers or relabelers may be required to submit a 510(k) if they significantly change the labeling or otherwise affect any condition of the device. Significant labeling changes may include modification of manuals, such as adding a new intended use, deleting or adding warnings, contraindications, etc. Operations, such as sterilization, could alter the condition of the device. However, most repackagers or relabelers are not required to submit a 510(k).

4. Foreign manufacturers/exporters or U.S. representatives of foreign manufacturers/exporters introducing a device to the U.S. market.

Please note that all manufacturers (including specification developers) of Class II and III devices and select Class I devices are required to follow design controls (21 CFR 820.30) during the development of their device. The holder of a 510(k) must have design control documentation available for FDA review during a site inspection. In addition, any changes to the device specifications or manufacturing processes must be made in accordance with the Quality System regulation (21 CFR 820) and may be subject to a new 510(k). Please see our guidance, "Deciding When to Submit a 510(k) for a Change to an Existing Device."

When a 510(k) is Required

A 510(k) is required when:

1. Introducing a device into commercial distribution (marketing) for the first time. After May 28, 1976 (effective date of the Medical Device Amendments to the Act), anyone who wants to sell a device in the U.S. is required to make a 510(k) submission at least 90 days prior to offering the device for sale, even though it may have been under development or clinical investigation before that date. If your device was not marketed by your firm before May 28, 1976, a 510(k) is required.

2. You propose a different intended use for a device which you already have in commercial distribution. The 510(k) regulation (21 CFR 807) specifically requires a 510(k) submission for a major change or modification in intended use. Intended use is indicated by claims made for a device in labeling or advertising. Most, if not all changes in intended use will require a 510(k). Please note that prescription use to over the counter use is a major change in intended use and requires the submission of a new 510(k).

3. There is a change or modification of a legally marketed device and that change could significantly affect its safety or effectiveness. The burden is on the 510(k) holder to decide whether or not a modification could significantly affect safety or effectiveness of the device. Any modifications must be made in accordance with the Quality System regulation, 21 CFR 820, and recorded in the device master record and change control records. It is recommended that the justification for submitting or not submitting a new 510(k) be recorded in the change control records.

A new 510(k) submission is required for changes or modifications to an existing device, where the modifications could significantly affect the safety or effectiveness of the device or the device is to be marketed for a new or different indication for use. SeeIs a new 510(k) required for a modification to the device?for additional information.

When a 510(k) is Not Required

The following are examples of when a 510(k) is not required.

1. You sell unfinished devices to another firm for further processing or sell components to be used in the assembling of devices by other firms. However, if your components are to be sold directly to end users as replacement parts, a 510(k) is required.

2. Your device is not being marketed or commercially distributed. You do not need a 510(k) to develop, evaluate, or test a device. This includes clinical evaluation. Please note that if you perform clinical trials with your device, you are subject to theInvestigational Device Exemption(IDE) regulation (21 CFR 812).

3. You distribute another firm's domestically manufactured device. You may place a label on the device, "Distributed by ABC Firm" or "Manufactured for ABC Firm," (21

CFR 801.1) and sell it to end users without submission of a 510(k).

4. In most cases, if you are a repackager or a relabeler you are not required to submit a 510(k) if the existing labeling or condition of the device is not significantly changed. The labeling should be consistent with the labeling submitted in the 510(k) with the same indications for use and warnings and contraindications.

5. Your device was legally in commercial distribution before May 28, 1976 and you have documentation to prove this. These devices are "grandfathered" and havePreamendment Status. You do not have to submit a 510(k) unless the device has been significantly modified or there has been a change in its intended use.

6. The device is made outside the U.S. and you are an importer of the foreign made medical device. A 510(k) is not required if a 510(k) has been submitted by the foreign manufacturer and received marketing clearance. Once the foreign manufacturer has received 510(k) clearance for the device, the foreign manufacturer may export his device to any U.S. importer.

7. Your device is exempted from 510(k) by regulation (21 CFR 862-892). That is, certain Class I or II devices can be marketed for the first time without having to submit a 510(k). A list of the Class I and II exempted devices can be found onMedical Device Exemptions 510(k) and GMP Requirements. However, if the device exceeds the limitations of exemptions in .9 of the device classification regulation chapters (e.g., 21 CFR 862.9, 21 CFR 864.9), such as the device has a new intended use or operates using a different fundamental scientific technology than a legally marketed device in that generic type of device, or the device is a reprocessed single-use device, then a 510(k) must be submitted to market the new device.

Preamendment Devices

The term "preamendments device" refers to devices legally marketed in the U.S. by a firm before May 28, 1976andwhich have not been:

- significantly changed or modified since then;and
- for which a regulation requiring a PMA application has not been published by FDA.

Devices meeting the above criteria are referred to as "grandfathered" devices and do not require a 510(k). The device must have the same intended use as that marketed before May 28, 1976. If the device is labeled for a new intended use, then the device is considered a new device and a 510(k) must be submitted to FDA for marketing clearance.

Please note that you must be theownerof the device on the market before May 28, 1976, for the device to be grandfathered. If your device is similar to a grandfathered device

and marketedafterMay 28, 1976, then your device does NOT meet the requirements of being grandfathered and you must submit a 510(k). In order for a firm to claim that it has a preamendments device, it must demonstrate that its device was labeled, promoted, and distributed in interstate commerce for a specific intended use and that intended use has not changed. SeePreamendment Status for information on documentation requirements.

Third Party Review Program

The Center for Devices and Radiological Health (CDRH) has implemented a Third Party Review Program. This program provides an option to manufacturers of certain devices of submitting their 510(k) to private parties (Recognized Third Parties) identified by FDA for review instead of submitting directly to CDRH. For more information on the program, eligible devices and a list of Recognized Third Parties go toThird Party Review Program Informationpage.

Appendix B: The PMA Process

The information in this appendix is taken directly from the FDA?s official description of the Premarket Approval process (http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/PremarketSubmissions/PremarketApprovalPMA/default.htm).

Overview

Premarket approval (PMA) is the FDA process of scientific and regulatory review to evaluate the safety and effectiveness of Class III medical devices. Class III devices are those that support or sustain human life, are of substantial importance in preventing impairment of human health, or which present a potential, unreasonable risk of illness or injury. Due to the level of risk associated with Class III devices, FDA has determined that general and special controls alone are insufficient to assure the safety and effectiveness of class III devices. Therefore, these devices require a premarket approval (PMA) application under section 515 of the FD&C Act in order to obtain marketing clearance. Please note that some Class III preamendment devices may require a Class III 510(k). See "Historical Background" for additional information.

PMA is the most stringent type of device marketing application required by FDA. The applicant must receive FDA approval of its PMA application prior to marketing the device. PMA approval is based on a determination by FDA that the PMA contains sufficient valid scientific evidence to assure that the device is safe and effective for its intended use(s). An approved PMA is, in effect, a private license granting the applicant (or owner) permission to market the device. The PMA owner, however, can authorize use of its data by another.

The PMA applicant is usually the person who owns the rights, or otherwise has authorized access, to the data and other information to be submitted in support of FDA approval. This person may be an individual, partnership, corporation, association, scientific or academic establishment, government agency or organizational unit, or other legal entity. The applicant is often the inventor/developer and ultimately the manufacturer.

FDA regulations provide 180 days to review the PMA and make a determination. In reality, the review time is normally longer. Before approving or denying a PMA, the appropriate FDA advisory committee may review the PMA at a public meeting and provide FDA with the committee's recommendation on whether FDA should approve the submission. After FDA notifies the applicant that the PMA has been approved or denied, a notice is published on the Internet (1) announcing the data on which the decision is based, and (2) providing interested persons an opportunity to petition FDA within 30 days for reconsideration of the decision.

The regulation governing premarket approval is located in Title 21 Code of Federal Regulations (CFR)Part 814, Premarket Approval. A class III device that fails to meet PMA requirements is considered to be adulterated under section 501(f) of the FD&C Act and cannot be marketed.

When a PMA is Required

PMA requirements apply to Class III devices, the most stringent regulatory category for medical devices. Device product classifications can be found by searching the Product Classification Database. The database search provides the name of the device, classification, and a link to the Code of Federal Regulations (CFR), if any. The CFR provides the device type name, identification of the device, and classification information.

A regulation number for Class III devices marketed prior to the 1976 Medical Device Amendments is provided in the CFR. The CFR for these Class III devices that require a PMA states that the device is Class III and will provide an effective date of the requirement for PMA. If the regulation in the CFR states that "No effective date has been established of the requirement for premarket approval," a Class III 510(k) should be submitted.

Please note that PMA devices often involve new concepts and many are not of a type marketed prior to the Medical Device Amendments. Therefore, they do not have a classification regulation in the CFR. In this case, the product classification database will only cite the device type name and product code. If it is unclear whether the unclassified device requires a PMA, use the three letter product code to search thePremarket Approval (PMA) database and the 510(k) Premarket Notification database. These databases can also be found by clicking on the hypertext links at the top of the product classification database web page. Enter only the three letter product code in the product code box. If there are 510(k)?s cleared by FDA and the new device is substantially equivalent to any of these cleared devices, then the applicant should submit a 510(k). Furthermore, a new type of device may not be found in the product classification database. If the device is a high risk device (supports or sustains human life, is of substantial importance in preventing impairment of human health, or presents a potential, unreasonable risk of illness or injury) and has been found to be not substantially equivalent (NSE) to a Class I, II, or III [Class III requiring 510(k)] device, then the device must have an approved PMA before marketing in the U.S. Some devices that are found to be not substantially equivalent to a cleared Class I, II, or III (not requiring PMA) device, may be eligible for the de novo process as a Class I or Class II device. For additional information on the de novo process, see the guidance "New section 513(f)(2) - Evaluation of Automatic Class III Designation: Guidance for Industry and CDRH Staff" as well as the Evaluation of Automatic Class III Designation (De Novo) Summaries webpage.

Devices Used in Blood Establishments

The Center for Biologic, Evaluation, Research (CBER) has expertise in blood, blood products, and cellular therapies as well as the integral association of certain medical devices with these biological products. To utilize this expertise marketing and investigational device submissions (Premarket Notification, Premarket Approval, and Investigational Device Exemption) for medical devices associated with the blood collection and processing procedures as well as those associated with cellular therapies are reviewed by CBER. Although these products are reviewed by CBER, the medical device laws and regulations still apply. Thelist

of medical devices reviewed by CBERare available on the Internet. In addition to CDRH guidance on Premarket Approval,specific medical device guidance for devices reviewed by CBERis available at online or by contacting:

Center for Biologics Evaluation and Research
Office of Communication, Training and Manufacturers Assistance (HFM-43)
1401 Rockville Pike, Room 200N
Rockville, MD 20852-1448 U.S.A.
Telephone Number: 301-827-2000 or 800-835-4709
Fax Number: 301-827-3843

Data Requirements

A Premarket Approval (PMA) application is a scientific, regulatory documentation to FDA to demonstrate the safety and effectiveness of the class III device. There are administrative elements of a PMA application, but good science and scientific writing is a key to the approval of PMA application. If a PMA application lacks elements listed in the administrative checklist, FDA will refuse to file a PMA application and will not proceed with the in-depth review of scientific and clinical data. If a PMA application lacks valid clinical information and scientific analysis on sound scientific reasoning, it could impact FDA's review and approval. PMA applications that are incomplete, inaccurate, inconsist, omit critical information, and poorly organized have resulted in delays in approval or denial of PMA applications. Manufacturers should perform a quality control audit of a PMA application before sending it to FDA to assure that it is scientifically sound and presented in a well organized format.

Technical Sections: The technical sections containing data and information should allow FDA to determine whether to approve or disapprove the application. These sections are usually divided into non-clinical laboratory studies and clinical investigations.

Non-clinical Laboratory Studies Section: Non-clinical laboratory studies section includes information on microbiology, toxicology, immunology, biocompatibility, stress, wear, shelf life, and other laboratory or animal tests. Non-clinical studies for safety evaluation must be conducted in compliance with21CFR Part 58(Good Laboratory Practice for Nonclinical Laboratory Studies). To assist you in determining the appropriate preclinical bench studies for your device, refer to the applicable guidance documents and standards identified in theProduct Classificationdatabase for your device. You may also seek input from the review branch via thePre-Submission Program.

Clinical Investigations Section: Clinical investigations section includes study protocols, safety and effectiveness data, adverse reactions and complications, device failures and replacements, patient information, patient complaints, tabulations of data from all individual subjects, results of statistical analyses, and any other information from the clinical investigations. Any investigation conducted under anInvestigational Device Exemption (IDE)must be identified as such.

Like other scientific reports, FDA has observed problems with study designs, study

conduct, data analyses, presentations, and conclusions. Investigators should always consult all applicable FDA guidance documents, industry standards, and recommended practices. Numerous device-specific FDAguidance documents that describe data requirements are available. Study protocols should include all applicable elements described in the device-specific guidance documents.

Appendix C: Code of Federal Regulations Title 21

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

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

PART 807 – ESTABLISHMENT REGISTRATION AND DEVICE LISTING FOR MANUFACTURERS AND INITIAL IMPORTERS OF DEVICES Subpart E–Premarket Notification Procedures Sec. 807.92 Content and format of a 510(k) summary.
(a) A 510(k) summary shall be in sufficient detail to provide an understanding of the basis for a determination of substantial equivalence. FDA will accept summaries as well as amendments thereto until such time as FDA issues a determination of substantial equivalence. All 510(k) summaries shall contain the following information:

(1) The submitter's name, address, telephone number, a contact person, and the date the summary was prepared;

(2) The name of the device, including the trade or proprietary name if applicable, the common or usual name, and the classification name, if known;

(3) An identification of the legally marketed device to which the submitter claims equivalence. A legally marketed device to which a new device may be compared for a determination regarding substantial equivalence is a device that was legally marketed prior to May 28, 1976, or a device which has been reclassified from class III to class II or I (the predicate), or a device which has been found to be substantially equivalent through the 510(k) premarket notification process;

(4) A description of the device that is the subject of the premarket notification submission, such as might be found in the labeling or promotional material for the device, including an explanation of how the device functions, the scientific concepts that form the basis for the device, and the significant physical and performance characteristics of the device, such as device design, material used, and physical properties;

(5) A statement of the intended use of the device that is the subject of the premarket notification submission, including a general description of the diseases or conditions that the device will diagnose, treat, prevent, cure, or mitigate, including a description, where appropriate, of

the patient population for which the device is intended. If the indication statements are different from those of the legally marketed device identified in paragraph (a)(3) of this section, the 510(k) summary shall contain an explanation as to why the differences are not critical to the intended therapeutic, diagnostic, prosthetic, or surgical use of the device, and why the differences do not affect the safety and effectiveness of the device when used as labeled; and

(6) If the device has the same technological characteristics (i.e., design, material, chemical composition, energy source) as the predicate device identified in paragraph (a)(3) of this section, a summary of the technological characteristics of the new device in comparison to those of the predicate device. If the device has different technological characteristics from the predicate device, a summary of how the technological characteristics of the device compare to a legally marketed device identified in paragraph (a)(3) of this section.

(b) 510(k) summaries for those premarket submissions in which a determination of substantial equivalence is also based on an assessment of performance data shall contain the following information:

(1) A brief discussion of the nonclinical tests submitted, referenced, or relied on in the premarket notification submission for a determination of substantial equivalence;

(2) A brief discussion of the clinical tests submitted, referenced, or relied on in the premarket notification submission for a determination of substantial equivalence. This discussion shall include, where applicable, a description of the subjects upon whom the device was tested, a discussion of the safety or effectiveness data obtained from the testing, with specific reference to adverse effects and complications, and any other information from the clinical testing relevant to a determination of substantial equivalence; and

(3) The conclusions drawn from the nonclinical and clinical tests that demonstrate that the device is as safe, as effective, and performs as well as or better than the legally marketed device identified in paragraph (a)(3) of this section.

(c) The summary should be in a separate section of the submission, beginning on a new page and ending on a page not shared with any other section of the premarket notification submission, and should be clearly identified as a "510(k) summary."

d) Any other information reasonably deemed necessary by the agency.

[57 FR 18066, Apr. 28, 1992, as amended at 59 FR 64295, Dec. 14, 1994]

Appendix D: Historical Firm Ownership Data

Problem: Evaluate MedTech provides detailed firm ownership data as of most recent status.

Solution: Evluate MedTech also provides a full history of deals (acquisitions, majority stake purchases, etc.).

We are able to use the historical deal data to retrospectively re-assign correct "parent companies" (firms), such that each device clearance is paired with the available financial data of the firm that was responsible for bringing that device to market. As noted above, this correction resolves any concerns falsely attributing a new device to the firm that would later acquire it. This type of false attribution of a device to its future (but not current) parent firm would create measurement error to the extent that there is selection in the probability of firms being acquired, which we believe is likely to be the case.

Appendix E: Additional tables and results

Appendix Table 1: Machine readable documents by sample year

Year	Readable Documents	Total Products	% Readable
2002	2573	2587	99.5%
2003	2565	2579	99.5%
2004	2476	2505	98.8%
2005	2338	2364	98.9%
2006	2430	2450	99.2%
2007	2245	2318	96.9%
2008	2333	2382	97.9%
2009	2287	2333	98.0%
2010	2168	2242	96.7%
2011	2405	2452	98.1%
2012	2466	2502	98.6%
2013	2404	2428	99.0%
2014	2509	2552	98.3%
2015	2334	2408	96.9%
2016	2261	2328	97.1%
Total	35794	36496	98.1%

Notes: based on 8 most common medical specialty areas (classes)

Appendix Table 2: Keywords and overlap of each with MTI classification of software devices

Keyword (& acronyms thereof)*	Total devices	% Flagged by MTI as "software"
data	18,894	20%
internet	9,840	17%
software	6,788	73%
imaging	5,470	49%
display	5,107	50%
interface	3,728	40%
digital	3,249	47%
computer	2,779	58%
screen	2,278	49%
transmission	1,798	41%
platform	1,361	47%
network	1,187	62%
wireless	906	48%
database	757	57%
server	731	70%
programmable	714	48%
microprocessor	593	33%
digitally	464	27%
bit	418	58%
processor	381	48%
analog	359	39%
digitalimage	312	54%
ethernet	291	58%
bluetooth	287	35%
cpu	277	50%
LAN	232	66%
datastorage	223	57%
datacollection	221	45%
informationsyste	193	69%
touchscreen	183	31%
download	180	59%
online	161	48%
IT	133	39%
digitaldata	125	54%
harddisk	116	73%
bandwidth	110	63%

Notes: this list includes all keywords found in >100 unique product descriptions