Essays on Productivity and Innovation

A dissertation presented
by
Cirrus Foroughi
to
the Strategy Unit

in partial fulfillment of the requirements
for the degree of
Doctor of Business Administration
in the subject of
Strategy

Harvard Business School
Boston, Massachusetts
May 2020

Thesis advisor: Juan Alcácer Cirrus Foroughi

Essays on Productivity and Innovation

Abstract

This dissertation explores themes surrounding digital innovation and its effects on industries and firm- and employee-level productivity. Using novel and often proprietary sources of data, I take an empirical approach to asking how digital technology has transformed entire industries (e.g. medical devices and medical care), as well as how employees interact with tasks (e.g. technology-enabled remote work and telemedical care), with a focus on unpacking mechanisms driving each result. In each of these chapters, I focus on phenomena with large-scale impacts, from the digital transformation of a $150 billion dollar medical device industry in the US, to increased productivity at the US Patent and Trademark Office that could lead to $1.3 billion in value as a result of new patent approvals, to potentially saving 70,000 lives a year as a result of telemedical intervention in the intensive care sector in the US.

Contents

For Olivia

Acknowledgments

I CAN STILL REMEMBER THE FIRST TIME I KNEW I LOVED ECONOMICS—it was in your office, Doug Staiger, pouring through old California healthcare data (stored on compact discs) and writing my first STATA code. Thank you for sending me on this unforgettable journey. Bruce Sacerdote, Eric Zitzewitz, and Jon Skinner—thank you for your dedication and kindness. Dartmouth would not be the same without you, and I wouldn't be who I am either.

To Heidi Williams—thank you for teaching me the importance of rigor, and for your patience and kindness as I transitioned from college to full time research. And to the lifelong friends I made at the National Bureau of Economic Research — thank you for creating an environment that will forever feel like home.

I've been blessed with a wonderful and supportive committee and set of mentors over the past few years. Ariel, I would have been completely lost in grad school had you not taken me under your wing. You've always been there and have been incredibly generous with your time and advice. Juan, thank you for introducing me to the Strategy unit and believing in me. Shane—some of my fondest memories in grad school were made sitting in your office, going off on tangents about technology. And Amitabh, thank you so much for being a fantastic collaborator, mentor, and friend.

To Saurabh Chandra and the rest of the team at Northwell Health — this dissertation would not have been possible without your kindness, patience, and generosity.

I've built an incredible set of friendships over the course of grad school. Brandon, thanks for years of laughs and a sense of purpose built from a WeWork in Central Square. Ben, Ian — thank you for being fantastic roommates and for making what could have been a transient living situation feel permanent. I'll always look back fondly upon memories of staying up late and somehow mixing laughter with talk of econometrics. And Tamar — thank you for one my closest and longest friendships, from the moment we walked into the NBER back in 2013.

Elise, thank you for bringing joy to my life while I was putting this all together over the past few months.

Olivia, this last year of grad school has been full of highs and lows. Thank you for your love, optimism, support, and most of all, thank you for bringing laughter into my life. Even though we may have grown up in adjacent towns, I'm forever thankful that our paths crossed here at Harvard.

And finally, to my parents, Kaivan and Nafeesa, for raising me to be who I am today, for teaching me the value of education, for sacrificing so much, and for all of the love you had to give.

0

Introduction

This dissertation explores themes surrounding digital innovation and its effects on industries and firm- and employee-level productivity. Using novel and often proprietary sources of data, I take an empirical approach to asking how digital technology has transformed entire industries (e.g. medical devices and medical care), as well as how employees interact with tasks (e.g. technology-enabled remote work and telemedical care), with a focus on unpacking mechanisms driving each result. In each of these chapters, I focus on phenomena with large-scale impacts, from the digital transfor-

mation of a $150 billion dollar medical device industry in the US, to increased productivity at the US Patent and Trademark Office that could lead to $1.3 billion in value as a result of new patent approvals, to potentially saving 70,000 lives a year as a result of telemedical intervention in the intensive care sector in the US.

Chapter 1, titled "Who Drives Digital Innovation: Evidence from the U.S. Medical Device Industry" (joint with Ariel D. Stern), asks whether the large-scale technological change that is characteristic of an industry-wide digital transformation entrenches industry leaders or enables the rise of new entrants? We study the medical device industry, a unique setting in which we observe all new product commercialization over several years and in which the introduction of software has created fresh opportunities for new product development. Pioneering a new application of text analysis, we consider over 35,000 new medical devices that came to market in the United States from 2002 to 2016 in order to identify digital products. We examine the relative importance of within-firm know-how, geography, and financial resources in predicting digital new product development. We find that prior product-area commercialization experience and location in a region of concentrated expertise reinforce one another as predictors of digital innovation. Access to financing through public capital markets and venture capital are also positive predictors, but the magnitudes of these effects are smaller and do not appear to compensate for past product experience or geography. We conclude that the digital transformation of the medical device industry is disproportionately driven by product area and geographic incumbents.

Chapter 2, titled "Work-from-anywhere: The Productivity Effects of Geographic Flexibility" (joint with Raj Choudhury and Barbara Larson), asks how technology-enabled remote work programs affect productivity. While traditional "work-from-home (WFH)" programs offer the worker temporal flexibility, "work-from-anywhere (WFA)" programs offer both temporal and geographic flexibility. We study the effects of WFA on productivity at the United States Patent and Trademark Office (USPTO) and exploit a natural experiment in which the implementation of WFA was driven

by negotiations between managers and the patent examiners' union, leading to exogeneity in the timing of individual examiners' transition to WFA. WFA resulted in a 4.4% increase in output compared to when the worker was on WFH, without affecting the incidence of rework. We employ micro-data on geographic peers, an exogenous mandate to use IT, and proxies for examiner effort to shed further light on mechanisms. Finally, we study how flexibility affects the location choice of WFA, noting that employees move to lower cost-of-living areas, as well as noting correlations between career stage and advantageous locations (such as the Eastern District of Texas).

Chapter 3, titled "IT and Productivity: Evidence from Telemedicine", addresses the impact of IT on skilled-worker productivity by studying the implementation of IT in a non-routine, skill-intensive setting, the provision of intensive medical care by doctors, and asks whether IT is outcomes-enhancing and induces workers to shift effort across tasks. Using data from a large hospital network, I estimate the effects of a network-wide telemedical intervention in which in-person critical care physicians are almost entirely replaced by software algorithms and remote treatment. Leveraging detailed patient medical records and clinically-validated health risk measures, I estimate a 16% average reduction in mortality. This average effect masks substantial heterogeneity associated with IT usage: the healthiest and absolute sickest patients experience small increases in mortality, while more moderate patients benefit. Treatment rates appear flat, however, physicians redirect treatment towards patients that benefit most on the margin, pointing towards a more optimal allocation of effort with IT. Further analysis points to the importance of software—both the average effects and heterogeneity appear driven by the use of algorithms in monitoring and directing care. Complementary assets are also important—improvements associated with automation and technology use are greatest in technology-focused locations and diminish for community hospitals lacking on-site resources. Back-of-the-envelope calculations imply that adoption of this technology at scale across the United States could lead to 70,000 lives saved per year, at a value of $20 to $135 billion annually. Ultimately, implementation of telemedical technology at this hospital network is labor-substituting,

with the network reporting $3 million in annual cost reductions, comprised almost entirely of reductions in physician staffing needs for hospitals that have implemented eICU technology.

1

Who Drives Digital Innovation: Evidence from the U.S. Medical Device Industry

In recent years, major industries ranging from manufacturing and inventory management to entertainment to health care have undergone a "digital transformation," in which key aspects of both day-to-day business and new frontiers of product devel-

opment have migrated to a primarily digital (i.e., software-driven) context. This scenario raises questions about how new opportunities for digital product development impact both new entrants and experienced industry leaders. Does widespread technological change strengthen incumbent power or does it provide greater opportunity for new entry? The answer is not clear *ex ante*: incumbent firms have substantial experiential and resource advantages that could allow them to both weather and take advantage of large-scale technological change, while younger, leaner firms are not burdened by existing research and development (R&D) approaches and are often characterized by creativity and flexibility[75].

This study implements our research question in a novel setting: the digital transformation of the medical device industry. An advantage of this setting is that all product commercialization is directly observable. The existence of a centralized, national regulatory approval process, combined with detailed databases that assign all devices to standardized product areas, allows for precise and comprehensive categorization of new products, which can, in turn, be directly linked to commercializing firms and locations. Using text analysis and document classification techniques, we characterize over 35,000 medical devices brought to market from 2002 to 2016 and identify the digital products among them. We describe the commercialization of these products across medical specialty areas, across types of innovator firms, and over time.

We find that within-firm experience and geographic expertise are the strongest predictors of digital product commercialization. While a firm's previous experience with digital devices is broadly relevant, experience that is specific to a particular medical specialty area (product class) is *additionally* predictive of follow-on innovation (i.e., the commercialization of new versions of already-marketed digital products). This finding points to likely within-firm spillovers from experience in this context. Geography—in particular, a firm's location within a cluster—is a universally strong predictor of digital innovation. Interestingly, this effect is specific to digital experience within a product class, suggesting that geographic expertise is most relevant for digital innovation when knowledge is spe-

cialized. Overall, we find that digital innovation is led by *generalist* firm experts, but is more likely to occur in *specialized* clusters.

With respect to financial resources, we find that access to capital is associated with higher rates of digital innovation, but that these effects are smaller than those associated with experience and geographic expertise. Public firms are more likely to commercialize digital products, but these firms are, by definition, already incumbents. Among newer entrants, our estimates suggest that an order-of-magnitude increase in venture capital (VC) funding would be needed to offset the positive effects of being in a specialized product cluster. Further, the results imply that a doubling of VC funding would be necessary to offset just a single-digit percent increase in a firm's recent experience with related digital devices.

Finally, we observe that different types of access to capital matter in slightly different ways: publicly listed firms are particularly likely to engage in digital innovation of durable medical devices (those used multiple times, often based in hospitals; e.g., ultrasound equipment), whereas VC-funded firms are more likely to pursue digital innovation in single-use devices (those used for only one patient; e.g., pacemakers and insulin pumps). These findings are consistent with stronger inter-temporal spillovers from digital innovation among established firms.

Previous studies have highlighted the importance of software and digitization in determining how firms innovate (Arora et al. [24]; Branstetter et al. [39]) and perform [43]. Our study builds in many ways on the literature linking software and networking capabilities to innovative activity; it differs, however, from previous studies, in that our primary measure of innovation goes beyond patenting activity to assess the precise and complete set of new products that are 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 tail end of the innovation process, focusing on the final phase of new product development. Furthermore, this study is distinct in that it models digital innovation as a *dependent* variable, whereas other studies have frequently

treated the use of software as an independent variable.

1.1 Background

In the United States, health care spending makes up nearly 18% of the economy[9], offering a large potential market for new technologies and a variety of opportunities for innovators to build and grow businesses around new products. A growing segment of the health care market is "digital health," which is broadly defined to include companies and products at the intersection of healthcare and technology.[1] The digital health space includes health care IT and information systems, as well as a host of companies that build and sell technologies such as wireless sensors, software-enabled diagnostic and imaging devices, and artificial intelligence software programs with health care applications. In recent years, there has been dramatic growth in funding for digital health[162], with notable private and public initiatives emerging to fund research and investment.[2]

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. Common examples include digital blood-glucose monitors and nearly all contemporary radiology devices, which combine equipment for imaging with software for image processing and display. Modern medical devices incorporate software for tasks ranging from simple blood pressure monitoring to the processing and analysis of computed tomography (CT) data. Today, digital medical technology is commonplace and its use is inescapable for health care delivery professionals: a recent report found that U.S. hospitals use an average of

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

[2] For example, Rock Health describes itself as "the first venture fund dedicated to digital health" (https://rockhealth.com/about) and the state of Massachusetts launched the Massachusetts Digital Healthcare Initiative 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).

10 to 15 "connected"—i.e., networked—digital devices per patient bed [132]. Yet until the late 20th century, few software-driven medical devices existed and it wasn't until the late 1990s that the U.S. Food and Drug Administration (FDA) first issued guidance on the incorporation of software into regulated medical devices (FDA [1]).

While the 21st century has seen rapid digitization of medical data as a result of the growth of electronic health records [168] and the creation of a multi-billion-dollar digital health investment space [161], the growth of software in medical devices has not yet been characterized across products or firms, nor is it tracked directly by regulators. The implications of digitization for shaping patterns of innovation and commercial leadership in this sector have therefore not yet been studied at scale.

1.1.1 Regulated Medical Devices

The FDA is the only 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, it regulates over two trillion dollars' worth of products annually, including all medical technologies [31]. 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).[3] Within the CDRH, the Office of Device Evaluation reviews new products.[4]

Devices are wide-ranging in their complexity and their risk to patients. They range from low-risk devices 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 replace-

[3] 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).

[4] Since 1976, the regulation of new medical devices has been governed by the Medical Device Amendments (MDA) to the Federal Food, Drug, and Cosmetic Act of 1938.

ment heart valves. While devices of the lowest risk are subject only to so-called "general controls" of labeling and of compliance with the FDA's good manufacturing practices,[5] moderate-risk and high-risk devices must submit applications to the FDA for regulatory clearance or regulatory approval, respectively.[6] The administrative data from these regulatory processes, along with each new product's formal description, are made publicly available at the time a device completes regulatory review. These documents constitute the main source of new product data used in our analyses.

As described in detail below, a growing number and share of devices now contain software. Such features allow for additional functionality, such as allowing physicians to remotely diagnose and monitor patients. For example, the CardioMEMS™ HF System allows for remote, wireless heart failure (HF) monitoring, which has been shown to reduce HF hospital admissions for at-risk patients.[7] Despite recognition of the increasingly digital nature of medical devices,[8] the FDA does not formally track the use of software in medical devices in its product-level regulatory data. As a result, the prevalence and growth trajectory of digital products and their distribution across medical specialty areas have not yet been broadly described. The first portion of this paper is therefore dedicated to using information embedded in the text of medical device summaries to identify digital medical devices and quantify their growth. Using text analysis and an off-the-shelf natural language processing tool for medical topic identification, we analyze 15 years of medical device product summaries. We then turn to a set of empirical exercises that model the drivers of digital innovation across firms in this industry.

---

[5] https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpcd/315.cfm

[6] https://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/Overview

[7] https://www.sjm.com/en/sjm/cardiomems

[8] See, for example, the FDA's growing list of guidance documents related to software in medical devices (FDA [1]; FDA [3]; FDA [6]; FDA [8]).

1.1.2 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 law that established this regulatory pathway. One important component of the 510(k) application is 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" along with "an explanation of how the device functions [and] the scientific concepts that form the basis for the device." The summary also describes "significant physical and performance characteristics of the device, such as device design, material used, and physical properties," making it a clear source of information on the product's key technological characteristics.[9] It is these summaries (and their equivalents for high-risk devices) that are used to construct the text database described below.

A sample 510(k) Summary can be seen in Appendix Exhibit 1. Appendix A provides additional detail on the 510(k) process.

1.1.3 High-risk Devices and the PMA Process

High-risk (Class III) devices are regulated 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.[10] Evidence from trials is presented to the FDA as one part of the PMA package [104].[11]

Like the 510(k) process, the PMA process includes a product-specific summary document, which

[9] https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=807

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

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

is made publicly available at the time the device is approved.[12] Much like 510(k) summaries, PMA summary documents contain information on indications for use and 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"—among other components.[13]

Appendix B contains additional detail on the PMA process; a sample PMA summary can be seen in Exhibit 2.

1.1.4 Software in Medical Devices

The integration of software into medical devices is a relatively recent phenomenon. The first traces of regulatory interest in software in medical devices go back to 1999, when the FDA released its first guidance document, outlining expectations and standards for software embedded in new medical technologies (FDA [1]). The FDA's guidance has been augmented and updated several times since (e.g. FDA [3]; FDA [2]; FDA [2]; FDA [8]) and today, medical devices that not only incorporate software but also functionally *rely* on it, are commonplace. Thousands of patients and their physicians have come to depend on software-enabled medical devices, ranging from imaging devices for radiology to software-enabled insulin pumps to implantable heart failure monitors capable of wireless transmission.

1.1.5 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 use and adoption of elec-

[12] These summaries are used along with their moderate-risk device equivalents in the analysis below.

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

tronic health records (EHRs), one of the primary ways in which software and information technology have impacted health care delivery in the past decade (e.g. Adler-Milstein et al. [18]; Agha [20]; Dranove et al. [63]; Lee et al. [108]; Lin et al. [113]). These studies have documented the ongoing adoption of EHRs along with the heterogeneous (and typically limited or delayed) impacts on patient outcomes.

Our study is also related to a small literature on the adoption and use of software and information technology elsewhere in health care delivery. For example, [25] find that basic digitalization of emergency services ("911") increased the short-term survival rate of patients in cardiac distress. Other researchers have considered subtler regulatory factors in health care, such as data privacy laws [125], in order to understand how new technologies are adopted and used by patients. In the context of telehealth, [62] describe the major trends with a particular focus on home and mobile device applications. Yet beyond these studies, management and economics research at the intersection of digitization and health care is scant and the impacts of digitization on health care innovation have not been rigorously examined.

1.1.6 Determinants of Innovation

Firm experience and incumbency have been shown to drive innovative activity in contexts ranging from biotechnology [95] and pharmaceuticals (Nerkar & Roberts [130]; Morton [128]) to computer and IT hardware (Bayus & Agarwal [33]; King & Tucci [98]). In various settings and competitive environments, 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 [128], productivity spillovers in R&D activities [95], complementarities among technological and product-market experience [33], and experience with the process of new-market entry *itself* [98]. Using detailed commercialization histories, we are able to revisit the role of firm experience in the context of an industry undergoing

digitization.

A number of studies in management and economics have also highlighted the role of geography in innovative activity.[70] study the competing effects of colocation and coagglomeration of invention, 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[96] suggests similar dynamics; namely, local knowledge spillovers leading to geographic clustering of patent citations. In the related health care context of biotechnology,[119] highlights the role of knowledge spillovers and agglomeration economies in research-intensive sectors.

Finally, firm financial resources are thought to explain firms' innovation activities.[53] reviews the literature on this topic and concludes that in many—but not all—settings, cash flow is associated with higher R&D spending, noting that at least for smaller firms, the causality is thought to run from the former to the latter[92]. In the medical technology setting specifically, data suggest that small firms are particularly capital-constrained and rely on capital flows from both larger firms and venture capitalists to finance costly new product development[117]. Thus, our set-up also considers firms' access to capital (in particular, public markets and VC funding) as specific financial resources that may drive innovation.

1.2 Conceptual Framework

We outline a simple conceptual framework for predicting how firms make decisions to pursue costly new product development projects, given heterogeneity in location, prior experience, and financial resources. In particular, we emphasize that the necessity of a regulatory approval process, along with the accompanying time, financial investments, and institutional know-how required for successful new product development, generate differences in the relative costs of commercialization activities for different types of firms. Our framework focuses solely on the supply-side decision to enter a new

market.

A typical feature of digital products is low (or zero) marginal cost of provision to additional customers[81], however the cost of developing the digital technology in the first place may be quite large. We build on this intuition, noting that in the case of a multi-purpose technology such as software (for example, 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. Further, financial resources are known to shape R&D investments at the firm level (Cohen [53]; Hall & Lerner [90]; Kortum & Lerner [102]; and many others). In a setting where multiple factors can influence the cost of new product development, financial resources are another important lever that would be expected to impact the costs associated with digital innovation.

1.2.1 Framework for firm decision-making

A simple framework for considering firm investments can be seen in the following stylized two-period model: consider a firm, f, from geography g, facing a decision in period 1 ($t = 1$) regarding commercialization of a product in product class s. Commercializing a product involves costs, C_{fstg}, which include manufacturing and production costs, M_{fstg}, and financing costs, I_{fst}, e.g., for product design and R&D. That is, $C_{fstg} = c(M_{fstg}, I_{fst})$.[14]

Commercialization of a product results in expected revenues in period 2, r_{fst+1}. Firms will invest in commercializing new products when $C_{fstg} < r_{fst+1}$; that is, whenever net expected profits from a given product are positive:[15] $r_{fst+1} - C_{fstg} = \pi_{fstg+1} > 0$.

[14] Financing costs may vary by firm and product class and over time, but should not *further* vary by geography after accounting for other factors.

[15] A more detailed model could also account for the relevant discount rate. 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, which could also be included in a more detailed model; however, we note that since the average product lifecycle is just 1.5–2 years [171], it is realistic to assume that products should achieve profitability on a very short time horizon in order to justify commercialization.

We consider variation at the firm, product class, time, and geography levels in the empirical models that follow. In the remainder of this section, we leave off subscripts for simplicity.

1.2.2 Hypotheses

To preview, our conceptual framework predicts that the costs of commercializing a new product will vary with firm know-how, location, and financial inputs to R&D activities. C will therefore be decreasing in E and increasing in I. As a direct corollary, expected revenues in period 2 for firm f commercializing a given device in period 1, π_{t+1}, would be increasing in E and decreasing in I, making firms with more experience, those located in clusters, and/or those with lower financing costs more likely to pursue innovation. In this setup, four assumptions, which are consistent with both the theoretical and empirical literature, are required in order to take into account variation in firm commercialization decisions over time. All cross-partial derivatives of C and π can then be signed, leading to a set of testable hypotheses.

Assumption 1: Manufacturing costs are a function of a firm's labor costs, raw material costs, and prior commercialization experience, such that $M_{ftg} = m(L_{tg}, R_{tg}, E_{ft})$. We assume that all firms can access the same local labor and raw materials markets such that the remaining variation in the cost of manufacturing is only related to differences in know-how, E (that is, prior commercialization experience).

Assumption 2: We can further disaggregate E to allow for both firm- and geography-specific differences in know-how. More precisely, we allow for firm-level and area-level variation in prior commercialization experience such that $E = e(\alpha, \gamma)$, where α is the level of within-firm expertise and γ is the level of local geographic expertise. The importance of within-firm expertise has been extensively documented (Bayus & Agarwal [33]; Henderson & Cockburn [95]; King & Tucci [98]; Nerkar

& Roberts [130]; Morton [128]; and many others),[16] as has the role of regional expertise and geography in predicting innovative activity (Delgado et al. [59]; Forman et al. [70]; Mariani [119]; Jaffe et al. [96]; and many others).

Additionally, we note that within-firm and local geographic expertise can be categorized as general or class-specific, such that general experience considers a firm or location's commercialization experience across all software-driven medical devices, while *class-specific* experience considers only a firm or location's software-driven product commercialization experience within a specific FDA product class, such as digital radiology products. That is, E can be divided into general and class-specific components: $E = e(\alpha, \gamma, \alpha_s, \gamma_s)$.

Assumption 3: We expect that class-specific expertise is more transferrable to new product development than expertise outside the focal product class such that class-specific expertise reduces commercialization costs more than general expertise does. Thus, $\frac{\partial E}{\partial \gamma} < \frac{\partial E}{\partial \gamma_s}$ and $\frac{\partial E}{\partial \alpha} < \frac{\partial E}{\partial \alpha_s}$. Our hypotheses then can be stated as follows:

- **Hypothesis 1**: $\frac{\partial C}{\partial E}\frac{\partial E}{\partial \alpha s} < \frac{\partial C}{\partial E}\frac{\partial E}{\partial \alpha} < 0$.
  - **1a:** Within-firm know-how (previous experience) decreases commercialization costs.
  - **1b:** It does so in a way that is increasing in the specificity of within-firm experience.
- **Hypothesis 2**: $\frac{\partial C}{\partial E}\frac{\partial E}{\partial \gamma s} < \frac{\partial C}{\partial E}\frac{\partial E}{\partial \gamma} < 0$.
  - **2a:** Local geographic expertise (being located within a cluster) decreases commercialization costs.
  - **2b:** It does so in a way that is increasing in the specificity of local expertise.

[16] This is also consistent with the theory of economies of scope as described by [133] and as seen in empirical studies such as [95].

Assumption 4: We assume that a firm's financing cost, I, is correlated with its access to external capital, through either public capital markets or venture capital financing. This is consistent with literature linking firm performance and innovation to access to finance and financial constraints (Cohen [53]; Cohen & Klepper [54]; Hao & Jaffe [92]; Stern [156]). We can therefore write I as a decreasing function of (a) being publicly listed (having access to public capital markets), φ and (b) being VC-funded, v. We can then define I as a function i, where $I = i(\varphi, v)$ and $\frac{\partial I}{\partial \varphi}$ and $\frac{\partial I}{\partial v}$ <0.

The next set of hypotheses therefore address the implications of financial resources on expected patterns of commercialization, in which smaller and more capital-constrained firms will face higher costs of pursuing digital innovation:

- **Hypothesis 3**: $\frac{\partial C}{\partial I}\frac{\partial I}{\partial \varphi} < 0$. The cost of digital new-product development will be lower for publicly listed companies, making them more likely to commercialize new products.
- **Hypothesis 4**: $\frac{\partial C}{\partial I}\frac{\partial I}{\partial v} < 0$. The cost of digital new-product development will be lower for firms with venture capital funding, making them more likely to commercialize new products.

Importantly, we expect these dynamics to emerge as novel manifestations of firm advantage in digital new product commercialization. That is, the digital-product-specific components of firm experience, geography, and access to financing should matter *above and beyond* the advantages that these factors may confer already, when considering firms' commercialization patterns more broadly. We test each of these hypotheses in the analyses described below.

1.3 Data, Classification, and Summary Statistics

1.3.1 Summary

This project draws on four main sources of data. We begin with administrative data on all FDA-regulated moderate-risk and high-risk medical devices that came to market over 15 recent calendar years; namely 2002 to 2016, inclusive. For each device, we collect and analyze the text of the accompanying product summary or statement.

Using an automated script and two different types of supervised document classification, we identify and characterize digital (software-driven) devices. First, we document the incidence and frequency of 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)[17]—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 along with historical data about the location of a given product application and firm-level financial data, we characterize commercializing firms at the time each medical device in our dataset came to market.

1.3.2 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 came to market in the United States after 1996. Moderate-risk devices—such as hearing aids, blood pressure monitors and echocardiograph devices—are the largest category of devices regulated by the FDA, while high-risk devices—

[17] https://ii.nlm.nih.gov/MTI

such as pacemakers and drug eluting stents—make up a smaller share of new products. Moderate-risk device clearance happens through a process called "510(k)," while high-risk device approval occurs through the PMA process. Both processes are described briefly above and in detail in Appendices A and B, respectively. These processes are the final step of the research and development process, after which a cleared/approved 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 [115].

The FDA's 510(k) clearance database[18] and PMA approval database[19] include the full set of device names, product codes (three-letter classifications that categorize devices according to site of use and purpose), and submission and FDA decision dates for all products cleared/approved for marketing. 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). For each of these classes, there were over 2,000 unique new device approvals between January 1, 2002 and December 31, 2016.[20] Due to availability of product descriptions (as described below) this is our period of analysis. Over this period, 35,794 new regulated devices came to market in the United States. Each class of devices includes multiple product codes and (typically) multiple unique devices within each product code. Figure 1 presents a simple example of the hierarchy of the classification system.

[18] https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm

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

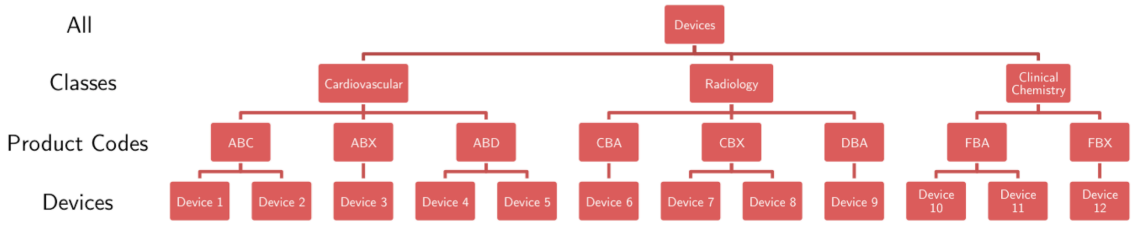
[20] These eight classes are defined using the full set of FDA clearance records available and therefore represent the universe of newly approved, FDA-regulated devices.

Table 1: Total counts of products by medical specialty (class) and overall

	Unique devices, full sample		Unique devices, regression sample*	
	n	%	n	%
Cardiovascular	6,092	17.0	3,884	17.5
Clinical Chemistry	2,353	6.6	1,318	6.2
Dental	3,942	11.0	2,282	10.9
Gastroenterology, Urology	2,571	7.2	1,647	8.2
General Hospital	3,779	10.6	1,920	9.3
General, Plastic Surgery	4,959	13.9	3,201	14.5
Orthopedic	7,228	20.2	5,346	22.4
Radiology	4,870	13.6	2,693	11.1
Total	35,794	100.0	22,291	100.0

*post 2004, US only

Figure 1: Device classification (example)



1.3.3 Rich Text Data

The second data source is a novel database of text files made up of the device summaries (standardized product descriptions). At the time of 510(k) clearance or PMA approval, a "summary" or "statement" is published for each device. As noted above, the summary must contain "a description of 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" (such as

design and physical properties).[21] In less than 10% of cases in our sample years, a related document called a "statement" was published in lieu of a summary,[22] in which case, we used the text from the statement instead. While somewhat less detailed than summaries, statements also contain relevant information about the content of products (for example, several included use of the word "software") and therefore provide the type of text information that is relevant for product classification in this study.[23] We hereafter use the term "summary" broadly to refer to both types of document.

Device summaries are published as online PDF documents following a standardized URL-format and we use an automated script to batch download all posted 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 summaries available. Using Abbyy FineReader optical character recognition (OCR) software, we convert downloaded documents into machine-readable text files. In total, ninety-eight percent of the product summaries could be converted to a machine-readable format giving us 35,794 device-text pairs.[24] We have no systematic concerns regarding selection or time trends in missing text data: the machine-readability of online PDFs is not statistically different across medical specialties overall, in any year, or over time. For all years, at least 97% of all digital documents were machine-readable following OCR document processing. Appendix C, Table I presents the number of machine-readable summaries in our sample by calendar year.

Although the use of text-based data—for example, categorizing phrases to document firm exten-

[21] https://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/PremarketSubmissions/PremarketNotification510k/ucm142651.htm; as noted above, Exhibits 1 and 2 present examples of 510(k) and PMA summaries, respectively.

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

[23] In theory, the use of statements could lead to under counting digital products if their text files are less detailed. Therefore, in robustness tests (unreported), we confirm that all results hold when considering the sample of product summary documents only.

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

sions into new products and services, as in [87]—has a well-established history in empirical analysis, the automation of these exercises is a relatively nascent phenomenon. [76] describe several techniques for 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." In recent years, text data has been used in studies ranging from sentiment analysis of policy uncertainty [32] to labor economics [60] and in the analysis of patent data [129]. Here, we demonstrate the utility of automated classification of product types at scale for understanding the content and functionality of new medical devices.

We process text files 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.[25] (A list of the 36 most frequently used keywords—each of which were found in over 100 unique product descriptions—can be found in Table II of Appendix C.) Unsurprisingly, the use of "software" and several related keywords has increased over time (Appendix C, Figure I). Because "software" is the most common among our search terms and is highly correlated with others, we rely on its inclusion in a product's description as our first indicator to identify digital products.

Categorizing products by keyword use is our first application of simple document classification to identify digital devices. It has the advantage of being simple and highly transparent, but the disadvantage of being somewhat *ad hoc*. However, this method is very successful 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 text related to features that are not included in the device itself. To put it simply, keywords such as "software" will not appear in the product description if they do not relate to aspects of the device's

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

functionality (see Appendix 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: we drew a random sample of 120 devices (eight per calendar year) that had been flagged for including "software" and manually inspected each of these devices' summaries. In this sample, 100% of devices flagged as including "software" were found to be correctly coded (that is, a 0% rate of type I error in this random subsample).

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 device descriptions using the MTI and generate an indicator for whether the algorithm assigned the MeSH code for *software* to the product.[26] The MeSH code for "software" broadly covers "sequential operating programs and data which instruct the functioning of a digital computer," a slightly higher bar for classifying digital products than searching for the keyword "software" alone.

The MTI algorithm has the advantage of having been externally validated by the NIH and by several years of use by the National Library of Medicine, but has two clear disadvantages. First, as noted above, we believe that the bar may be higher for flagging product descriptions for software inclusion (that is, identifying digital devices), since the MTI will require a *discussion* of software programs in the text, beyond simply using the keyword "software." For this reason, our expectation is that the MTI may identify a more software-intensive subset of products in our sample. Second, the MTI is non transparent in how it assigns concepts to text, since the algorithm itself has not been published.[27]

---

[26] 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.

[27] The MTI algorithm is not directly observable/open source; we batch-process text files through the algorithm and record the subject headings that the MTI returns as output.

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 software. However, as expected, not all summaries using the keyword "software" are identified by the MTI. The rightmost column of Table II in Appendix C presents a cross-tabulation of our *ad hoc* keyword-based document classification versus 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 sense of incorporating *any* software, we focus on the keyword-based definition for our primary analysis; however, for all regression models, we test the alternative (MTI-based) definition and present alternative versions of all key tables in Appendix C. The choice of definition does not appear to change the sign or statistical significance of the main results below, but magnitudes are attenuated roughly proportionally to the decrease in the number of software devices included in the MTI-defined sample.

Figure 2 presents the growth of new digital devices over our observation period. Figure 3a shows the growth in digitized product codes—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 has been 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 this period, while non-software product codes grew by only about 150% (albeit off a higher baseline). Figure 4b breaks down the growth of digital devices across medical specialty classes, revealing interesting heterogeneities. Although all classes show growth in digital products, the share of new products that are digital varies dramatically across medical specialty classes.

Figure 2: Number of newly approved digital devices by year

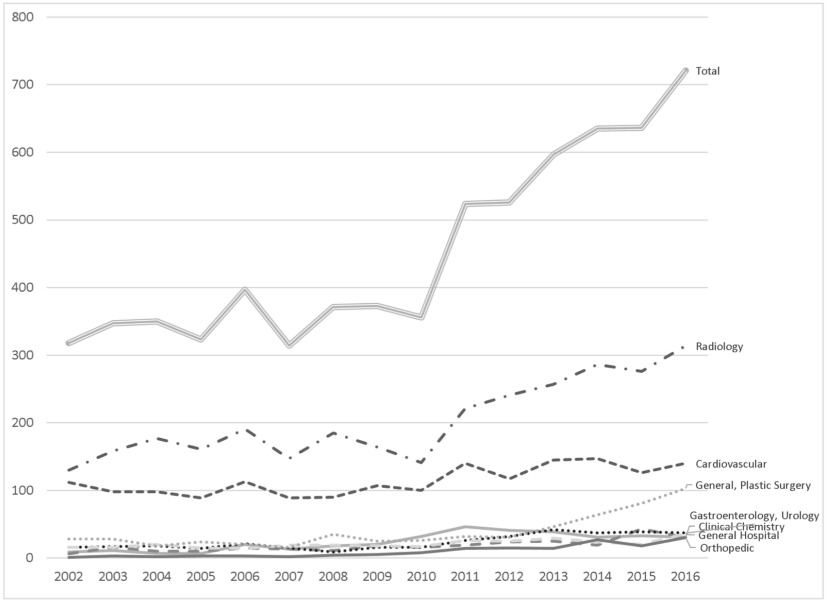


Figure 2: Number of newly approved digital devices by year

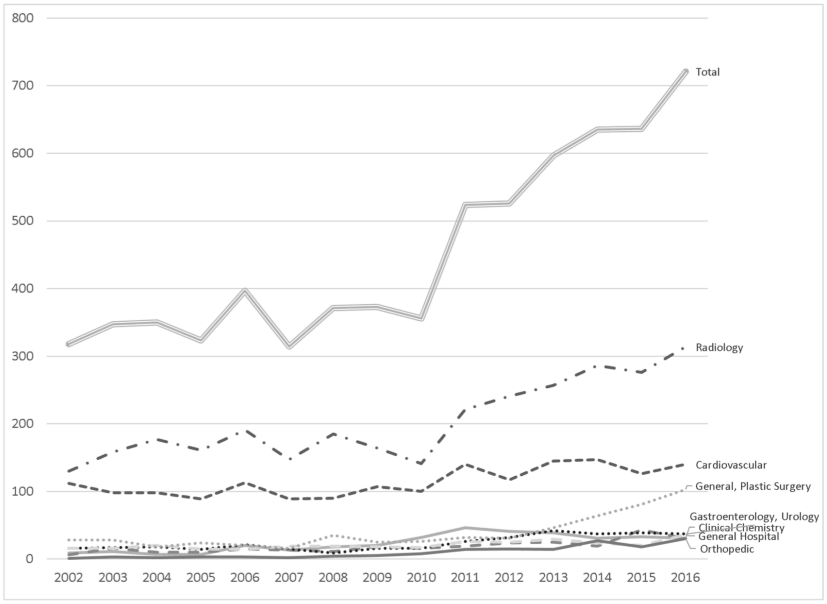


Figure 2: Number of newly approved digital devices by year

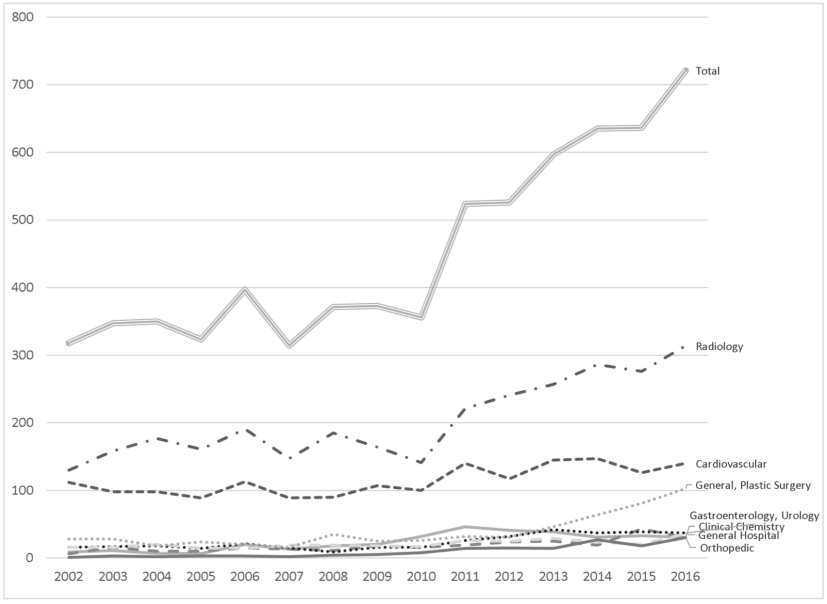
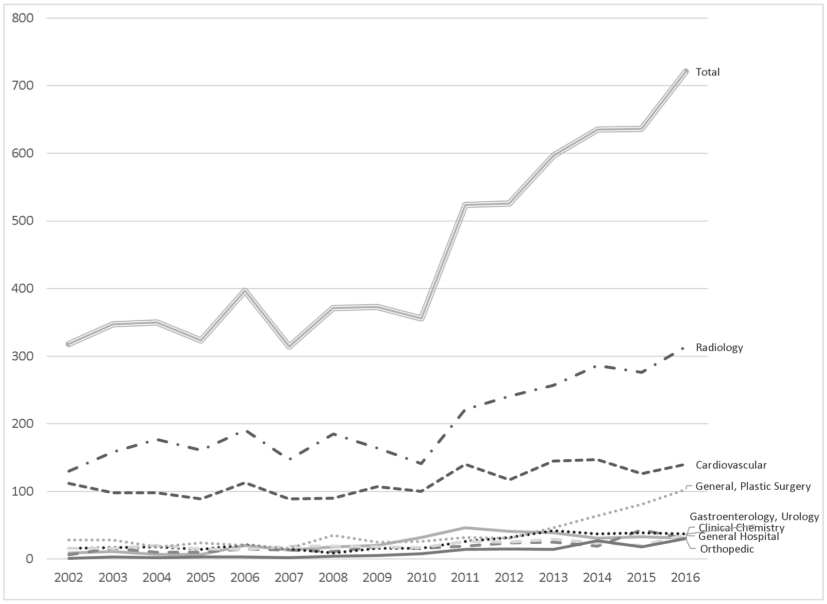


Figure 2: Number of newly approved digital devices by year



1.3.4 Measuring Firm Experience and Geographic Clusters

We characterize firms' digital device experience along two dimensions. First we calculate the count of digital devices the firm commercialized in the three years *prior* to the year of observation. This window is consistent with the medical device development period, which may run over two years [171]. Second, we calculate the count of digital devices within each *class* that the firm commercialized over the same three-year window of time (up to, but not including, the year of observation). We calculate total (digital and non-digital) devices commercialized by each firm in the same manner to be used as measures of overall firm experience.[28]

With respect to geography, we characterize firms as to whether or not they are in a "digital cluster" in three (increasingly specific) ways.[29] Each of these definitions requires limiting the sample to U.S.-based applications in order to operationalize a consistent definition of state-based geographic clusters. Notably, many of these applications are from U.S. offices of firms headquartered outside of the United States, so many large, international firms are in the final sample.

First, we consider local labor market expertise. Using annual data for 2016 from the U.S. Bureau of Labor Statistics (BLS), we compile data on each state's share of software engineers in the labor force[30] in order to consider whether there is a relationship between the characteristics of the skilled IT workforce in a state and the likelihood of digital innovation emerging from that state. Because each application includes an address, we can see the location of the facility from which a device application was submitted. Appendix Figure II presents a set of sample states. While there is some variation over time within states, the primary source of variation in the share of software engineers is

[28] in robustness tests, we also consider a two-year and a five-year window of past firm experience

[29] We also define overall device commercialization clusters in analogous ways to be used as controlls.

[30] Data were available for 2005 to 2015 inclusive; however, our analysis sample spans 2005 to 2016. In order to estimate 2016 data, we impute 2016 state employee shares based on a linear projection from the prior five years, which is similar to the method that BLS uses in its own projections.

across states.

Next we consider two types of state cluster for digital innovation, as defined by where device commercialization took place in preceding years. We define digital clusters by identifying the top 20% of states for digital device commercialization, based on a three-year moving average of the number of digital products brought to market in the period leading up to the year of observation. We then consider a class-specific version of this definition, in which we define the top 20% of states for digital device commercialization within each product class. Based on each of these definitions, we create an indicator variable for whether or not a device originated from a cluster. The sample used in regression analysis is limited to the years 2005 to 2016 (inclusive), to facilitate a three-year look-back on regional product expertise. For example, a data point in the year 2005 uses data from the years 2002, 2003, and 2004. Using these two definitions, Appendix Figures III and IV show the share of digital devices originating from clusters versus those not originating from clusters.

By using a three-year look-back to define clusters and to define firms' digital device experience, our analysis sample includes only the set of products that came to market in the years 2005-2016 (although data from the years 2002 to 2004 are used to define important firm measures in the early years of the analysis sample). Further, we limit our analysis sample products commercialized in the United States (including those commercialized by foreign firms with U.S. regulatory submission addresses), in order to tractably and consistently define geographic clusters. The final analysis sample includes 22,291 observations at the product level (Table 1). Table 2 presents summary statistics of all variables used in regression models.

Table 2: Summary statistics

Metric	Sample Mean (± SD)
Firm experience	
Prior digital devices (past 3 years)	3.65 ± 12.54
Prior digital devices in class (past 3 years)	2.24 ± 9.05
Geography	
In digital device cluster (general, past 3 years), %	15.30
In digital device cluster (class-specific, past 3 years), %	44.39
Financial resources	
Publicly listed, %	32.32
VC funded (applicant), %	16.63
Total venture funding, cumulative	6.29 ± 23.91
Total venture funding if any, cumulative	37.81 ± 47.42

Non-binary variables are given as mean ± SD
n=22,291
n=3,706 for cases with any total venture funding
Prior digital devices calculated using keyword-based definition

1.3.5 Firm Financial Data

Each device is linked by its commercializing entity to detailed firm financing data. We first link commercializing entities to a panel of firm acquisitions created using data from *EvaluateMedTech*[31] in order to account for subsidiary ownership and introduce the notion of *child* (acquired) and *parent* (acquirer) firms. These child and parent firms are then separately linked to data on each firm's public listing status and venture capital data. In order to link firm-level datasets, we use 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. Firm names were cleaned using a consistent set of rules to account for suffixes and abbreviations.[32]

Data on venture capital funding are assembled from *EvaluateMedTech* and *Preqin*, with prece-

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

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

dence given to the latter.[33] We observe deal dates and funding amounts for each linked firm, which we use in creating (a) lagged binary indicators for whether a firm was ever venture-funded or venture-funded prior to product commercialization and (b) running totals for dollar values of venture funding.

Data on firm public listing were collected from *EvaluateMedTech* and *Capital IQ*, with precedence given to the former, as it has broad coverage of the medical device industry.[34] These data allow us to create a binary indicator for whether the commercializing firm was publicly listed at the time a given product came to market. Appendix Figures V and VI show the share of digital devices that were commercialized by venture-capital–funded firms (in the sample of all privately-held firms) and by public firms, respectively.

1.4 Estimation and Results

In the estimation exercises that follow, we test the hypotheses outlined earlier. First, we explore evidence for Hypotheses 1 (firm experience) and 2 (geographic clusters) and the likelihood of a firm engaging in digital device innovation. Next, we explore evidence for Hypotheses 3 and 4 by modeling the relationship between firm financial resources (public capital and VC funding, respectively) and digital innovation. In combined models, we consider all factors simultaneously and explore mechanisms.

[33] Preqin is widely considered the best publicly available dataset for venture funding and has been used in a variety of recent studies (e.g. Harris et al. [93] and Korteweg & Nagel [101]).

[34] We validate *EvaluateMedTech* data using *Capital IQ*, long considered a primary source for detailed firm financials. See, for example, [16]; [38]; [151].

1.4.1 Overall Estimates

Trends in digital innovation in medical technology and the observed variation across medical specialty classes can be seen in Figures 2 to 4. Notably, there is significant heterogeneity across classes in the volume (Figure 2) and share (Figure 4b) of digital innovation. There are also clear time trends, with the number of new digital products growing over time. Among other things, the descriptive findings point to the importance of using product class and year controls in empirical models. We model the likelihood of digital innovation, D, as:

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

where the regressors include:

- Indicators of expertise (within-firm experience and geographic), each of which can be general or class-specific.
- Indicators of firm financial resources, including whether a product emerged from a publicly listed firm or a VC-funded firm.
- Controls for:
  - Clearance year, in order to capture time trends in software inclusion over time.
  - 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.
  - The firm's overall level of recent device innovation (that is, all product commercialization), such that any additional statistical relationships identified represent *additional* effects seen for digital devices.

  - The firm's location in a general device cluster (that is, all devices, not just digital products), such that any additional statistical relationships identified represent *additional* effects seen for digital devices.

- An error term, ε.

In the regression models below, all specifications include controls for the focal firm's volume of recent commercialization activity, controls for the locations of general medical device clusters, and year and class fixed effects, with standard errors clustered at the product-code level in acknowledgement of potential differences across product type (for example, as a result of differences in innovation behavior or regulatory burden). All tables report marginal effects from logit models, facilitating a more direct interpretation of statistical relationships.[35]

Table 3 presents a full set of controls. As expected, there are statistically significant differences across classes and over time. For example, all else equal, radiology devices are over 61 percentage points more likely to be digital than orthopedic devices (the omitted category), while the time trend indicates that, all else equal, the likelihood of a new device being digital increases by roughly 1.3 percentage points with each passing year. Column 1 uses year fixed effects, while Column 2 includes a time trend. Notably, the coefficients on the class controls are very similar across the two samples. The pseudo-R-squared values are trivially higher in the models using year fixed effects rather than a time trend, so we use the full specification in Column 1 as controls in all subsequent regressions (however, results are stable regardless of the convention chosen for including this set of control variables).

[35] A full set of corresponding linear probability models (excluded due to length and redundancy) lead to the same conclusions as those presented below.

Table 3: Control variables: year and product class

Logit model: digital device commercialization		
	(1)	(2)
Cardiovascular	0.240***	0.239***
	(0.045)	(0.046)
Clinical Chemistry	0.154***	0.156***
	(0.047)	(0.047)
Dental	0.054*	0.054*
	(0.024)	(0.024)
Gastroenterology, Urology	0.126***	0.126***
	(0.024)	(0.024)
General Hospital	0.092**	0.091**
	(0.033)	(0.033)
General, Plastic Surgery	0.097***	0.097***
	(0.023)	(0.023)
Radiology	0.615***	0.616***
	(0.079)	(0.080)
Year Fixed Effects	X	
	.	
Clearance year		0.013***
		(0.001)
N	22,291	22,291
Pseudo-R^2	0.2419	0.2402

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

Logit model results for years 2005-2016, inclusive. Column 1 includes year fixed effects; Column 2 includes a linear time trend. Omitted class = Orthopedic Devices; omitted year (Column 1) = 2005, marginal effects reported. Digital devices defined based on keyword method.

1.4.2 Geographic and Within-firm Know-how

Table 4 presents results predicting digital innovation at the product level, specifically evaluating Hypotheses 1 and 2.

Table 4: Geographic and within-firm expertise

Logit model: digital device commercialization								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ln of digital device experience	0.122***		0.077***					0.072***
	(0.009)		(0.012)					(0.011)
Ln of same-class digital device experience		0.123***	0.056***					0.043**
		(0.012)	(0.016)					(0.013)
Ln of state software employment				0.007			0.010*	0.005
				(0.005)			(0.005)	(0.004)
In digital device cluster (general)					0.080***		0.041***	0.010
					(0.013)		(0.009)	(0.008)
In digital device cluster (class-specific)						0.142***	0.133***	0.104***
						(0.010)	(0.009)	(0.008)
N	22,291	22,291	22,291	22,291	22,291	22,291	22,291	22,291

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

Firm experience and cluster variables are defined based on three most recent calendar years. All models control for volume of firm commercialization activity in past three years and state-level device clusters (representing all state-level medical device device commercialization). All models also include a full set of time and product class fixed effects. Marginal effects reported; standard errors are clustered at the product code level. Digital devices defined based on keyword method.

Columns 1 to 3 consider the role of within-firm experience in predicting digital innovation. Column 1 shows a strong, statistically significant relationship between general digital device experience and the current likelihood of digital innovation. Relative to the sample mean, a one-standard-deviation increase in general firm experience is associated with a 33% increase in the likelihood of commercializing a digital product.[36] Column 2 shows a class-specific relationship between firm experience and the current likelihood of digital innovation. Relative to the sample mean, a one-standard-deviation increase in class-specific firm experience is associated with a nearly 37% increase in the likelihood of commercializing a digital product. Column 3 indicates some attenuation of the effect sizes seen in Columns 1 and 2 for the obvious reason that class-specific experience is a subset of general experience; however, the effects are individually and jointly highly significant and suggest large magnitudes for the relationship between a firm's recent experience in digital product commercialization and its current likelihood of digital innovation.[37]

Columns 4 to 7 consider the role of geography in predicting digital innovation. We find that although a state's share of software engineers (Column 4) is not a statistically significant predictor of digital innovation (although the coefficient is positive, as expected), other measures of geographic expertise are associated with a much higher likelihood of digital innovation. Dummy variable indicators for being in either a general digital device cluster (Column 5) or being in a class-specific digital device cluster (Column 6) are both strongly associated–both individually and jointly–with higher probabilities of digital innovation in new products (Column 7). In the combined model, which suffers least from potential omitted-variable bias, we observe that being in a general digital de-

[36] Natural logarithms of the values presented in Table 2 were used for this and the following calculations.

[37] These findings are also consistent with the experience of a former medical device industry executive who was interviewed about these findings *ex post*. He described a relatively siloed R&D process within the companies he was most familiar with, where knowledge was likely to spill over first within business units. Medical device firms' business units tend to be organized by medical specialty area (e.g., interventional cardiology) and such specialties largely correspond to the classes used here. The class-specific nature of the spillovers documented in Table 4 were described by the executive as "unsurprising."

vice cluster is associated with a 4.1 percentage point increase in the likelihood of digital innovation, while being in a class-specific cluster increases that probability by a further 13.3 percentage points.[38] Further, these magnitudes are large: devices in the sample were roughly 30% digital in 2016, the final year of observation. This implies that being located in a general digital device cluster would (conservatively) increase the baseline likelihood of digital innovation by 13.7% while being in a class-specific cluster would increase it by over 44%.[39]

Column 8 of Table 4 presents an "all-in" model that combines both experience and geography. The magnitudes of the coefficients are similar to those in Columns 3 and 7, with the exception of general digital device clusters becoming an insignificant predictor of digital innovation. Column 8 therefore suggests that after controlling for factors related to firm experience, only location in a class-specific cluster is additionally predictive of digital innovation. The similarity of the remaining coefficients across specifications in Table 4 indicates that within-firm experience and class-specific geographic expertise are mostly independent of one another and have largely orthogonal impacts in these predictive models. The findings in Table 4 support Hypotheses 1a, 1b, and 2b and partially support Hypothesis 2a.

1.4.3 Firm Financial Resources

Table 5 presents results from regressions designed to evaluate Hypotheses 3 and 4.

[38] In this model, the state's share of software engineers is also predictive of digital innovation.

[39] Being in a digital device cluster is associated with a 4.1 percentage point increase in the likelihood of digital innovation off a mean of 30%, for a 13.7% increase. A similar calculation was done to arrive at the 44% figure for being in a class-specific cluster.

Table 5: Financial resources

Logit model: digital device commercialization						
	(1)	(2)	(3)	(4)	(5)	(6)
Publicly listed firm	-0.004			-0.002	-0.003	0.004
	(0.010)			(0.010)	(0.010)	(0.012)
VC-funded firm		0.030*		0.030*		
		(0.012)		(0.013)		
Total VC funding, $ (Ln)			0.012**		0.012**	0.012**
			(0.004)		(0.004)	(0.004)
N	22,291	22,291	22,291	22,291	22,291	22,291

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

All models include a full set of year and product class fixed effects. Column 6 includes additional controls for volume of firm commercialization activity in past three years and state-level device clusters (representing all state-level medical device device commercialization), as in Table 4. Marginal effects reported; standard errors are clustered at the product code level. Digital devices defined based on keyword method.

We first consider whether public firms (Column 1) and VC-funded private firms (Column 2) are more likely to engage in digital innovation. We find that VC funding is on its own a statistically significant predictor of digital innovation, with VC-funded firms roughly 3 percentage points more likely to innovate digitally, all else equal. Column 3 presents results when using the natural logarithm of the cumulative dollar value of VC funding up to the year of commercialization as a predictor. These results indicate that a one-standard-deviation increase in a firm's VC funding is associated with a 4.1 percentage point higher likelihood of digital innovation. [40] Columns 4 and 5 present results from combined regression models that consider public status and VC funding (or funding amounts) simultaneously, finding again that only venture-capital–funding indicators (both as a binary status and as a cumulative funding total) are statistically significant predictors of new digital commercialization. Column 6 presents the same model as Column 5, but uses the full set of firm experience and product cluster controls used in Table 4; results are unchanged.

Based on Table 5 alone, only Hypothesis 4 is broadly supported by the data. However, we note that if either measure of firm financing is correlated with omitted variables, such as geography, the

[40] This calculation uses the natural logarithms of values presented in Table 2.

models in Table 5 will suffer from omitted-variable bias. Therefore, in the next set of regressions, we explore fully specified models.

1.4.4 Further Regression Analysis and Mechanisms

Table 6 presents a set of combined models in which Hypotheses 1 to 4 are evaluated simultaneously, with further *post-hoc* extensions to consider drivers of novel versus follow-on innovation and to consider potential differences between durable versus single-use devices.

Column 1 presents an all-in predictive model using the entire regression sample. Differences between the results presented in Column 1 of Table 6 and those seen in Tables 4 and 5 therefore indicate the size and direction of any omitted-variable bias that may have been unintentionally introduced by assessing individual hypotheses separately. We find that the main difference in this all-in analysis is that being a publicly listed firm is now associated with a higher likelihood of digital innovation, thus providing conditional support for Hypothesis 3.

In these specifications, access to capital is associated with more digital innovation; however, these effects are smaller than those associated with experience and geographic expertise. After controlling for other factors, public firms are more likely to commercialize digital products by roughly four percentage points (Column 1), but these firms are, by definition, already incumbents. Among new entrants, VC continues to be a positive (albeit less powerful) predictor of digital innovation. Back-of-the envelope calculations suggest that a firm would need an exponential increase in its VC funding in order to compensate for the benefit afforded by simply being located in a specialized product cluster. Alternatively, the results imply that a doubling of VC funding would be necessary in order to offset just an 8.5% increase in a firm's recent class-specific experience with other digital products– the equivalent of roughly one additional digital device's worth of recent commercialization experience.

Table 6: Decomposition by device types

Logit model: digital device commercialization						
	Full Device Sample	Novel D.Devices	Follow-On D. Devices	All Non-Chem Devices	Single-Use Devices	Durable Devices
	(1)	(2)	(3)	(4)	(5)	(6)
Ln of digital device experience	0.078***	0.007***	0.075***	0.098***	0.022***	0.146***
	(0.011)	(0.002)	(0.011)	(0.011)	(0.006)	(0.026)
Ln of same-class digital device experience	0.040**	0.002	0.040**	0.021	0.016**	0.022
	(0.012)	(0.002)	(0.013)	(0.012)	(0.005)	(0.025)
In digital device cluster (general)	0.013	0.003	0.012	0.017*	-0.003	0.030
	(0.008)	(0.002)	(0.007)	(0.008)	(0.005)	(0.017)
In digital device cluster (class-specific)	0.104***	0.011***	0.098***	0.097***	0.028***	0.147***
	(0.007)	(0.002)	(0.007)	(0.008)	(0.005)	(0.017)
Publicly listed firm	0.039***	0.011***	0.034***	0.033***	0.007	0.053**
	(0.008)	(0.002)	(0.008)	(0.008)	(0.006)	(0.017)
Total VC funding, $ (Ln)	0.010***	0.002***	0.008**	0.011***	0.008***	0.010
	(0.003)	(0.001)	(0.003)	(0.003)	(0.002)	(0.006)
N	22,291	18,508	22,027	20,973	12,835	8,133

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

Firm experience and cluster variables are defined based on three most recent calendar years. All models control for volume of firm commercialization activity in past three years and state-level device clusters (representing all state-level medical device device commercialization). All models also include a full set of time and product class fixed effects. Marginal effects reported; standard errors are clustered at the product code level. Digital devices defined based on keyword method.

In additional *post-hoc* analysis presented here, we split the sample into novel digital devices (first-of-their-kind) and follow-on digital devices (new versions of already-marketed digital products), as in Columns 2 and 3. We see only one major difference between these two types of digital product commercialization: the importance of class-specific experience in the overall analysis appears to be driven by its role in predicting follow-on innovations. In other words, first-of-their-kind innovations are less reliant on specialized within-firm experience than follow-on innovations, a result that is intuitive and, in part, mechanical. The fact that product-class–specific experience predicts follow-on innovation strongly points to likely within-firm spillovers from experience in this context and is consistent with notions of information friction [163] and asset complementarity [164].

In a final *post-hoc* analysis, we observe that different types of access to capital matter in slightly different ways for considering durable medical devices (those used multiple times and often based in hospitals, such as ultrasound equipment) versus single-use medical devices (those used in only one patient, such as pacemakers or insulin pumps). Columns 5 and 6 indicate that publicly listed firms are over five percentage points more likely to engage in digital innovation of durable devices, whereas VC-funded firms are more likely to pursue digital innovation in single-use devices.[41] These findings are consistent with stronger intertemporal spillovers from digital innovation among established firms. Class-specific experience is also a stronger predictor of single-use versus durable digital devices. This may indicate that class-specific learnings and spillovers are less important for durable medical equipment, where general purpose digital components such as digital monitors or digital data storage are more likely to be relevant.

[41] Column 4 presents results from the same statistical model as used in Column 1, but on a limited sample that excludes "clinical chemistry" devices, which, for technical reasons, could not be easily classified into durable versus single-use products and were therefore excluded from the analysis in Columns 5 and 6.

1.4.5 Robustness Tests and Alternative Specifications

We also undertake a series of exercises as robustness tests of our main findings. The primary analysis presents results using geographic and within-firm expertise variables constructed using three-year look-backs; this was chosen as a result of the natural 2+ year product development cycle for medical devices discussed above. However, we also estimate all models using versions of these measures constructed based on two-year and five-year look-backs. For brevity, we do not report coefficients, although all of our findings remain highly similar using those alternate constructions of experience.

Second, we consider our alternative form of document classification for the identification of digital medical devices in order to verify the results generated using our supervised, keyword-based classification method. As discussed earlier, we use the National Library of Medicine's Medical Text Indexer (MTI) algorithm to identify medical devices whose product statements or summaries are flagged as discussing software. Tables 4, 5, and 6 of Appendix C present MTI-based analogues of our keyword-based results in Tables 4, 5, and 6 in this paper. We find that all of our main results are robust to this different classification method, with magnitudes only somewhat attenuated due to the potentially more conservative nature of the MTI algorithm.

1.5 Discussion and Conclusions

In this study, we describe the digital transformation of the medical device industry and consider how new opportunities for digital product development have been pursued by both new entrants and incumbents. In this setting, we observe all new product commercialization over a 15-year period and document several trends in the digitization of medical technology and their implications for the industry.

We first characterize the growth of digital products over time and across medical specialties, finding important differences. For example, by 2016, there were over twice as many digitized product

types and more than three times as many new digital product approvals for cardiovascular devices than for orthopedic devices. These descriptive findings are novel; to our knowledge, no other studies have comprehensively characterized the digitization of products in this industry. Further, we develop and validate a method for using supervised document classification to analyze the contents of product descriptions. We use multiple methods to collect indicators of the use of software in product descriptions of new medical devices and cross-validate our findings.

We then turn to unpacking the factors that predict which firms drive digital innovation in this setting and find several pieces of evidence that point to significant incumbent advantages. We observe a strong relationship in which both geographic digital product clusters and prior digital product commercialization experience–above and beyond existing *general* clusters and *overall* new product commercialization experience–predict digital innovation. Class-specific firm experience is even more strongly associated with digital innovation, pointing to the importance of product area experience when commercializing new digital devices.

The importance of firm experience is evocative of other studies of the medical device industry, such as[49], that emphasize the importance of regulatory knowledge, marketing knowledge, and understanding of market opportunities in the medical device industry. The results are also similar to those seen in other settings, where the important role of "complementary know-how" in a changing industry has been well documented[94]. More broadly, our findings are consistent with the evidence that acquired know-how has positive spillovers not only within firms, but also across firms in a local labor market, as summarized by[30]. Interestingly, within-firm experience and class-specific geographic expertise are mostly independent of one another and have largely orthogonal impacts in these predictive models. This suggests that being advantageously located can compensate to some extent for lack of within-firm know-how (and vice versa).

We also consider how access to capital may support digital innovation and we find positive–but comparatively small–effects of financial resources on predicting digital innovation. After control-

ling for other factors, public firms are more likely to commercialize digital products, as are those with VC funding, however we find that in order to have a comparable likelihood of commercializing a digital device, a firm would need orders of magnitude more VC funding in order to compensate for the benefits of being in a specialized product cluster. Considering differences between novel versus follow-on innovations, we find that first-of-their-kind digital products appear to be less reliant on within-firm experience than are follow-on innovations, for which there appear to be stronger spillovers from medical-specialty-class–specific activities. Finally, we observe that, in this setting, publicly listed firms are more likely to commercialize durable medical devices, which are used over longer periods of time and are often hospital-based, relative to privately-held and/or VC-funded firms, which are more likely to pursue single-use products. These findings support stronger intertemporal spillovers from digital innovation among established firms.

Taken together, our results suggest that industry incumbents–by multiple definitions–are driving digital innovation in the U.S. medical device industry. We observe within-firm and within-cluster spillovers from past digital innovation into future digital innovation as well as a positive role of large, publicly listed firms (which are, by definition, established players). Venture capital funding does appear to play a small role in supporting digital innovation, but this role is dwarfed by that by other factors. We conclude that in this industry setting, where the costs of entering new product markets are high, digital innovation favors firms with an incumbent advantage.

An important caveat to this study is that we have characterized just one industry. Our setting is advantageous because data on all new product commercialization can be observed and databases are detailed and provide rich product detail. However, it is a setting in which entry barriers shape the relatively high costs associated with entering new product markets. We therefore expect that our findings are most likely to be relevant in other settings–whether regulated industries or not–in which the cost of entering a new product market is non trivial and where supply-side costs are therefore relevant predictors of new market entry behavior.

These findings have important managerial and policy implications. For firms considering digital market entry, our results suggest that in settings with significant entry costs, incumbent firms are likely to play a more significant role in digital product development than new entrants. More specifically, firms with digital product experience are at an advantage relative to firms with only general product experience. However, the geographic concentration of digital new product development in specific product areas points to advantages for *both* new entrants and incumbent firms located in these clusters. As noted above, these effects appear to be largely orthogonal: being advantageously located can to some extent compensate for a lack of digital product experience, and contrariwise. Thus, new entrants may strategically benefit from co-locating with early digital leaders in their industry.

For policy-makers, our findings suggest the importance of prior experience when undertaking digital product commercialization. To the extent that policy-makers want to support new entrants, clear guidance on best practices for developing digital products may serve as a substitute for prior experience. As regulators increasingly devote attention to clarifying expectations for digital devices– for example, through the FDA's new "digital health software pre-certification" program,[42] which allows a small group of technology leaders to commercialize new software products more efficiently– it will be important to keep such considerations in mind.

[42] https://www.fda.gov/MedicalDevices/DigitalHealth/DigitalHealthPreCertProgram/default.htm

2

Work-from-anywhere: The Productivity Effects of Geographic Flexibility

2.1 Introduction

Human capital has been documented as a critical source of firm competitive advantage (Campbell et al. [46]; Coff & Kryscynski [52]; Ganco et al. [74]; Starr et al. [155]). Within this literature, a growing body of work documents the role of nonpecuniary incentives in shaping motivation of workers (Campbell et al. [46]; Carnahan et al. [47]; Kryscynski [105]; Sauermann & Cohen [148]). However, from the perspective of the firm, it is also critical to study the productivity effects of provisioning such incentives. An example of a nonpecuniary incentive that has attracted much debate is remote work, in which an employee is allowed to work outside of the office, either part or full time. Despite a few high-profile retreats from remote work by companies like Yahoo! and IBM (Simons [152]; Swisher [159]), many organizations such as Amazon, Apple, American Express, and Glassdoor offer remote work programs to employees (Team [160]). Demand for remote work and other flexible work arrangements is increasing (Gallup [10]) and the value that employees place on remote work arrangements is driven in large part by the costs of commuting, childcare, and eldercare faced by a population increasingly comprised of dual-career families (CEA [4]).

To date, research on remote work, including the recent Bloom, Liang, Roberts, & Ying (2015) study (Bloom et al. [36]), has focused largely on the effects of working from home (WFH), in which

the employee may conceivably still be living within commuting distance of the office. In recent years, however, another form of remote work—work-from-anywhere (WFA)—has begun to emerge. Here, workers are no longer required to live in the same geographic location as the firm and have greater flexibility to choose where to live. Organizations with WFA policies include Gitlab, Akamai, Github, NASA, and DataStax, among others (Fatherly [67]; Moore [126]; NASA [11]; Reynolds [141]). However, to the best of our knowledge, there is no research on the productivity effects of WFA policies. One of the comprehensive literature reviews on telecommuting states (emphasis added), "Home was the primary location for telecommuting in nearly all the studies included in this meta-analysis" (Gajendran & Harrison [72]). In other words, the prior telecommuting and remote work literature focused on studying the effects of moving the worker from one workspace (within the firm's office), to an alternative workspace (within the home of the worker, typically in the same geographic location as the firm office). In contrast, WFA gives the worker the additional choice of deciding to move to a geography of their choice.

We argue that WFA is fundamentally different from WFH in how it might affect worker productivity. Previous research on WFH has identified benefits to employee performance via mechanisms such as reduced commute times and fewer sick days (Bloom et al. [36]), which can be attributed to increased temporal flexibility (Evans et al. [66]). WFH also allows workers to control ambient workspace elements such as clothing, layout, music, ventilation, etc. (Gajendran & Harrison [72]). WFA goes further by eliminating the traditional link between the geography of home and work location, resulting in geographic flexibility, in which a worker can remain employed at a firm without needing to live within commuting distance of the firm's office location. In the case of WFA, employers cede to workers control of the geography in which they choose to live, in addition to ceding the temporal control afforded by WFH (Evans et al. [66]). In this sense, WFA is perhaps better titled "Live Anywhere," as the benefits that distinguish it from WFH derive from employees' ability to choose where they live. Contingent on the firm not adjusting wages downwards for workers transitioning into a

WFA program and the worker relocating to a lower cost-of-living location, real income might rise. This addresses concerns about the rising cost of living for knowledge workers (Moretti [127]). In addition, geographic flexibility might enable the knowledge worker to move to a location affording the worker greater psychic benefits such as better climate or proximity to family (Greenwood [88]).

These differences between WFA and WFH, along with the general increase in both worker demand for and employer provision of WFA policies, lead us to our main research question: How does the geographic flexibility provided by WFA affect individual worker productivity? Bloom et al. (2015)'s [36] research in a Chinese travel agency shows positive causal productivity effects of moving a worker from an in-office setting to a WFH regime. We ask whether there are causal productivity effects of moving a worker from a WFH regime to a WFA regime. We also attempt to shed light on the mechanisms underlying the effects of WFA on knowledge worker productivity.

Our setting, the United States Patent and Trademark Office (USPTO), and in particular, the job of patent examiner, is in many ways the ideal setting for our research questions. First, our setting allows us to exploit a natural experiment related to the implementation of a WFA policy. The bureaucratic processes governing the implementation of WFA at the USPTO allow us to mitigate endogeneity concerns related to worker selection into the WFA policy. More specifically, the implementation of WFA was driven by negotiations between USPTO managers and the union of patent examiners, leading to a monthly enrollment quota that created exogeneity in the timing of individual examiners' transition to WFA. Second, the role of a patent examiner is relatively independent. Third, examiners in our sample had spent at least two years in the USPTO office and additional time in a traditional WFH program before taking on a WFA assignment. These conditions help us in three ways. First, the independent nature of the task performed by patent examiners and the mandate to spend two years in the office help us (at least partially) control for adverse effects of remote work (e.g., effects of additional coordination costs and reduced learning effects from colocated peers) that might lead to confounding concerns in a more general setting. Second, given that all

WFA employees in our study first transition from being an "in-office worker" to a "WFH worker" before further transitioning into a "WFA worker", we are able to isolate a productivity effect of geographic flexibility awarded by WFA vis-à-vis WFH. Third, the exogenous timing of transitioning from WFH to WFA enables us to estimate a causal comparison of productivity under the two regimes. These conditions not only present a clean empirical setting, but also serve as important boundary conditions to our findings and suggest a future research agenda.

To preview, we exploit this bureaucratic policy-induced variation and employ examiner fixed effects, finding that examiners enjoy an increase in work output of 4.4 percent when in the WFA program compared to the baseline of when the worker was in the WFH program, with no significant increase in the amount of rework. It is important to point out that to the best of our knowledge, with the exception of the Bloom et al. (2015)[36] study, there are no other studies in the remote work/teleworking literature that document causal productivity results. Furthermore, while Bloom et al. (2015)[36] documents causal results related to WFH, this study documents causal results related to transitioning from WFH to WFA. Our secondary analysis compares WFH productivity to in-office productivity, finding an increase in productivity similar to that identified in Bloom et al. (2015)[36]. These two analyses give a sense of the stepwise progression that can take place (at least in some organizational contexts) as employees move from in-office, to WFH, and ultimately to WFA work.

We also exploit institutional details of our setting to isolate WFA mechanisms. First, our analysis shows a correlational relationship between examiners relocating to below-median cost-of-living locations and greater productivity increases, suggesting that one of the motivating benefits of WFA for workers could be an increase in real income. Second, we present correlational results showing that workers with greater tenure are more likely to choose Florida, arguably a "retirement-friendly" destination where they might derive additional psychic benefits. Crucially, we also show that the productivity effects of WFA cannot be explained by alternative mechanisms such as reduced commute

time and reduced monitoring. We also document illustrative evidence of how geographic flexibility affects productivity using field interviews.

In supplementary analyses, we also examine work practices associated with enhanced productivity for the sample of WFA examiners. In particular, we exploit a bureaucratic mandate requiring WFA employees to utilize informational technology tools for online coordination and find that mandating this practice does result in an increase in productivity for WFA employees whose work needs to be certified by a supervisor. Finally, we test whether measures related to examiner effort and leniency change when an employee transitions to WFA: We find no evidence of increased leniency or reduced effort as measured by examiner-added citations.

Our findings make an original contribution to the literature on remote work. While prior literature has documented robust productivity effects of working from home (Bloom et al. [36]), our study goes a step beyond in documenting the productivity effects of working from anywhere compared to WFH, and demonstrates that the benefits of WFA derive from additional mechanisms not triggered by WFH policies alone. Our results also contribute to the literature on the effects of nonpecuniary incentives (Kryscynski [105]; Sauermann & Cohen [148]; Stern [157]), demonstrating that the application of a WFA policy can provide employees both direct economic value, in the form of reduced cost of living, and intrinsic value, in the form of increased psychic benefits from geographic flexibility, while also increasing firm productivity and profits.

2.2 Work-from-anywhere and geographic flexibility

In this paper, we examine the productivity effects of moving workers from a traditional WFH regime to a WFA regime that grants workers geographic flexibility, i.e. the flexibility to choose a geographic location in which to live, which we theorize as differing from both the temporal flexibility and flexibility to design the workspace granted by a WFH policy. In this section, we situate

WFA within the larger body of research on nonpecuniary incentives and identify mechanisms that we argue distinguish WFA from WFH.

Nonpecuniary incentives such as authority, independence, and company reputation have been linked to employees' decisions to stay or leave a firm (Agarwal & Ohyama [19]; Cable & Turban [45]; Fehr et al. [69]; Sauermann & Stephan [149]). Kryscynski (2011) [105] argues that incentives encompass any aspects of the employment relationship valued by workers, regardless of whether those aspects are directly or indirectly bestowed, created, or tied to individual membership, effort, or performance. Certain non-pecuniary incentives provided by firms to workers—such as challenge and autonomy (Sauermann & Cohen [148]), information on quality of work (Kolstad [100]), and tolerance of early failure (Azoulay et al. [29]; Ederer & Manso [64])—have been shown to influence worker productivity in ways that are incremental to the effects of pecuniary incentives. Sauermann & Cohen (2010) [148] argue that pecuniary incentives are designed to appeal to employees' extrinsic motivations, while non-pecuniary incentives appeal to employees' intrinsic motivations by enabling them to gain greater satisfaction from the work itself.

Remote work programs such as WFH and WFA are non-pecuniary incentives designed to offer the worker flexibility. Evans et al. (2004) [66] define flexibility in the employment relationship as "ceding control to workers over the circumstances of their work by enabling them to vary those circumstances to address personal and family needs and uncertainties". WFH policies are an increasingly common means of granting temporal flexibility to employees, among other practices that give employees more control over the hours in which they complete their work (Briscoe [42]). The benefits of granting temporal flexibility (such as flexible work hours) to employees have been well-documented in the research on family-friendly work policies and WFH policies, with particular emphasis on improvements in work-life balance and reductions in work-family conflict. Bloom et al., (2015) [36] found evidence that WFH led to a 13 percent performance increase (compared to working from office), of which 9 percent was due to fewer breaks and sick days, and 4 percent was due

to a "quieter and more convenient" work environment (p.165). Work-life balance is generally seen to improve when employees are able to work from home (Gajendran & Harrison [72]), though some negative impacts have been noted in the areas of work-life boundary maintenance (Kossek et al. [103]) and family-to-work conflict (Golden et al. [80]). Improved work-life balance can increase the intrinsic motivation of employees (particularly employees whose identity and motives are tied to family), helping maintain a balance between work and personal life (Akerlof & Kranton [21]; Sauermann & Cohen [148]). Research on remote work has previously linked temporal flexibility to increased perceptions of job autonomy (Bélanger [34]; Gajendran & Harrison [72]), and autonomy has been related to increased motivation on the job (Richard & Oldham [142]; Ryan & Deci [145]).

A WFA policy affords employees all of the benefits of a WFH policy, but additionally awards the worker geographic flexibility. We argue that there are two significant additional advantages of WFA to workers that are not covered by the provision of WFH. The first of these is the ability to live in a location with a lower cost of living than the location of the employer, enabling the employee to enjoy higher real income without an increase in nominal salary. Moretti (2013)moretti2013real deflates nominal wages using a location-specific cost-of-living index and finds that the difference between the wage of college graduates and high school graduates in the U.S. is lower in real terms than in nominal terms; at least 22 percent of the documented increase in college premium is accounted for by spatial differences in the cost of living. Given the opportunity to relocate anywhere in the U.S., knowledge workers might choose to move to a lower cost-of-living region, thus allowing the worker to enjoy greater real income, holding nominal income constant—especially when the employer is based in an urban area with a relatively high cost of living. There is also a literature summarized by Leana & Meuris (2015) [107] that documents how income is related to worker satisfaction.

The second benefit is that if allowed geographic flexibility, employees may also choose to move to a location that awards the worker "psychic benefits," such as living in a location with a more attractive climate (Greenwood [88]). There is also a nascent literature that looks at revealed preferences of

scientists, engineers, and entrepreneurs to choose work that is close to home (Dahl & Sorenson [57], Dahl & Sorenson [58]). Using panel data on the Danish population, Dahl & Sorenson (2010a) [57] estimate a strong revealed preference of scientists and engineers to live close to family and friends. In another paper, they note that "one commonly cited reason for why people do not move more often is that they value being near family and friends, or at least the more frequent and more extended interactions that propinquity allows" (Dahl & Sorenson, 2010b) [58]. A related concept in economics is the construct of psychic costs of migration (Sjaastad [153]; Schwartz [150]). These studies suggest that the provision of geographic flexibility should benefit employees in ways that are incremental to the benefits of temporal flexibility. While temporal flexibility allows employees to spend more time with immediate family, geographic flexibility enables employees to relocate to a location that has lower cost of living and/or where the worker experiences psychic benefits. In addition to more attractive weather and being closer to family and friends, a geographic location of choice could also offer other psychic benefits to a worker, such as pursuit of a personal interest outside of work (e.g., skiing in Colorado). As another example, early-career workers wishing to raise families can move to a family-friendly locale. However, from the perspective of the firm, provisioning policies such as WFH and WFA could lead to additional costs, including increased coordination costs. The organization of workers into a firm has been viewed as a system to coordinate effort and communicate knowledge across multiple intrafirm actors (Grant [86]; Srikanth & Puranam [154]; Thompson [165]). Altering the spatial distribution of employees changes the means of coordination, limiting the ability of workers to rely on tacit coordination mechanisms (Srikanth & Puranam [154]), and potentially leading to increased coordination costs via difficulties in knowledge sharing (Cramton [56]; Gibson & Gibbs [78]). Second, social and professional isolation is a well-documented challenge in the research on remote workers (Cooper & Kurland [55]; Golden et al. [79]). Managers and organizations can help mitigate these challenges through the provision of structures that facilitate social interaction among remote employees, and ground rules for the use of information technology tools, to facilitate communica-

tion (Makarius & Larson [116]).

In summary, prior research on incentives has argued that employers should design incentives to best attract an ideal employee; WFA is an example of such an incentive that offers workers both geographic and temporal flexibility. However, given the constraints of coordination and isolation, an important question remains from the perspective of firms: whether the provisioning of such an incentive creates value for the firm. The hedonic wage analysis literature predicts a "negative trade-off between wages and 'positive' job attributes, attributes like status or flexibility in hours of work" (Lazear & Shaw [106]). Indeed, empirical research demonstrates at least some willingness to exchange wages for non-monetary benefits (Stern [157]). Mas & Pallais (2017) [122] find that on average, workers are willing to accept 8 percent lower wages in exchange for a remote work option, suggesting that remote work policies are perceived as a valuable non-pecuniary benefit by employees. However, in some cases (such as the USPTO), the firm does not decrease wages for employees choosing a WFA regime. As stated earlier, this raises an interesting question for scholars of strategic human capital, economists, and practitioners alike: holding wages equal, when workers are moved from a WFH regime to a WFA regime, does the additional geographic flexibility provided lead to higher productivity?

2.3 Exploratory Fieldwork and Research Context

Because of the nascent stage of WFA research (Edmondson & McManus [65]), we undertook some exploratory qualitative, inductive work to better understand the research context, and to identify potential mechanisms underlying the productivity effect of switching to a WFA regime. This exploratory work included 26 interviews with various USPTO managers, patent examiners, and the Patent Office Professional Association (POPA; labor union) leaders. We also gathered online reviews posted by current and former patent examiners on the review site Glassdoor.com.

The USPTO is the federal government agency authorized to evaluate patent and trademark applications. It is headquartered in Alexandria, VA, and employs about 13,000 people, including slightly more than 8,000 patent examiners. Patent examination comprises roughly 90 percent of the USPTO's work; in 2015, the USPTO received 629,647 patent applications and granted 325,979 patents across many industries and technologies (Choudhury et al. [51]).

A patent application specifies a set of "claims" that defines the invention the applicant wishes to protect. Applications are assigned to examiners based on the required area of technical expertise (software, chemicals, mechanical, etc.). Examiners are organized into nine "technology centers," each made up of smaller "art units." Within a given art unit, a supervisory patent examiner (SPE) assigns each new patent application to a patent examiner (Lemley & Sampat [110]). The examiner is then responsible for reviewing the claims and moving the application through the examination process, with minimal supervisory oversight. Examiners must determine whether patent claims in applications meet the criteria of "novelty" and "nonobviousness" in order to be patentable. In order to determine the validity of claims in an application, the patent examiner uses several proprietary search tools to review the body of publicly available work (called "prior art," it includes existing patents, published patent applications, academic and trade journal articles, and other publications). In order to determine "novelty," the examiner must determine that the claims within the application are not already wholly addressed by another single patent or published work.

Once the examiner has (to her knowledge) exhausted the existing prior art, she issues a "first office action," (FOA) which can be an "allowance," accepting all claims as patentable or, more commonly, a "nonfinal rejection," which indicates that some or all claims are not patentable, and gives the basis for rejection. Applicants can respond by withdrawing, narrowing, clarifying, or providing further evidence to support the rejected claim. The examiner then reviews the response, accepts additional claims as applicable, and issues another office action. This process continues until the examiner believes that no further response will change the outcome of an application, at which point she

issues a "final action." Upon receiving a final action, the applicant has the choice of abandoning all remaining rejected claims, appealing the action to a board of appeal, or restarting the application process by paying an additional $1,200 fee to initiate a "request for continued examination" (RCE). The RCE restarts the entire examination process, but is carried out by the same examiner and takes into account all prior communication. There is no limit on the number of RCEs an applicant may file, and approximately one-third of all applicants file at least one RCE, though few file more than three.

Patent examiners are typically highly educated, holding undergraduate degrees in science and engineering, and some holding advanced degrees in technical fields. New employees are hired at a specific grade level (in line with hiring and employee ranking procedures at all federal agencies) based on their experience and skills. At the USPTO, examiners are hired at the civil servant "grade levels" GS-5, GS-7, GS-9, GS-11, GS-12, GS-13, GS-14, or GS-15, with pay and responsibilities increasing with each grade. During labor negotiations, examiners are represented by the USPTO's union, POPA. Examiners advance up to GS-13 automatically, based largely on tenure. Upon reaching GS-13, an examiner can enter into a signatory review program in which the examiner's work is evaluated. Upon passing this review, the examiner is designated a partial signatory (PS) and can sign nonfinal office actions. After six months of PS status, examiners are eligible for a second-round work review. Upon passing this review, the examiner attains full signatory (FS) status, indicating that the examiner can sign all decisions (including FOAs and final actions).

The USPTO measures examiner productivity using the number of actions completed by an examiner within a given period of time in relation to an expected productivity level, which is determined based on examiner grade level (a proxy for experience) and examiner-specific case mix—examiners in more nuanced or complex fields are granted more time to examine a given application. Following the USPTO's measures, we take the number of actions in a given period as the measure of examiner output. We consider the number of RCEs in a given period to serve as a measure of

rework. While we recognize that this is an imperfect measure (an inventor is well within rights to doggedly pursue a patent claim via an unlimited number of RCEs, regardless of the accuracy and quality of the examiner's ruling), an RCE mechanically leads to rework, as the examiner must search prior art again to write the next decision. Furthermore, our field interviews further support the assumption that, on balance, a greater number of RCEs for a given examiner is likely to indicate a greater need for rework.

The process of patent examination is largely an individual exercise, but with some supervisory constraints. At lower grade levels, patent examiners are typically newer and less experienced in their fields and, therefore, must obtain approval on their actions from either their assigned SPE or a senior patent examiner. However, given the independent nature of the task, even for junior examiners, there is relatively little coordination that needs to be managed between the examiner and his or her supervisor (Choudhury et al. [51]; Lemley & Sampat [110]).

To further enrich our understanding of the examiners' perspective on their jobs, we gathered qualitative data from 258 online employee reviews at Glassdoor.com. The reviews contained a number of consistent themes. Temporal and geographic flexibility are both cited as highly valuable aspects of the examiner job that enable a desirable level of work-life balance.

Furthermore, there are frequent mentions of the independent nature of the job, giving further confirmation that our research context is one of pooled interdependence (Thompson, 1967). The job is also described as highly routine and repetitive, suggesting that routineness is a further scope condition of our findings. There is extensive discussion of the emphasis placed on meeting performance targets for actions. This theme further supports our use of the number of total actions as a valid measure of employee productivity in this context.

2.3.1 Remote work programs at the USPTO

We will focus on two prominent telework programs at the USPTO: WFA (i.e., the Telework Enhancement Act Pilot Program or TEAPP program) and WFH (i.e., the Patents Hoteling Program or PHP program). The USPTO introduced the voluntary PHP in January 2006 with an initial cohort of 500 patent examiners. PHP provides eligible employees with equipment and remote access to systems and allows them to WFH up to four days per week. When they report to the office, they reserve desk space through an online system. PHP is a classic WFH program that offers temporal flexibility, but less geographic flexibility than WFA. In addition, participants must have worked at the USPTO for at least two years and demonstrated "satisfactory performance." Eventually, the PHP program grew to include two subprograms: (1) the "PHP within 50 miles" program (i.e., those examiners who lived within the 50-mile radius of the USPTO headquarters in Alexandria and reported to the USPTO headquarters at least once per week); and (2) the "PHP greater than 50 miles" program (i.e., those examiners who lived at least 50 miles from headquarters but were still required to report to the USPTO headquarters at least once a week).

In December 2010, President Barack Obama signed the Telework Enhancement Act, which set standard rules and regulations for remote work at federal government agencies. In early 2011, the USPTO began planning to pilot a WFA program (i.e., TEAPP), allowing employees to work anywhere in the contiguous U.S. (greater than 50 miles from the USPTO) and travel to headquarters periodically at their own expense, thus awarding eligible patent examiners geographic flexibility. Importantly for our purposes, the USPTO did not adjust wages for employees opting to participate in either the WFH or WFA programs; this helps us test the net impact on firm productivity of the WFA benefit in the absence of any offsetting reduction in wages.

Employees were eligible to participate in WFA if they: (1) were already enrolled in the "PHP > 50 miles" program; (2) had access to the Internet and USPTO systems; (3) agreed to change their "duty

station" (i.e., home office) to a location greater than 50 miles from USTPO headquarters; and (4) waived their rights to travel reimbursement for required trips back to headquarters. The USPTO capped the number of trips that teleworking employees would need to make to headquarters at 12 days and/or five trips during a fiscal year. The USPTO also provides WFA workers with online communication tools such as Microsoft Lync, WebEx webinar services, and Cisco Voice over Internet Protocol (VoIP). On January 30, 2012, the USPTO officially launched the WFA program. On June 24, 2013, the USPTO (in negotiation with POPA) amended the WFA agreement to include the following: "the above tools (instant messaging, document/desktop sharing, virtual meeting, video communication, and a presence indicator) would be mandatory for...full-time teleworkers," noting that "the purpose of requiring the use of these tools was to encourage collaboration" (Chu, Bergrud, Lavigna, McGrath, Reeder, 2015, p. 70). Employees who had been located in the Alexandria headquarters for at least two years were eligible for the "PHP > 50 miles" program.

2.4 Hypotheses: WFA and Productivity

We first examine the productivity effects of moving a worker from a WFH regime to a WFA regime. As discussed earlier, in the former regime, the worker experiences temporal flexibility and control of ambient workspace elements. In the latter regime, the worker additionally experiences geographic flexibility, which might enable the worker to move to a lower cost-of-living location and raise real income. Geographic flexibility might also help the worker relocate to a location which affords the worker higher psychic benefits, such as being in a location with better climate (Greenwood[88]. Given this, we hypothesize that remote work offering both geographic and temporal flexibility (i.e., WFA) has greater productivity benefits compared to remote work offering temporal flexibility alone (i.e., WFH).

Generally, the provisioning of WFA could lead to increased coordination costs, isolation, and

fewer opportunities to learn from colocated peers. However, we argue that in our setting neither of these things are likely to affect rework, especially given that patent examiners were allowed to self-select into a WFA program only when they had spent at least two years at the USPTO headquarters and had achieved a baseline level of task-specific human capital. As prior research (Argote & Miron-Spektor [23]; Katila & Ahuja [97]; Rosenkopf & McGrath [144]) has shown, learning by doing is accrued through the experience of performing a task repeatedly.

Given their preexisting tenure of at least two years at the USPTO headquarters and the routine nature of patent examination, experienced patent examiners are likely to have already developed the requisite absorptive capacity (Cohen [53]) and task-specific human capital (Gibbons & Waldman [77]) to perform tasks such as a prior art search. Second, as described earlier, patent examiners carry out their tasks (researching, searching for prior art, writing decisions, and communicating with applicants) independently, and there are relatively few requirements to coordinate with peers. In this pooled-interdependence setting, patent examiners reach out to peers mainly to seek advice on relevant prior art. Experienced examiners could continue to leverage their intraorganizational social ties even after migrating to a WFA program to mitigate isolation, and our field interviews yielded examples of experienced WFA examiners calling peers (based in Alexandria or elsewhere) and sharing computer screens on the videoconferencing calling tool WebEx to ask: (1) "Do you have a search for me?" (that is, have you searched this topic previously and, if so, could you share the results?); or (2) "Can you take a look at my drawings and suggest prior art?" In summary, experienced examiners have already developed firm- and task-specific human capital and have the technological means to reach out and seek advice from prior colocated colleagues.

Given this, we anticipate that, on balance, even as the amount of output increases, the amount of rework will not increase after such employees move to a WFA regime. We hypothesize: Hypothesis 1a. For workers in a pooled interdependence (low coordination) setting with a baseline level of task-specific human capital, moving the worker from a WFH to a WFA regime leads to an increase in

output.

Hypothesis 1b For workers in a pooled interdependence (low coordination) setting with a baseline level of task-specific human capital, moving the worker from a WFH to a WFA regime does not lead to an increase in rework. As our previous hypotheses focus on the progression from WFH to WFA, it is important that we also document the relationship between these workers' WFH productivity and their previous in-office productivity in order to eliminate the possibility that productivity had previously declined when the worker was moved from an "in-office" regime to WFH. We expect that the benefits identified in Bloom et al. (2015)[36] would be mirrored in our context with an implementation of WFA resulting in increased work output compared to when the worker was in-office:

Hypothesis 2 For workers in a pooled interdependence (low coordination) setting with a baseline level of task-specific human capital, WFH is associated with greater productivity than working in the office.

2.5 Data

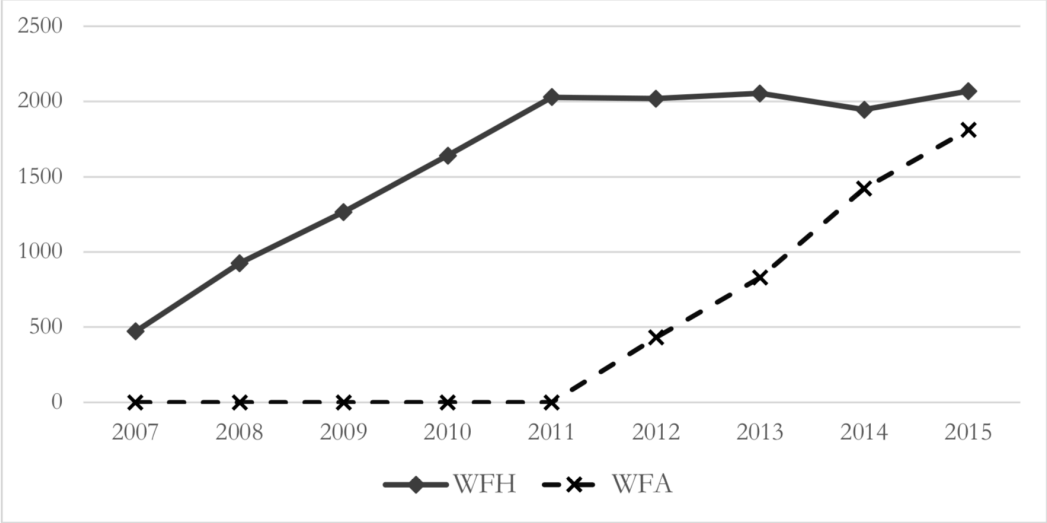
This paper draws on multiple sources of data. We begin with a unique administrative dataset obtained from the USPTO for the years 2007–2015 that reports, annually, all patent examiners on the USPTO payroll, their general schedule (GS) pay level, and a benchmark measure of productivity used for promotion decisions (as a function of the "United States Patent Classification" or USPC class of their examined patents). We link this data to a separate administrative dataset, again obtained from the USPTO, that identifies which examiners are enrolled in each remote work program, their current home office location, and when they began remote work. From here, we link the combined examiner datasets to publicly available USPTO data on applications and transactions (such as RCEs) to quantify examiner-level output and rework.

2.5.1 Examiner Personnel Data

The first dataset used for this study is an annual record of all patent examiners active at the USPTO from 2007–2015, with 9,210 unique examiners over these eight years, inclusive. This data also provides the GS of every USPTO examiner, data that is otherwise not public. As described earlier, the GS of an examiner is of particular importance: it serves as a natural hierarchy for promotions, it is mechanically correlated with tenure and experience, and higher-grade examiners have increasing levels of autonomy in their workflows. Hence, controlling for GS is important to account for unobservable task-specific human capital of examiners (Gibbons Waldman, 2004).

We also utilize a second unique, USPTO-provided, personnel dataset specifically focused on remote workers. This dataset includes examiner identifiers, as well as the remote work program(s) in which the examiner enrolled: WFA and PHP (<50 miles and >50 miles combined). Figure 1 shows the growth in remote working across the three programs from 2007–2015; WFA appears to gain an increasing share of the teleworking population as examiners substitute away from PHP programs. The examiner-specific start date for each specific telework program is also available to us, allowing us to track an examiner across programs. This data also identifies the city and state of a teleworking examiner (as of August 2016), which is important for spatial analyses (to be described later).

FIGURE 1 Growth in Number of Remote Workers at the USPTO



This figure illustrates the annual number of examiners enrolled in two remote work programs at the USPTO: WFA (TEAPP) and PHP.

2.5.2 USPTO Patent Data

Data on patents and patent application-level transactions were collected from a combination of two publicly available datasets: USPTO's Public Patent Application Information Retrieval (PAIR) dataset and PatentsView. Application data collected includes the name of the examiner assigned to a patent, the examiner's art unit, and the USPC classification of the application. For each patent, we then collected data on all transactions executed by an examiner, focusing on two specific metrics of productivity: total actions (measure of output) and RCEs.5 Total actions is a measure of aggregate output delivered by an examiner, and aligns with the PTO's internal performance measure of expectancies. The second measure, RCEs, are a measure of rework.

2.5.3 Spatial Data

City and state data on the most recent location of teleworking patent examiners was obtained through the USPTO administrative dataset on teleworkers. This data was then geocoded using commercially available GIS tools, and measures of the spatial concentration of WFA examiners were calculated.

2.5.4 Identification Strategy: Natural Experiment

To provide robust econometric estimates related to how the implementation of WFA affected output and rework, we exploit a natural experiment within the USPTO. As noted earlier, the implementation of WFA was driven by negotiations between the USPTO management and POPA. Specifically, these negotiations resulted in a monthly quota for eligible examiners transitioning to WFA in the first 24 months of program implementation. Our field interviews indicated that the monthly quotas were oversubscribed, and eligible examiners often had to wait for several months to [1] transition into the WFA program. While it is likely that observable and unobservable factors determine whether or not examiners transition into WFA, we circumvent these concerns by focusing on the sample of examiners who selected to transition into the WFA program over the first 24 months

[1] We assume here that shirking—another possible negative outcome associated with increased autonomy—is reflected in the productivity measure, given that we are using an objective measure of productivity. Concerns about shirking were addressed at the USPTO in a contemporaneous time frame, with claims of "examiner fraud" and "attendance abuse" made by The Washington Post (Rein [139]), based on critical findings from the U.S. Department of Commerce's Office of the Inspector General. However, all of these findings related to either (1) overreporting of hours worked or (2) shifts in the timing of work completed, such as backloading at the end of a calendar quarter, which raised concerns about the accuracy and quality of work completed. USPTO Office Director Michelle K. Lee told lawmakers that she and her team "do not tolerate any kind of attendance abuse" (Rein [139]). Our measure of productivity is only output dependent, so overreporting of time worked would not affect this measure. Second, our measure of rework—while not a perfect proxy for quality—should capture any substantial degradation in work quality due to backloading or other timing shifts. In robustness checks (available upon request), we also employ month fixed effects to test our causal results, and results remain robust.

and exploit variation in when (i.e., which month) the examiner could transition into WFA, variation that is exogenous given the monthly quotas determined by the USPTO management and POPA. Below we provide further details of how the implementation of the WFA program lends itself to a natural experiment.

As a result of the negotiations conducted between the USPTO management and POPA, the USPTO planned to enroll participants in the WFA program in phases. Additionally, and importantly for the purpose of identification, there was an exogenous quota imposed for eligible examiners enrolling in the WFA program. The number of slots was decided by a committee comprising management and union members. If a slot was not available, the prospective enrollee was placed on a waiting list. Our field interviews indicated that all slots allocated for the first several months were exhausted, implying that even if an examiner was eligible for WFA, he or she would have had to wait an unknown length of time before transitioning to WFA. As such, the timing of an eligible examiner's transition to WFA was relatively exogenous. Our field interviews indicated that prior tenure, experience, or performance were not considered in allocating slots to eligible examiners.

To validate our natural experiment and the insights generated by the field interviews, we test whether the variation in WFA transition time was truly exogenous by regressing the time it took an eligible examiner to transition to WFA on observable measures of past performance. As our main results analyze productivity (and include a measure of the expected work as a control), we regress "months to WFA" on measures of total examiner-level output, rework, and expectancy (a measure of expected output in the previous year). Results from these analyses are reported later in the paper. We find no evidence of selection on prior performance (or other observables), validating our principal identification strategy.

2.6 Estimation and Results

We focus on utilizing the natural experiment and limit our sample to examiners who enrolled in WFA in either 2012 or 2013. Within this sample, we exploit bureaucratic process-induced variation in enrollment dates to identify the effect of receiving WFA earlier than another examiner. As both examiners in this exercise must be eligible and have selected into the program, we avoid the traditional identification issues that arise from self-selection—all examiners in our sample can be thought of as treated, varying only in the amount of time they have had to wait to be exposed to the treatment (WFA). Moving forward, we refer to this sample as the "WFA sample." The WFA sample comprises 831 examiners (out of the 9,210 examiners). Table 1 reports summary statistics for the full sample.

TABLE 1 Descriptive Statistics and Correlation Matrix: Full Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1) Total Action	1.000	0.431	0.589	0.734	0.457	0.093	0.118	0.016	-0.242	0.023	0.032	0.291	0.261
(2) Total RCE		1.000	-0.074	0.260	0.239	0.098	-0.028	0.019	-0.006	-0.061	-0.010	0.082	-0.004
(3) FOA			1.000	0.433	0.177	-0.044	0.129	0.008	-0.247	0.078	0.040	0.168	0.246
(4) Reject				1.000	0.722	0.159	0.016	-0.022	-0.165	-0.055	0.016	0.111	0.064
(5) Examiner Cites					1.000	0.127	-0.097	0.001	-0.084	-0.123	-0.009	-0.076	-0.129
(6) WFA (TEAPP)						1.000	-0.325	-0.164	0.025	-0.456	0.029	0.143	0.053
(7) PHP (<50 miles)							1.000	-0.288	-0.031	0.642	0.006	0.383	0.357
(8) PHP (>50 miles)								1.000	0.014	-0.402	0.027	0.084	-0.024
(9) Mean Expectancy									1.000	-0.034	-0.074	0.027	-0.005
(10) Nearby Examiners										1.000	-0.038	0.183	0.220
(11) Distant Examiners											1.000	0.026	0.031
(12) GS												1.000	0.690
(13) Primary Examiner													1.000
Mean	11.375	1.347	4.306	6.500	15.238	0.076	0.198	0.070	23.183	1258.06	149.586	12.552	0.403
Std Dev	7.039	1.581	3.931	4.329	13.835	0.264	0.399	0.255	5.143	1131.38	362.223	1.919	0.490
Min	1	0	0	0	0	0	0	0	6.6	0	0	0	0
Max	167	18	160	58	208	1	1	1	31.6	2399	2263	15	1
n (non-missing)	576,274	576,274	576,274	477,305	477,305	576,274	576,274	576,274	576,274	322,155	322,155	576,274	576,274

Notes. Observations are at the examiner-month level. The full sample uses all examiners in our dataset, regardless of their remote work status. The "WFA causal sample" (reported in Appendix 4) refers to examiners who transitioned to WFA in 2012 or 2013.

2.6.1 Causal Estimation: WFA vs WFH

Hypotheses 1a and 1b state that for workers in a pooled-interdependence (low-coordination) setting with a baseline level of task-specific human capital, moving the worker from a WFH to a WFA regime leads to an increase in output but does not an increase in rework. We utilize the natural experiment described above, employing the following examiner month-level specification to test these hypotheses:

$$Output_{it} = \alpha_i + \beta_{it} * WFA + \xi_i t + \gamma_t + \lambda_i + \varepsilon_i$$

where WFA is a binary indicator that turns on (and stays on) when an examiner enrolls in WFA during the 2012–2013 timeframe. As described earlier, we use two different measures of individual-level output: for individual output using total actions and for individual rework using the number of RCEs. ξ_{it} is a vector of controls that includes examiner month-specific grade level and examiner month-specific expectancy, while γ_t is a full set of time (month) fixed effects and λ_i is an optional set of examiner fixed effects. Standard errors are clustered at the art unit level to account for concerns regarding intra-art unit correlation of error terms, particularly as they relate to unobserved routines. Columns 1–4 of Table 2 provide the focal set of results evaluating the effect of WFA on productivity.

TABLE 2 Causal Estimates of WFA on Productivity

Variables	(1) Total Actions	(2) Total Actions	(3) Total RCEs	(4) Total RCEs
WFA	0.509	0.574	-0.0540	0.00123
	p = 0.014	p = 0.000	p = 0.293	p = 0.973
Controls:				
Expectancy	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Grade Fixed Effects	Yes	Yes	Yes	Yes
Examiner Fixed Effects	No	Yes	No	Yes
Observations	65,694	65,694	65,694	65,694
Adjusted R-squared	0.358	0.553	0.143	0.279

Notes. Standard errors are clustered at the art unit level. Observations are at the examiner-month level and utilize the "WFA sample" of experienced examiners for Columns (1) through (4)—a subset of the main dataset that is limited to examiners who transitioned to WFA in 2012 or 2013. WFA is an indicator variable that turns on for examiner-months that transitioned into the WFA (i.e. TEAPP) program. Controls are indicated in the table above. All columns utilize data from 2007–2015.

Columns 1 and 2 report results relevant to output. Specifically, Column 1 identifies the effect of WFA on the total number of actions completed by each examiner in a given month, with Column 2 including a set of examiner fixed effects to identify the effect not just within the sample of examiners transitioning to WFA in 2012 and 2013, but also within each examiner. There is a positive, highly significant effect of WFA on overall output of 0.574 actions (p-value = .000), roughly corresponding to a 4.42 percent increase in the total number of actions on a mean of 12.97 per month. Columns 3 and 4 present results indicating that WFA does not increase the amount of RCEs an examiner engages in (with or without examiner fixed effects, p-values = .293 and .973, respectively). In summary, Hypotheses 1a and 1b are both supported. It is important to note that since workers had to first transition into the WFH program prior to transitioning to the WFA program, the baseline level of productivity here is productivity of the examiner while on WFH.

2.6.2 Baseline comparison of WFH productivity and in-office productivity

Hypothesis 2 states that for workers in a pooled-interdependence (low-coordination) setting with a baseline level of task-specific human capital, WFH is associated with greater productivity than work-

ing in the office. To recap, the USPTO experimented with a series of remote work programs. There was a WFA program (i.e., TEAPP) that allowed eligible examiners to live and work at any location in the U.S., and the USPTO implemented remote work programs such as PHP that offered examiners less autonomy on location choice and were akin to WFH programs. Given that the bureaucratic assignment to remote work is valid only for WFA, we can no longer rely on the natural experiment in this setting, and we estimate the specification below within the full sample of existing examiners across all months (576,267 examiner-months from 2007–2015):

$$Output_{it} = \alpha_i + \beta_{1it} * WFA + \beta_{2it} * PHP_{<50} + \beta_{3it} * PHP_{>50} + \xi_{it} + \gamma_t + \lambda_i + \varepsilon_i$$

where $\beta_{1it} * WFA$, $\beta_{2it} * PHP_{<50}$, and $\beta_{3it} * PHP_{>50}$ are indicator variables for when an examiner enrolled in either of the three programs, indicators that remain on until the examiner switches programs. As before, ξ_{it} is a vector of controls that includes examiner month-specific grade level and examiner month-specific expectancy, while γ_t is a full set of time (year) fixed effects and λ_i is a set of examiner fixed effects, which are of particular importance in this exercise as they allow us to track examiners as they switch from program to program. As before, standard errors are clustered at the art unit level. Table 3 provides results from this estimation exercise:

.**TABLE 3** Degree of geographic flexibility – WFA vs. WFH

Variables	(1) Total Actions	(2) Total Actions
PHP (<50 Miles)	1.339	1.035
	p = 0.000	p = 0.000
PHP (>50 Miles)	1.131	0.487
	p = 0.000	p = 0.000
WFA	1.792	1.022
	p = 0.000	p = 0.000
Controls:		
Expectancy	Yes	Yes
Year Fixed Effects	Yes	Yes
Grade Level Fixed Effects	Yes	Yes
Examiner Fixed Effects	No	Yes
Observations	576,267	576,267
Adjusted R-squared	0.354	0.562

Notes. Standard errors are clustered at the art unit level. Observations are at the examiner-month level and utilize the full sample of examiners. WFA is an indicator variable that turns on for examiner-months that transitioned into the TEAPP WFA program. PHP <50 and >50, respectively, are indicator variables that identify examiner-months that have transitioned into the two PHP programs. The two PHP programs are akin to a traditional WFH program, with less geographic flexibility than a WFA program. Controls are indicated in the table above and may include year fixed effects, grade level (GS) fixed effects, expectancy (a measure of expected effort/output on an examiner-month level), and examiner fixed effects).

Column 2 reports the most restrictive specification with examiner fixed effects. As this model includes examiner fixed effects, we note that the coefficients are semi-additive: WFA captures the effect of remote work above and beyond PHP (>50 miles), as examiners must enroll in the latter before being eligible for the former. Hence, in this model, all telework programs incrementally increase productivity compared to working in the office, validating Hypothesis 2. The traditional WFH program, titled PHP (>50 miles), having the lowest productivity increase, while the other traditional WFH program, i.e. PHP (<50 miles), has roughly twice the impact as PHP(>50 miles) (p- value = .000 and .000 respectively). The impact of WFA, when interpreted additively with PHP (>50 miles), is far beyond the WFH programs (p-value = .000). It is important to note that we interpret these results in the context of one another rather than as causal estimates; the full sample regressions illuminate the relative differences between the remote work programs rather than the absolute improvements themselves. 23

2.6.3 Evidence on Mechanisms: Fingerprint of Geographic Flexibility

We now turn to establishing a fingerprint for mechanisms through which geographic flexibility can affect productivity. In doing so, we work within the constraints of available data and acknowledge that unobservable mechanisms might be in play. Our field interviews indicated that geographic flexibility benefitted individuals in a myriad of ways. To quote one examiner, "I'm a military spouse, which means I live in a world with frequent moves and personal upheavals that prevent many spouses from pursuing lasting careers, especially careers of their choice. WFA has been the most meaningful telework program that I have encountered in the military social sphere, as it allows me to follow my husband to any state in the U.S. at a moment's notice, and... pursue my own aspirations to contribute both to my home and to society. " Another examiner explained the benefit of geographic flexibility to his family as follows: "I have a daughter with a medical condition that, because of WFA, my family and I were able to search the northeast looking for the ideal location that would provide the services and supports for my daughter that we felt were best for her. As a result, we moved to Pennsylvania a little over two years ago. I cannot fathom what it must be like to uproot one's family AND have to find a new job in the process. I feel so lucky that I was able to make the move... to get the care my daughter needs and be able to keep the job I love doing." Our field interviews also indicated that moving to lower cost-of-living locations was a common benefit awarded by geographic flexibility. To quote another examiner, "I selected the Patent Office as D.C. seemed an interesting place to live with the understanding that I would make a lateral move to a private law firm in the D.C. area to improve my professional experience and to enhance my chances of leaving the D.C. area when I was ready to start a family. After three years, the Office began offering full-time telework schedules and I saw some of my colleagues depart D.C. to move to areas that were considerably more affordable....I have been a TEAPP worker for the last 4 years living in Alabama with my wife and two children."

We next turn to empirical analysis and first examine whether WFA examiners relocate to counties that lower their cost of living and, in effect, increase their real income. Utilizing previously described county-level cost-of-living data, we estimate the effects of telework on an examiner's current home cost-of-living index relative to Alexandria, VA, within both the full sample and the sample of examiners transitioning to WFA in 2012-2013. We estimate:

$$Cost\_of\_Living\_Reduction_{it} = \alpha_i + \beta_{1it} * WFA + \beta_{2it} * PHP_{<50} + \beta_{3it} * PHP_{>50} + \xi_{it} + \gamma_t + \varepsilon_i$$

where $Cost\_of\_Living\_Reduction_{it}$ is an examiner-specific measure of the reduction in the county cost-of-living index relative to Alexandria, VA, while WFA, PHP<50, and PHP>50 are indicator variables defined as before. This model similarly includes controls for year, grade level, and expectancy, but does not include examiner fixed effects, as those would absorb all time-invariant, examiner-specific variation in cost-of-living reductions. Table 4, Column 1 reports results from regressions utilizing the full sample of examiners[2]. We find evidence of substantial cost reductions associated with PHP (>50 miles) and WFA, on the order of two standard deviations in the distribution of cost reductions across all teleworking examiners (p-value = .000). As expected, PHP (<50 miles) does not show evidence of cost reductions, as those examiners must live in and around Alexandria, VA. The results remain robust in the WFA sample (Column 2, p-value = .000).

[2] As reported in Table 1, this sample has 576,274 examiner-month-level observations. We dropped a few observations, corresponding to examiners without worker location data

TABLE 4 Cost-of-Living Reduction

Variables	(1) Cost-of-Living Reduction	(2) Cost-of-Living Reduction
PHP (<50 Miles)	0.00335 p = 0.938	
PHP (>50 Miles)	18.57 p = 0.000	
WFA	18.54 p = 0.000	3.030 p = 0.000
Controls:		
Expectancy	Yes	Yes
Year Fixed Effects	Yes	Yes
Grade Level Fixed Effects	Yes	Yes
Examiner Fixed Effects	Yes	Yes
Observations	576,002	65,437
Adjusted R^2	0.828	0.761

Notes. Standard errors are clustered at the art unit level. Column (1) reports results from a regression of Cost-of-Living Reductions, indexed to 0 for Alexandria on dummies for being in either PHP program and being in WFA. Column (1) utilizes the full sample of examiners. In order to align with our main results, Column (2) reports results from the "causal sample" of examiners who transition to WFA in 2012 to 2013. Both columns limit samples to those locations with cost-of-living data and include controls for expectancy, year, and grade level, as well as examiner fixed effects. All columns utilize data from 2007–2015.

Next, we turn to studying whether cost-of-living benefits are correlated with output, given that wages and wage dispersion are arguably related to satisfaction and productivity of knowledge workers (Leana & Meuris [107]; Pfeffer & Langton [135]). A simple test of this potential mechanism is to compare the productivity of examiners that relocate to below-median cost-of-living locations with those that relocate to above-median cost-of-living locations, within the causal sample of WFA employees. More specifically, we estimate

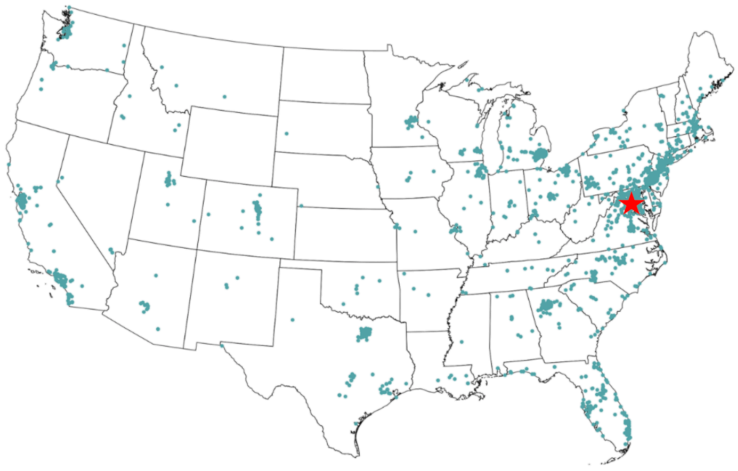
$$Output_{it} = \alpha_i + \beta_{1it} * WFA + \beta_{2it} * WFA * below\_median\_COL_i + \xi_{it} + \gamma_t + \lambda_i + \varepsilon_i$$

where $below\_median\_COL_i$ is an examiner-specific identifier that equals 1 when the examiner moves to a below-median cost-of-living location. This model includes controls for year, grade level, and examiner fixed effects. Table 5, Column 1 shows that examiners relocating to below-median

cost-of- living locations may experience increased output gains ($p < .006$)[3].

We next study WFA workers' geographic locational choices and their potential effects on productivity. As part of our examination of the data on WFA examiner location (Figure 2), we noticed clusters of examiners in a number of major metropolitan areas, including New York, Chicago, San Francisco, and Los Angeles, among others. These clusters can be expected given the concentration of population in these metropolitan areas. However, it also became clear there was a cluster of examiners in Florida, which cannot be explained by population alone.

FIGURE 2 WFA Examiner Locations



This figure illustrates the spatial distribution of WFA examiners at the USPTO as of August 2016. Each dot corresponds to a single unique examiner. Alexandria, Virginia (USPTO headquarters) is denoted by a red star.

We posited that a common reason for relocation to the coastal areas of Florida to seek alternate living arrangements when workers are close to retirement. We asked whether it was possible that more senior patent examiners were relocating to Florida at a higher-than-average rate, possibly as a first step toward retirement. Table 6, Column 2 indicates a positive and statistically significant corre-

[3] Heteroskedasticity-robust standard errors were used for this estimation exercise as there is little to no intra-art-unit variation across different geographies

lation between tenure at the USPTO and the probability of choosing to live in Florida (p-value = .001).

TABLE 5 Location drivers of productivity

Variables	(1) Total Actions	(2) Total Actions
WFA	0.311 p = 0.013	0.556 p = 0.000
Below-Median Cost-of-Living * WFA	0.302 p = 0.006	
Florida * WFA		0.592 p = 0.361
Controls:		
Year	Yes	Yes
Expectancy	Yes	Yes
Grade	Yes	Yes
Adjusted R^2	65,694	65,694
Observations	0.553	0.553

Notes. Observations are at the examiner-month level and utilize the 'WFA sample' of experienced examiners across both columns—a subset of the main dataset that limits to examiners that transition to WFA in 2012 or 2013. WFA is an indicator variable that turns on for examiner-months that have transitioned into the WFA (i.e. TEAPP) program. Controls are indicated in the table above, and include year fixed effects, grade level (GS) fixed effects, and expectancy (a measure of expected effort/output on an examiner-month level). All columns utilize data from 2007–2015. Heteroskedasticity-robust standard errors were used in this estimation exercise.

TABLE 6 Choice of Geographic Location by WFA Examiners

Variables	(1) In Florida	(2) In Florida
Grade 14 and Above	0.008 p = 0.579	-0.229 p = 0.156
Tenure (Years)		0.006 p = 0.001
Observations	2,064	2,064
Pseudo R-squared	0.001	0.007

Notes. Table presents coefficient estimates from two OLS regressions. Heteroskedasticity robust errors appear in parentheses. Observations are at the examiner level. The sample is restricted to examiners participating in the WFA (i.e., TEAPP) program.

We now look to see whether examiners relocating to a preferred location (e.g. Florida) perform equally with their peers. Results from this estimation exercise can be found in Table 5, Column

2, where we see that examiners relocating to Florida do not experience any statistically significant reductions (or gains) in productivity relative to their WFA peers (p-value = .361).

We also seek to identify whether the productivity improvements associated with a WFA regime are driven simply by mechanisms similar to those in WFH regimes, such as reductions in commute time and/or reduced monitoring. In order to estimate this, we compare examiners residing 50–75 miles from Alexandria, VA while working on the WFA regime to examiners residing over 75 miles away while working on the WFA regime. Examiners living 50–75 miles away from Alexandria, VA likely did not relocate as a result of moving from a WFH to a WFA regime and have likely self- selected into the geographic location of choice while being a WFH examiner. However, these examiners (i.e. living 50–75 miles away from Alexandria, VA) stopped commuting to the USPTO headquarters once they transitioned from a WFH regime to a WFA regime. To recap, on the WFH regime (i.e. PHP program), examiners were required to travel back to Alexandria, VA one day a week, incurring commute time and monitoring costs.

In contrast, examiners living over 75 miles away from Alexandria, VA while working on the WFA program have likely relocated beyond a reasonable commuting distance as they moved from the WFH to WFA regimes. These examiners too (like their peers on WFA in the 50–75 mile radius) experience a reduction in their weekly commute and monitoring. However, it is only when they move from WFH to WFA that they presumably experience the benefits of geographic flexibility for the first time. In other words, when examiners in the 50–75 mile radius move from WFH to WFA, they experience lower commute costs and less monitoring, given that they are presumably already in their geographic location of choice. In contrast, workers in the over 75 mile zone experience lower commute costs, less monitoring, and additionally experience the benefits of geographic flexibility for the first time when they move from WFH to WFA. Table 10 reports results for examiners residing 50–75 miles away from Alexandria, VA in Column 1, and examiners residing over 75 miles away in Column 2. We find that the WFA effect is driven entirely by examiners residing over 75 miles away,

pointing to productivity being driven by geographic flexibility, above and beyond the flexibility to reduce a commute (p-value = .000). We note that results are robust to other cutoffs, most notably 100 miles (rather than 75 miles).

While these results begin to paint a picture of geographic locational choices under a WFA regime, it is important to note that they do not capture the full range of possible mechanisms at play. For example, a worker may choose to relocate to a given location due to proximity to family (elderly parents, for example) or to return to a location where they have more friends and family (Dahl & Sorenson [57], dahl2010social). We expect there are a number of mechanisms not captured in the current analysis that could be highlighted by future research.

2.6.4 Work Practices that Enhance Productivity of WFA Workers

We conducted supplementary analyses to study productivity effects of work practices within the USPTO that might be correlated with work output of examiners in the WFA program. While an exhaustive examination of all relevant work practices is beyond the scope of this paper, our analysis is motivated by an observable work practice change where we could measure productivity effects. A USPTO directive in June 2013 mandated all teleworking patent examiners to utilize USPTO IT tools (e.g., logging into the USPTO virtual private network (VPN) and using USPTO messaging services). This provides us with the ability to measure the impact of IT tool use on productivity for a sample of WFA examiners. We postulate that the use of IT tools will enhance the productivity of WFA workers, especially WFA workers with a greater need for coordination (assistant examiners without signatory authority who had to coordinate with their supervisors). Research on remote work has indicated that the use of IT tools that foster situational awareness of the task helps in coordinating geographically dispersed workers (Malhotra & Majchrzak [118]). IT tools that are directed toward synchronous communication could arguably aid situational awareness and productivity of remote workers. Table 7 provides results from this estimation exercise. Column 1 reports results for

total actions, where the IT mandate improved output by 3 percent (off a mean of 12.9) exclusively for those examiners without signatory authority, that is, examiners who have to coordinate with their supervisors to get their work checked (p-value = .000). Column 2 reports results for RCEs, where we find no significant impact of the IT policy (p-value = .371).

TABLE 7 Effects of Mandated IT Use

Variables	(1) Total Actions	(2) Total RCEs
Examiner Needing Supervision	-1.119	-0.0311
	p = 0.000	p = 0.697
Mandated IT	-0.291	-0.0169
	p = 0.079	p = 0.758
Mandated IT * Examiner Needing Supervision	0.920	0.0659
	p = 0.000	p = 0.371
Controls:		
Expectancy	Yes	Yes
TEAPP Experience	Yes	Yes
Examiner Fixed Effects	Yes	Yes
Observations	19,255	19,255
Adjusted R^2	0.499	0.244

Notes. Standard errors appear in parentheses and are clustered at the art unit level. Observations are at the examiner-month level and are limited to those examiners on WFA before June 2013. Column (1) reports results from a regression of Total Actions, and Column (2) reports results for Total RCEs. Mandated IT is an indicator variable that turns on post-June 2013, and 'Examiner Needing Supervision' is an indicator for whether an examiner does not have "Full Signatory" authority and hence the work, while conducted alone, needs to be certified by a supervisor. Both columns include Expectancy, Year, and Examiner Fixed Effects, the latter of which absorbs nontemporal variation in Examiner status (i.e. needing supervision or not).

2.6.5 Robustness Checks

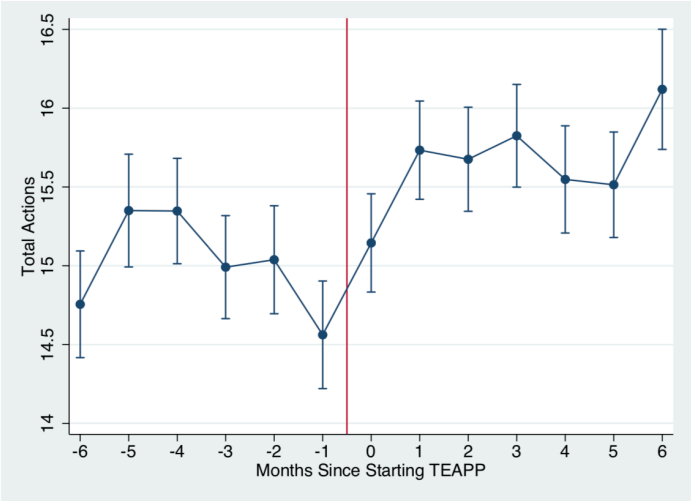
To test for concerns around time trends and reversion of performance to the mean post treatment (due to reciprocity or other unobserved mechanisms), we plot month-specific fixed effect coefficients in Figure 3 and find no evidence of post-treatment reversion to the mean. This analysis was repeated for a longer time window, and results remain robust. Given the point estimate of the month prior to treatment revealed in Figure 3, we additionally drop the two months prior to treatment from our regression analysis, and all results remain robust. Further, in order to validate our natural experiment, we look for evidence of selection in the time-to-WFA variation for those em-

ployees enrolling in WFA in 2012 or 2013. We estimate a model to determine whether previous performance, expected performance (expectancy), or rework is correlated with how soon an examiner receives WFA. In order to do so, we limit our sample to those examiners who obtained WFA in 2012 or 2013 and estimate variations on the following model:

$$Months_i = \alpha_i + \beta_{1it} * X_{it,<2012} + \xi_{it} + \varepsilon_i$$

Where $Months_i$ is an examiner-specific measure of the number of months (0–23) it took an eligible examiner to actually get in the program. X refers to total actions, total RCEs, or expectancy; hence, $X_{it,<2012}$ refers to an examiner's annual prior performance, rework, or expected performance. ξ_{it} is a set of controls for an examiner's GS at the month level. Table 8 presents results showing no evidence of previous performance, expected performance, or rework being correlated with the amount of time it took an examiner to transition to WFA, validating our identification strategy (all p- values > .10). We also conduct a placebo treatment test, explained and reported in Figure 4.

FIGURE 3 Difference-in-Differences Graph for Treatment



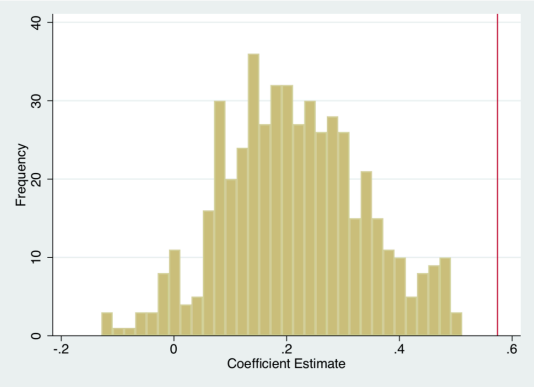
This figure plots the month-specific fixed effect coefficients estimated from a regression of total actions on controls for examiner, expectancy, grade level, and year. Standard errors are clustered at the art unit level. Treatment (TEAPP) is indicated with the red vertical line.

TABLE 8 Robustness tests for Selection into WFA

Variables	(1) Months to WFA	(2) Months to WFA	(3) Months to WFA
Total Action	0.00436 p = 0.452		
Expectancy		0.0867 p = 0.209	
Total RCE			0.0429 p = 0.147
Controls:			
Grade	Yes	Yes	Yes
Observations	2,771	2,771	2,771
Adjusted R-squared	0.002	0.003	0.003

Notes. All columns reflect regressions with the sample of examiners who received WFA in 2012 or 2013, limited to years prior to 2012 in order to observe pre-WFA performance. Observations are at the examiner year level, where Columns (1)–(3) estimate models testing whether prior output, expectancy, and rework are associated with the time it takes for an examiner to transition to WFA, the key source of causal variation in this study. Standard errors are clustered at the grade level.

Figure 4 Placebo Test



Notes. The analyses conducted here are as follows: There are 831 patent examiners in our dataset. For each of these examiners, we know the month they started participating in TEAPP. To perform one iteration of the placebo test, we shuffle the start month for all examiners—that is, we randomly assign, without replacement, when each examiner starts TEAPP. We merge these placebo treatments back into the panel dataset and recompute the WFA dummy that indicates whether the current month t is greater than or equal to examiner i's start month. We re-estimate the regression associated with Table 2 Column 2 with this new dummy variable (all other variables are the same), and we record the coefficient estimate associated with the synthetic treatment variable. We do this 500 times with different random shuffles of start date. Finally, we calculate a p-value by computing the proportion of the 500 iterations that yield coefficient estimates larger than what we find using the true data. None of the 500 estimates are greater than 0.574 ($p < 0.002$).

A potential concern is that examiners, upon transitioning to WFA, may scale back or distort effort relative to the quality of their work prior to being a WFA worker. For instance, while examiners may increase overall output, it is ex ante unclear whether leniency and/or effort change. Table 9 reports results from this exercise. Considering Columns 1 and 2, we find that the increase in first office actions is matched by a proportional increase in rejections (p-values = .038 and .032, respectively). We interpret this as evidence that examiners are no more or less lenient upon transitioning to WFA. Column 3 reports results for examiner-added citations—we are unable to distinguish from the null here (p-value = .401); there appears to be no reduction in examiner-added citations for those examiners transitioning to WFA.

TABLE 9 Robustness Tests for Examiner Effort and Leniency

Variables	(1) FOAs	(2) Rejections	(3) Examiner-Added Citations
WFA	0.135	0.194	-0.242
	p = 0.038	p = 0.032	p = 0.401
Controls:			
Expectancy	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
Grade Fixed Effects	Yes	Yes	Yes
Examiner Fixed Effects	Yes	Yes	Yes
Observations	55,791	55,791	55,791
Adjusted R^2	0.325	0.392	0.467

Notes. Standard errors are clustered at the art unit level. Observations are at the examiner-month level, where Column (1) utilizes first office actions (FOAs) as an outcome variable, Column (2) utilizes rejections as an outcome variable, and Column (3) utilizes examiner-added citations as an outcome variable. All regressions reflect analyses on the "WFA sample," limited to those with data on FOAs, rejections, and examiner-added citations. WFA is an indicator variable that turns on for examiner-months that have transitioned into the TEAPP WFA program. Controls are indicated in the table above, and include year fixed effects, grade level (GS) fixed effects, examine fixed effects, and expectancy (a measure of expected effort/output on an examiner-month level).

11. WELFARE ESTIMATES Using our estimates of a 4.4 percent increase in examiner-level production with no increase in the amount of RCEs, we make back-of-the-envelope calculations that suggest the following welfare gains: an increase in annual fees collected to the tune of $132 million, a one-time reduction of $0.7 million in hiring costs and a continuing annual cost savings of $2.75 million. In 2013, due in part to the agency's remote work options, the USPTO was ranked highest on the "Best Places to Work in the Federal Government" survey (USPTO[5]). Environmental benefits also accrue from the program; the agency estimated that in 2015, its remote workers avoided driving 84 million miles, thereby reducing emissions by more than 44,000 tons. Finally, in 2015, the USPTO estimated that it saved $38.2 million in real estate avoidance costs due to remote workers freeing up office space at headquarters (PTO[7]).

Finally, one particular feature specific to our setting is that the USPTO also helps set the rate of U.S. innovation, standing as one last bottleneck in the traditional innovation process. A 4.4 percent increase in patent grants could lead to innovation spillovers that amount to a total of $1.3 billion. We arrive at this estimate through back-of-the-envelope calculations. Choudhury et al. (2017)[51] indicate that the average number of patent grants from 2009–2012 was 211,973 patents per year;

this figure, taken into consideration with our estimated 4.4 percent increase in production, would lead to roughly 9,326 more patents being granted every year. Prior literature also indicates that the mean value for patents granted to U.S. patentees was $78,168 in 1992 dollars (Bessen [35]), with a median value of a patent to a U.S. assignee of $7,175 in 1992 dollars. We convert the mean and median values of a patent to a U.S. assignee to 2018 dollars and estimate that a 4.4 percent increase in production of patents at the USPTO creates $120 million in value for the U.S. economy (considering the median value of a patent in 2018 dollars) and $1.3 billion in value for the U.S. economy (considering the mean value of a patent in 2018 dollars).

2.7 Discussion

We study the relationship between geographic flexibility granted through a WFA program and worker productivity in a highly skilled work context. Our choice of setting presents us with two important empirical opportunities. First, the presence of a natural experiment originating from a bureaucratic policy allows us to mitigate the impact of endogeneity of selection into a WFA regime. Second, the dual mandate—to first spend two years in the office with other coworkers and then spend time in a traditional WFH program prior to becoming a WFA worker—allows us to control for the negative effects of remote work and to compare the productivity effects of WFH and WFA.

We find robust productivity effects, with a 4.4 percent increase in work output under WFA in comparison to WFH, and no effect on additional rework. In examining the increase to productivity under WFA, we conduct supplementary analyses that rule out WFH-related mechanisms such as lower commute time and reduced monitoring. These findings are important, as they suggest that WFA (and geographic flexibility) is a novel construct with unique benefits, not simply an extreme case of WFH. We provide evidence on mechanisms that could be driving the productivity increase under WFA. WFA examiners relocate to lower cost-of-living locations and we report a correlation

between relocating to a below-median cost-of-living location and productivity. We also study the choice of geographic location made by WFA workers and find a correlation between tenure at the USPTO and the likelihood of moving to a "retirement-friendly" location such as Florida. We also study conditions under which WFA worker productivity is further enhanced. Mandating IT usage appears to relax coordination constraints (and, thus, increase productivity) for employees requiring supervisory approval of work. A back-of-the-envelope calculation suggests that the increase in patents granted due to higher examiner productivity could result in $1.3 billion of added value to the U.S. economy in the best-case scenario. We also validate the Bloom et al. (2015)[36] result that WFH is related to higher productivity compared to working in office.

This paper makes contributions to research in the areas of remote work and non-pecuniary incentives. Our context of work-from-anywhere related to the construct of geographic flexibility distinguishes our study from prior research on remote work and working from home. WFH offers the worker temporal flexibility and flexibility in choosing working conditions (Gajendran & Harrison[72]). In contrast, WFA affords all of the benefits of WFH, plus the flexibility to relocate to a geographic location different from the location of the firm. As a result, while the WFH literature, notably Bloom et al. (2015)[36], identifies productivity-enhancing mechanisms such as reduced commute time, fewer work breaks, sick days, and the benefits of a quieter work environment, our focus on geographic flexibility points to additional mechanisms unique to WFA, such as the benefits of moving to a lower cost-of-living location, and increased psychic benefits to employees. As such, we argue that WFA needs to be studied as a separate form of remote work, with some underlying mechanisms similar to those of WFH, but with its own unique set of effects on workers and organizations. We also present a nuanced result related to the degree of geographic flexibility. Specifically, we find that a "middling" amount of geographic flexibility (i.e., PHP>50) is worse than very little flexibility (i.e., PHP<50) or a very strong case of geographic flexibility (i.e., WFA), evinced by the relative comparison of work output reported in Table 4. This finding has practical implications for

managers, suggesting that if a company hopes to enjoy the motivational benefits of increased perceived autonomy through the provision of a WFA regime, it must "cut the umbilical cord," giving employees true autonomy, rather than a piecemeal granting of autonomy.

We also contribute to the literature on incentives (Gambardella et al. [73]; Kryscynski [105]; Sauermann & Cohen [148]; Stern [157]). Our study suggests that the provisioning of an incentive such as WFA can create value for the firm while keeping wages constant, via an increase in worker productivity. In particular, we find not only that WFA workers who relocated were more likely to move to lower cost-of-living locations, but also that the workers who enjoyed higher-than-average cost-of-living savings (and thus a higher effective increase in real wages) demonstrated higher increases in productivity than relocating employees enjoying lower-than- average cost-of-living savings. These two findings taken together suggest a scenario in which a nonpecuniary incentive results in a pecuniary benefit to employees and firms alike. The pecuniary benefit of a reduced cost of living is equivalent to an increase in real income, which has been linked to increased employee satisfaction and productivity (Leana & Meuris [107]). While this result may not replicate for all types of workers in all organizations, it suggests that WFA could potentially be used as an effective firm-specific incentive to attract and retain skilled employees (Coff & Kryscynski [52]; Kryscynski [105][4].

Our study has several limitations. Similar to Bloom, et al. (2015)[36], our study is focused on a single organization. Additionally, it is plausible that in other settings where workers have greater dependence on coworkers and supervisors to accomplish their tasks, increased coordination costs might offset the gains from higher productivity. Future work should validate our findings in other settings that exhibit other forms of interdependence (i.e. sequential and reciprocal interdependence), where the worker might not have relevant prior task-specific human capital, and/or where the task is more or less routine compared to patent examination. Building on Kryscynski (2011)[105], we posit that nonpecuniary incentives such as WFA can and should be firm specific. For instance,

[4]Wiedner & Mantere [170] make a similar argument in the context of organizational separation

a firm could choose to provide a WFA option to experienced employees if the tasks performed by knowledge workers in the firm exhibit properties of pooled interdependence. However, WFA may not create value for other firms with stronger (i.e., reciprocal or sequential) interdependence regimes, and future research could examine this proposition.

Our results showing that differences in tenure at the USPTO are correlated to the worker's choice of geographic location open up avenues for future research. Our correlational finding that higher-tenured workers are more likely to choose a geographic location such as Florida (which is arguably better suited as a preretirement destination) suggests that future work can explore whether WFA could have career-extending benefits, motivating workers closer to retirement to remain in the workforce and be productive. Our research contributes to a very active managerial debate on the effectiveness of WFA. In February 2013, then-CEO Marissa Mayer famously rescinded the remote work program at Yahoo!, explaining in a company memo, "Some of the best decisions and insights come from hallway and cafeteria discussions, meeting new people, and impromptu team meetings. Speed and quality are often sacrificed when we work from home. We need to be one Yahoo!, and that starts with physically being together" (Swisher [159]). Yet, along with these highly visible retreats from WFA regimes, other employers continue (typically with less fanfare) to increase WFA opportunities and more generally support the concept of remote work. Akamai's "Akamai Anywhere" WFA policy is one such example (Mayer [123]7). In promoting the agency's WFA policy, NASA's Chief Technology Officer noted that, "The potential exists for... an employee's office to expand from a 12' by 12' room to virtually everywhere" (Porterfield [136]).

A series of empirical studies around WFA could help resolve this debate. It is plausible that the gains from WFA are restricted to settings where workers are approaching diminishing returns in learning from peers and/or are relatively less dependent on coworkers and supervisors to accomplish their tasks. Hence, it would be interesting to replicate our study in settings with varying degrees of worker interdependence (e.g., designers, software developers). Future research could also study the

duration of physical colocation required for new hires to acquire the tacit knowledge needed to perform the task with no increase in rework after moving to a WFA program. Similarly, study of the conditions (if any) under which workers could benefit from learning from other remote workers and knowledge spillovers among WFA workers could be beneficial. For example, it has been suggested that "innovation spaces," such as coworking spaces and incubators, are becoming a source of knowledge transfer that promotes innovation and collaboration (Wagner & Watch [169]). This argument suggests that there could be an optimal WFA policy that allows employees to interact to some degree with professional peers in a physical collaborative setting close to their chosen geography. These workers may experience increased productivity benefits from knowledge spillovers in their home geography, though this empirical question requires further exploration.

Thinking beyond the immediate debate around WFA and firm productivity, we believe that future research on WFA could also help inform managerial decision-making in the context of newer structures used to organize knowledge workers. A number of firms, primarily in the software and technology fields (such as Mozilla and Art Logic), are structured as virtual organizations in which WFA is the dominant form of work (Reynolds [140]). Many of these "all-remote" firms have also adopted new-generation technology tools; internal social tools such as Slack, Yammer, and Chatter, or embedded applications such as Microsoft Teams and JIRA are being implemented at a staggering rate (Leonardi & Neeley [111]). With these technologies further enabling WFA, researchers and firms will likely continue to explore the conditions under which geographic flexibility can contribute positively to remote worker productivity. Finally, the notion of geographic flexibility introduced in this study might have career-enhancing and career-extending effects; future research should study whether and when firms can extend the productivity of aging workers by giving them autonomy to relocate to "retirement-friendly" destinations.

As technology continues to expand avenues for communication and collaboration among virtual coworkers, and as major business centers grow more populous and congested, there is a need to

develop our understanding of how granting geographic flexibility via policies such as WFA affects productivity. To the best of our knowledge, our study represents the first set of robust econometric results on the productivity effects of moving workers from a WFH to a WFA regime and makes a contribution to the literature on remote work, nonstandard work, and nonpecuniary incentives.

3

IT and Productivity: Evidence from Telemedicine

3.1 Introduction

In recent years, there has been vibrant discussion of the ability for software to automate tasks as well as facilitate non-standard work arrangements, such as remote work. Most analysis has focused on routine, low-skilled tasks (Acemoglu & Restrepo [14], Autor & Dorn [26], Autor et al. [28], Goos & Manning [82]), with little empirical understanding about whether and how information technology interacts with skill-intensive, time-critical, non-routine tasks. And while there is a burgeoning literature on the productivity benefits of IT-enabled remote work, much of it focuses on either routine tasks or workers with little need to coordinate with one another (Bloom et al. [36], Choudhury et al. [50]). This study asks whether information technology is productivity enhancing in a skill-intensive, time-critical setting with high coordination requirements and non-routine tasks. Contemporary discussion has identified several industries and professions that face potential consequences as a result of the increasing role of technology and automation in skilled domains (e.g. finance and legal services)–among these, I study the provision of medical care by doctors. (Susskind & Susskind [158]).

In the past decade, spurred by government policy[1], digital technology adoption in U.S. health

[1] The Federal Health Information Technology for Economic and Clinical Health (HITECH) Act (part of

care has been comprised largely of systems to store and transmit electronic health records (EHRs). Looking forward, however, evidence points to telemedicine as the next major wave of technology adoption in healthcare. Defined as the "the remote diagnosis and treatment of patients by means of telecommunications technology" (Catalyst [48]), telemedicine is currently used by 76% of U.S. hospitals and is continuing to grow in share and intensity of use, with a current market size of $38.3 billion (AHA [12]). Unlike EHRs, which largely digitize analog records, telemedicine fundamentally changes the nature of the provider-patient relationship, and is often viewed as a solution to increasing concerns regarding patient access to specialty care and growing p shortages in the United States (Tuckson et al. [167]).

One field of medicine that has suffered from physician shortages for decades is critical (or intensive) care (Halpern et al. [91]). Responsible for the most severely ill patients in the hospital, critical care physicians are tasked with closely monitoring and stabilizing patients with acute risk of death. Each year, 500,000 people die in intensive care units (ICUs, the hospital departments associated with critical care) in the United States alone (Angus et al. [22]). For context, this amounts to approximately 18% of *all* deaths in the United States in 2018 (Kochanek et al. [99]). Aside from the tremendous mortality associated with ICU admissions[2], there are substantial financial costs of intensive treatment in the hospital as well. Overall spending on ICU care is roughly $260 billion annually, accounting for 20% of all hospital expenditure in the United States (Feeley [68]) and hence over 1% of national GDP (Martin et al. [121]). While ICU mortality has been declining in the past decade (Lilly 2017), the national average still hovers at approximately 15% (Dartmouth Atlas [13]), and ICU admission rates are *increasing* (Lilly et al. [112]). Given the enormous stakes at play, both in terms of human lives and financial cost, any intervention that improves the productivity of critical care could have staggering implications.

---

the broader American Recovery and Reinvestment Act of 2009)

[2] in 2014, the United States national average for ICU mortality was 14.7% (Dartmouth Atlas 2019)

I study the introduction of labor-augmenting technology in the ICU and its impact on outcomes. Using patient-level data from a large hospital network in the State of New York from 2013-2018, I estimate the effect of a sweeping telemedical intervention across 30 intensive care units in which in-person critical care physicians were almost entirely replaced by a system comprised of software algorithms and remote treatment, leaving only a few doctors to care for hundreds of ICU patients at a time. I control for ex-ante health via detailed, patient-encounter level medical records, information on the precise method of triage within the ICU, and clinically validated health scoring algorithms and estimate an average reduction in mortality of roughly 16% across a variety of specifications.

Importantly, the productivity impacts of IT are heterogenous: I find evidence that both the healthiest and absolute sickest patients experience a small increase in mortality, while patients near the center of the severity distribution enjoy the greatest improvements in mortality. While patient treatment rates don't appear to change on average, physicians appear to reallocate effort across patients, expending disproportionate effort near the center of the severity distribution and reducing effort near the tails.

Features of this setting allow me to disentangle the benefits that stem from the use of software algorithms and remote treatment—the gains attributable to this telemedical intervention appear driven by software algorithm use. Finally, complementary assets are critical in this setting—community hospitals lacking in staff, training, and IT expertise do not benefit from use of the eICU.

Telemedicine is ultimately productivity enhancing in this context, enabling physicians to take on far more patients while simultaneously improving outcomes. More specifically, the network was able to reduce physician workforce demand by roughly 70% in hospitals that transitioned to telemedicine, leading to cost savings of $12 million per year in labor across the entire network according to conversations with key stakeholders.

This study contributes to the growing body of literature on the productivity and outcomes ef-

fects of IT implementation in healthcare (Athey & Stern [25], Parente & McCullough [134], Miller & Tucker [125], Agha [20]). While the vast majority of studies of IT in healthcare focus on the adoption of electronic medical records (EMRs), which assist in patient data storage and transmission, this study examines the use of technology that fundamentally alters the provision of medical care itself. Additionally, unlike studies that use national aggregate data, this study uses detailed medical records for nearly 100K patient-encounters and precise details of the technology rollout to control for *ex ante* patient health and estimate the effect of telemedicine.

This study also contributes to the literature on technology implementation, automation, and task substitution. While there exists a rich literature on the effects of automation of routine, or low skilled tasks, there are few studies that empirically estimate the effects of technology on skilled labor. Those few studies tend to focus economy-level effects of technology on skilled *employment* (Autor et al. [28], Autor et al. [27]). And while there is a nascent literature on the effect of automation (namely AI) on skilled labor, the majority of empirical exercises use a task-based framework to analyze potential occupation-level effects (Acemoglu & Restrepo [14], Acemoglu & Restrepo [15], Brynjolfsson et al. [44]). By focusing within a large organization and observing detail on individual tasks (patients), this study is able to precisely estimate not just average effects, but also heterogeneity and more importantly can unpack drivers of benefits associated with technology and automation.

I proceed as follows: Section 2 provides detailed background on intensive care, telemedicine, and specific institutional detail regarding the intervention and my research setting. Section 3 presents a simple conceptual framework to fix ideas regarding the nature of productivity-enhancing technological change in medical care. Section 4 provides details on the data sample used for analysis, while Section 5 provides results from various estimation exercises. Section 6 concludes.

3.2 Background

This study focuses on an IT intervention in a setting characterized by complex, skill-intensive, non-routine tasks–the provision of intensive medical care. The research setting is a major hospital network in New York State; consisting of roughly 70,000 employees across 23 hospitals, this network is the largest private employer and healthcare provider in New York State. With geographic coverage of most of the New York City metropolitan area, the hospital network is responsible for the care of roughly 4 million unique patients each year, with annual revenues on the order of approximately $11 billion.

Like many hospital networks at the forefront of technology adoption, this network has invested substantially in the adoption of telemedicine as a result of patient demand, provider shortages, and potential cost savings. In 2013, the network began adoption of an electronic ICU system, a major shift in the modality of intensive care provision in which the work of in-person critical care physicians is largely substituted by investments in advanced hardware, software algorithms, and a few remote physicians providing coverage for particularly complex cases.

This section describes details regarding intensive care itself, the specifics of the electronic ICU and its rollout, as well as discussion of relevant background regarding IT and healthcare.

3.2.1 Telemedicine

Health care spending currently comprises 18% of the United States economy, with commensurately large opportunities for innovation and new technologies (Martin et al. [121]). In recent years, digital health investments at the intersection of technology and healthcare have rapidly grown in size and share of the overall market (Raphael [138]). Digital health has traditionally encompassed health IT, such as electronic medical records (EMRs), clinical decision support systems (CDSS), and has expanded to connected medical devices and complex software algorithms to monitor, diagnose, and

treat patients. While electronic medical records have been the focus of health IT investments in the past decade (Adler-Milstein & Jha [17]), telemedicine is now among the fastest growing segments in this space (AHA [12]).

Telemedicine—the remote diagnosis and provision of medical care by means of telecommunications technology—is a treatment modality in which in-person physicians or other providers are separated from the locus of medical care, using technology to facilitate advanced monitoring (in lieu of physical observation) and to direct care remotely. Promising increased efficiency and reach, telemedicine can provide care to patients that are otherwise too distant from a major healthcare hub as well as leverage economies of scale to provide care in facilities that cannot otherwise justify maintaining in-person coverage for certain specialties. Among the more common applications of telemedicine are telepsychology, telestroke, tele-urgent care, and recently, teleICU, or eICU.

3.2.2 Intensive Care Units

The intensive care unit (ICU) is a specialized department within a hospital that provides care to critically-ill patients . Staffed by specially trained nurses and physicians that are certified in critical care, the traditional ICU model relies on high staff-to-patient ratios and generally high-touch care in order to minimize mortality among a particularly risky population (Marshall et al. [120]). A traditional ICU has roughly 8 beds and is staffed by several nurses along with usually one (sometimes two) critical care physicians.

The principal job of the ICU is to stabilize and monitor patients with rapidly deteriorating health status, often due to an acute failure of key biological systems (e.g., respiration or circulation). Physicians and nurses coordinate in person to provide treatment and monitoring for these patients. Treatment within the ICU setting most often refers to the provision of respiratory assistance and active infusion of therapeutics (e.g. antibiotics, other pharmaceuticals). Monitoring entails the utilization of real-time data on patient vital signs ("vitals" e.g., heart rate, respiratory rate, oxygen

saturation levels) along with regular laboratory testing and in-person health status checks in order to ensure the stability of a given patient's health.

In my setting, patients enter the ICU from three sources, in the following order of incidence: (1) the emergency department, (2) post-surgical recovery, and (3) as inpatients. Patients are forwarded to ICU as a last resort, with the ultimate goal of restabilization for discharge either to the hospital itself, a skilled nursing facility, or to the patient's home. Upon admission to the ICU, detailed patient health indicators are measured, recorded, and used for triage purposes, including vitals, lab results, relevant diagnoses, and medical histories. An APACHE (Acute Physiology and Chronic Health Evaluation) score (an externally validated measure of predicted *ex ante* mortality) is calculated and used to triage care and attention from ICU staff.

3.2.3 eICU

As a response to rising costs of staffing and shortages of available critical care physicians, several hospital networks, including the one considered in this study, have implemented telemedicine within their intensive care units. Promising reduced monitoring overhead, more accurate triage, and more efficient distribution of physician workload (Doyle 2016), the electronic ICU (eICU) debuted in 2014 for three hospitals within the network, with subsequent rollout continuing through the present. The eICU is comprised of three major components: (1) an on-site hardware installation, (2) a series of software algorithms, and (3) a remote facility staffed by a handful of critical care physicians and triage nurses.

The primary goal of the on-site hardware installation is to collect and transmit high-frequency, real-time patient health data for use by software algorithms and the remote facility. This hardware suite replaces traditional bedside devices (e.g., pulse oximiters, respirometers, electrocardiograms, and oscillometers) with versions that provide higher-frequency data collection and integration with the software algorithms described below. In addition, cameras and video monitors are installed at

each bed to allow for remote monitoring and communication with the remote facility.

Third-party sourced, externally validated software algorithms available to both on-site and remote staff are the primary method of patient monitoring and triage. These algorithm takes into account real-time data from six major organ systems, generating a score that measures derangement (or deviations) from expected norms based on patient medical record inputs. In this iteration of the eICU, this system is rules based, using fixed parameters provided by the software vendor with no "learning" or adjustments since the initial rollout. Functionally, the system flags patients with immediate risk of physiological derangement in real-time, and physicians are thus able to make better informed triage and treatment decisions than would otherwise be possible with human data interpretation.

The dedicated remote eICU facility is staffed by both nurses and physicians, the availability of both being guaranteed 24/7. Each off-site physician is assigned three or four off-site triage nurses, who serve as the primary contact with on-site staff and assist in managing their off-site physician's "docket" as it fills with consult requests from hospitals all across the network. Under this new regime, each physician is no longer responsible for 8 beds as in a traditional ICU but rather roughly 60-80 patients remotely distributed at a given time.

3.2.4 eICU Implementation

Beginning in 2014, the hospital network began implementation of the eICU system at several hospitals. From field interviews with key stakeholders (including the Medical Director of Telehealth and the Chief Medical Informatics Officer), it was explained that the network intended to eventually study outcomes, and hence rolled out the IT system in a staggered, pseudo-random fashion.

More importantly, several key features of the rollout, and generally ICU care, minimize any concerns regarding selection into treatment in the first place. First, if a patient enters a given hospital, be it through the ER, via surgery, or as in inpatient, and experiences acute medical distress, the pa-

tient is almost exclusively sent to the "home" ICU (i.e., the ICU where they are an inpatient)–there are no common clinical processes by which patients are sent outside the "home" hospital and to an ICU in a different hospital (e.g., an "untreated" hospital). Second, because all ICU beds in a given hospital are converted to telemedical care simultaenously, there is no risk of patient sorting across traditional and electronic ICUs within the "home" hospital–neither the patient nor the physician have a choice as to where the patient is sent (it depends entirely on whether the patient's need for ICU care fell before or after the transition date). I empirically validate the above in Appendix A.

3.2.5 Technology Adoption in Healthcare

Prior Studies

As mentioned in Section 1, telemedicine is an increasingly common approach to healthcare delivery, although there are currently no large-scale studies in the management or economics literatures of the productivity or outcomes effects of the implementation of telemedical technology. Dorsey and Topol (2016)[62] provide a broad overview of the state of telemedicine, with a focus on trends and barriers to growth, but explicitly mention the dearth of empirical research on the benefits of telemedical care. And while the clinical literature has considered telemedicine at length, the majority of those studies focus on small sample cases at individual hospitals, if not individual units.

Outside of telemedicine, there is a growing, albeit small, body of literature studying the effects of IT and digitization on health care. The vast majority of these studies focus on one of the more salient forms of technology adoption in healthcare in the previous decades–the growing use of EHRs, a form of technology used for the storage and transmission of patient data. Generally speaking, most of these studies find limited benefits of EHR adoption and strong heterogeneity conditional on pre-existing firm resources. For instance, Miller and Tucker (2011)[125] and McCullough et al. (2010)[124] analyze the introduction of electronic medical records (EMRs) and find little benefit

when adopted alongside existing complementary resources. Agha (2014)[20] looks more broadly at healthcare IT, including not only EHR but simple CDSS, finding little to no benefits of IT adoption. McCullough et al. (2010)[124] finds evidence of heterogeneity such that only the most severe patients experience benefits. Few papers focus on topics outside the realm of EHR adoption. Most notably, Athey and Stern (2002)[25] observe that basic implementation of technology in emergency services (e911) leads to improved short-term patient outcomes for patients suffering from acute myocardial infarctions (AMIs, or heart attacks).

3.2.6 Potential benefits of telemedicine

The benefits of technology adoption, particularly telemedicine, are wide ranging. In addition to the empirical studies mentioned, there are clear benefits in terms of organizational processes that, while hard to isolate and estimate on their own, can help inform expectations over potential benefits.

For instance, technology on its own codifies organizational processes and reinforces protocols (Dewett and Jones 2001)[61]. Under the assumption that technology and the underlying organizational structure were designed to maximize outcomes, minimize risk, and improve efficiency, technology could ensure that employees maintain optimal performance, particularly in complex settings where task boundaries are uncertain (Roberts & Grabowski [143]). Put simply, information technology use can help ensure efficiency in complex organizational contexts.

Information technology has also long been discussed as a tool with the potential to improve physician efficiency, often due to biases surrounding time allocation across patients of varying severities (France et al. [71]). Further, technology that enables remote work physically separates clinicians from the patients themselves, leading to more reliance on objective measures and potentially more objective treatment (Lehoux et al. [109]). In the clinical literature, a focus on objectivity (particularly in acute care) has been associated with greater efficiency and better outcomes, and is the cornerstone of the modern push towards evidence-based medicine (Sackett et al. [146]). Therefore, one might ex-

pect information technology to lead to better outcomes and greater efficiency in case provision.

In addition to the benefits listed above, telemedicine implementation yields additional benefits as a result of the additional hardware and software used. The combination of real-time vital sign collection in conjunction with the software algorithms used for queue management (see section 2.3) have yielded improvements in the speed and accuracy of triage in a multitude of clinical studies across different settings (Görges et al. [84]; Raikhelkar & Raikhelkar [137]; Saffle et al. [147]; Young et al. [172]). Triage is one of the most important tasks undertaken by the ICU, both because of the scarcity of resources as well as the time-sensitivity of care (Gopalan & Pershad [83]), hence quick and accurate triage would directly contribute to better outcomes for individual patients as well as the ICU as a whole.

3.2.7 IT and Complementarities

There exists a large literature on complementary nature of IT efficacy and existing firm resources. Complementary assets, organizational adaptations, and changes are important in realizing productivity gains associated with IT investments (Bresnahan et al. [40]). More specifically, Bresnahan, Brynjolfsson, and Hitt (2002)[41] find that IT adoption is complementary with skilled labor and investments in work practices, among other resources. Bloom, Sadun, and van Reenan (2012)[37] find that firms in the United States enjoy larger gains in productivity stemming from IT than their global peers largely due to complementary "people management" skills. In healthcare, Dranove et al (2014)[63] finds strong evidence that complementary resources help health IT investments attain cost-effectiveness. These resources, such as software experience, can be internal to the firm or hospital, or can be drawn from local labor market expertise.

3.3 Conceptual Framework

3.3.1 Overview

This section presents a simple model illuminating the benefits of technology adoption in a healthcare setting.

3.3.2 Model

Physicians seek to maximize overall patient welfare subject to two constraints:

1. Minimum health
2. Time

To clarify, the minimum health constraint exists to ensure that physicians do not engage in practices that would leave the patient worse off than the alternative of no treatment, i.e. "do no harm". Additionally physicians are commonly constrained by time—they must make decisions regarding the allocation limited resources across patients presenting with varying severity.

Formally,

$$\begin{aligned} \theta^* \quad = \quad & \underset{\theta}{\text{argmax}} \quad \sum_{i=1}^{n} f(h,\theta)_i \\ & \text{subject to} \quad f(h,\theta)_i \geq h_i, \\ & \qquad\qquad \sum_{i=1}^{n} \theta = 1. \end{aligned} \tag{3.1}$$

where h refers to a patient's health upon admission to the ICU and θ refers to the level of treatment selected by the physician for a given patient, for which θ^* is the optimal level chosen for each patient to maximize overall patient welfare.

Built into this framework is the notion that patients present with varying levels of observable *ex ante* health, which I will henceforth refer to as severity. For simplicity, assume that there is no within-severity heterogeneity in treatment efficacy (i.e. similar patients respond similarly to identical treatment). Physicians provide treatment in order to maximize the objective function outlined in Equation (1).

However, consistent with the medical literature (Gruenberg et al. [89]), there is substantial cross-severity heterogeneity in treatment efficacy, such that $f(h, \theta)$ is not uniform, but rather follows an inverted U pattern. This pattern stems from two details:

1. There are decreasing returns to treatment for increasingly ill patients, as they are often past the threshold by which medical care can ensure their survival

2. There are low returns to treatment for particularly healthy patients, as they quite simply don't need as much treatment.

These patterns, in conjunction with the limited nature of medical resources (i.e., physician time), imply that accurate (and timely) triage is particularly important, a conclusion supported widely by the clinical literature (Truog [166], Gopalan & Pershad [83]).

3.3.3 Technology adoption

Technology adoption enters this model by enabling physicians to choose a θ closer to θ^* through three mechanisms, discussed in depth in section 3.2.6

1. Technology (telemedicine) removes physicians from the patient setting, mitigating biases in which physicians misallocate treatment effort (θ) across patients (e.g. spending too much time on the healthy or the terminally ill)

2. Technology (telemedicine) facilitates faster and more accurate triage, providing physicians with real time determination of the relative risk of death across their patients, and better information to make accurate decisions surrounding their choice of θ

The implications of technology adoption, in conjunction with the cross-severity treatment heterogeneity outlined in 3.2.1 are tested in Section 5.

3.4 Data and Summary Statistics

3.4.1 Summary

This project draws on three main sources of data. I begin with administrative data on all 394 ICU beds within 23 hospitals in the network and observe the date upon which each facility transitions to telemedicine from 2013 to 2018. I merge this data with detailed patient-encounter level medical records to identify whether a given patient-encounter was treated with the tele-ICU. These medical records also precisely identify the health status of each patient upon ICU admission. Finally, I utilize a third, separate administrative dataset on within-network transfers to identify whether patients were transferred into or out of a given ICU.

3.4.2 ICU Administrative Data

The first data source is a record of all ICUs from 2013 to 2018, along with dates of transition from traditional ICU care to eICU care for each facility. This data was linked to patient-encounter level records to identify whether a given patient-encounter was treated via a traditional ICU or the eICU. In this six year time span, 99,673 patient-encounters were recorded across 394 beds in 30 facilities. Separately, data on transfers out of a given ICU/eICU (to other locations) were also collected.

3.4.3 Patient Medical Records and APACHE scoring

In order to control for *ex ante* patient health and mitigate concerns of selection, each of the 99,673 patient-encounters above were linked to detailed EHRs provided by the hospital network, allowing access to the complete snapshot of patient health upon admission to the (e)ICU. This snapshot includes patient vital signs, relevant diagnoses, chronic conditions, and laboratory results.

In order to risk-adjust cases using detailed patient data (particularly lab results and vital signs) I closely followed the exact clinical process by which triage and treatment occur at the ICU by using APACHE scores to precisely control for patient health upon admission to the ICU rather than including all patient health covariates in an unstructured regression. APACHE takes into account 27 patient health variables and uses a series of predictive equations with predetermined, clinically validated weights (via extensive clinical trials) to output a score ranging from 0-292 that monotonically correlates with predicted mortality. These scores were calculated and recorded at the ICU itself as part of the intake process and were collected as part of the detailed medical records used in this study.

3.4.4 Outcomes Data

The principle outcome measure in this study is mortality—mortality is the most objective measure of patient outcomes in this setting for several reasons. First, mortality is not a diagnosis or treatment-specific outcome measure—for example, use of respiratory function as an outcome of interest for pneumonia patients would not be appropriate for patients suffering cardiac arrest. Second, avoiding mortality (and stabilizing patients on the brink of death) is the primary objective of ICUs; hence, measurement of mortality is a key internal benchmark that ICUs internally target. Third, as mortality is high among the intensive care unit patient population this provides substantial variation that can be used for study.

More specifically, this study focuses on hospital mortality rather than ICU mortality as use of hospital mortality alleviates concerns that an ICU might release a still-unstable patient back to the hospital, only for that patient to expire and not be captured in measures of ICU mortality.

Summary statistics for patient outcomes (mortality), as well as various covariates, including APACHE, are presented in Table 1, shown here.

TABLE 1

Variable	Sample Mean (± SD)	Min	Max
Mortality	0.104 ± 0.305	0	1
Treatment	0.548 ± 0.498	0	1
Male	0.547 ± 0.498	0	1
Age	66.2 ± 17.2	18	109
APACHE	58.0 ± 27.4	0	212
Transfer	0.0216 ± 0.145	0	1
Community	0.084 ± 0.278	0	1
Flagship	0.341 ± 0.474	0	1
Night	0.380 ± 0.485	0	1

n=99,673

3.5 Estimation and Results

3.5.1 Empirical strategy

Using data on treatment via electronic versus traditional ICUs in conjunction with detailed medical records to control for health and a variety of fixed-effects to account for unobservables, I estimate the effect of eICU treatment on mortality using the following general specification at the patient-encounter level:

$$Mortality_{ijt} = \beta_0 + \beta_1 * eICU_{it} + \beta_2 * APACHE_{it} + \chi_{it} + \gamma_{jt} + \nu_t + \varepsilon$$

where *eICU* is a patient-encounter indicator for treatment by the eICU, *APACHE* is the patient's APACHE health score, χ is a vector of patient controls, including patient age, gender, and diagnosis, γ is a set of facility fixed-effects, and ν is a set of year and month-of-year fixed effects. Here, β_1 is the coefficient of interest, measuring the outcomes effect of eICU implementation. Standard errors are clustered at the facility level for all results to follow.

3.5.2 Main results

The main estimation exercise seeks to identify the effect of eICU on patient-level mortality. Results can be found in Table 2, where column 1 includes controls for unit (i.e., department within a hospital), patient age, patient gender, patient APACHE scores, and diagnosis, while column 2 includes additional controls for year and month of admission. Focusing on column 2, the preferred specification, the coefficient on eICU implies a reduction in mortality of 1.70 percentage points off a mean of 10.4%, yielding a 16% percent reduction in mortality off of baseline mortality.

TABLE 2

Variables	(1) Mortality (Hospital)	(2) Mortality (Hospital)	(3) Active Treatment	(4) Active Treatment
eICU	-0.0202***	-0.0170***	0.0179	0.0176
	(0.00454)	(0.00436)	(0.0241)	(0.0248)
Controls:				
Facility, Gender, Age	Yes	Yes	Yes	Yes
APACHE	Yes	Yes	Yes	Yes
Diagnosis	Yes	Yes	Yes	Yes
Year and Month	No	Yes	No	Yes
Observations	99,673	99,673	99,673	99,673
Adjusted R-squared	0.277	0.280	0.189	0.191

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

Standard errors appear in parentheses and are clustered at the unit level. Observations are at the patient-encounter level and utilize the full sample of ICU admissions from 2013 to 2018. eICU is an indicator variable that turns on for those patient-encounters that were treated via the eICU. Controls are indicated in the table below, and include unit fixed effects, age, year, month, APACHE (an externally validated and medically relevant measure of severity), and diagnosis fixed effects.

Treatment Effort

The detailed patient medical records indicate whether a patient was actively treated or simply monitored by the eICU, and so I can test whether there are significant differences in treatment rates (i.e., whether a physician directed or otherwise provided treatment for a patient versus a baseline of monitoring) for eICU patients versus traditional ICU patients using the following specification:

$$Treated_{ijt} = \beta_0 + \beta_1 * eICU_{it} + \beta_2 * APACHE_{it} + \chi_{it} + \gamma_{jt} + \nu_t + \varepsilon$$

Again, the coefficient of interest is β_1. Table 2 above provides results from this exercise, where columns three and four provide increasing levels of added controls. The most saturated model, column 4, indicates that eICU patients don't receive any more treatment than non-eICU patients on

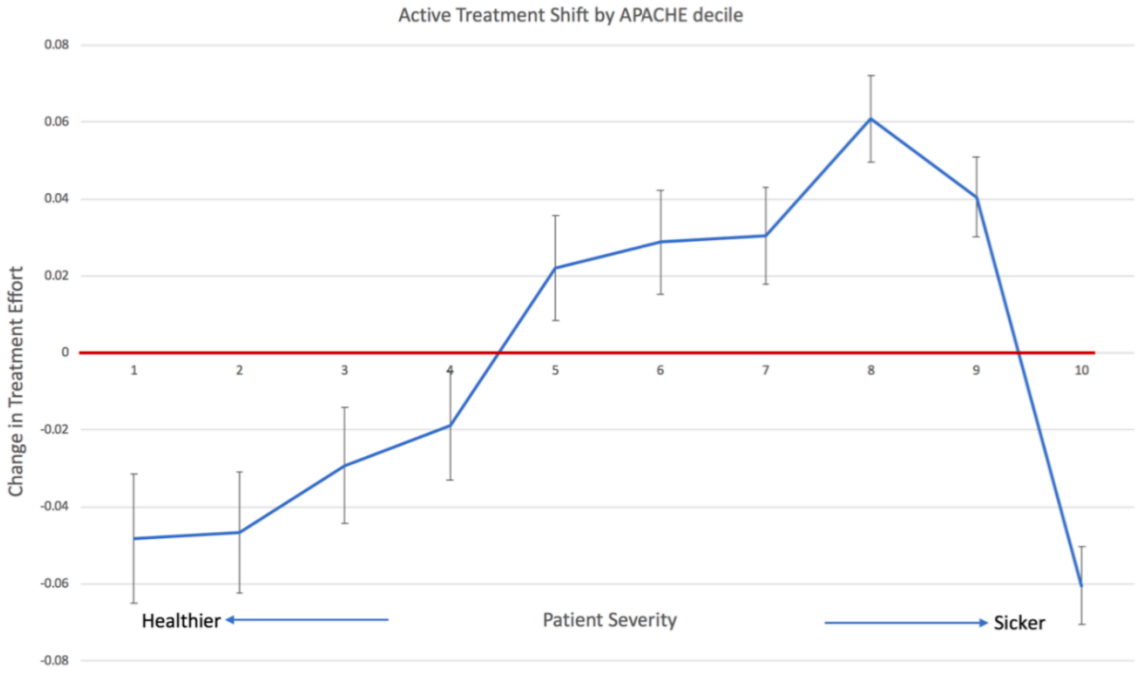
average.

3.5.3 Heterogeneity

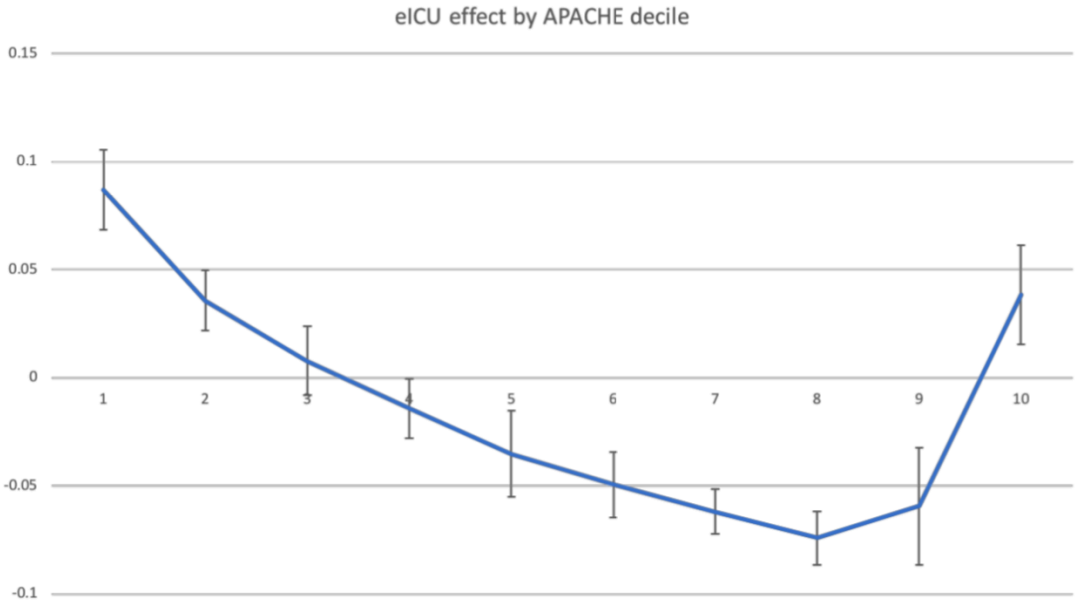
While physician treatment effort appeared flat on average, this could mask heterogeneity across patients of varying *ex ante* health status. To test for this, I estimate a model in which a dummy for treatment was regressed on deciles of APACHE scores as a proxy for patient severity. Results of this exercise are visualized in Figure 1, Panel A, where we see that treatment rates differ drastically by patient health, with the healthiest patients and the sickest upper decile seeing less treatment effort. These results, in conjunction with the average mortality improvements seen in 5.2, suggest that physicians are making choices to more optimally allocate time and effort that were not possible before the eICU implementation.

In order to further test the consequences of these reallocation decisions, I estimate a model similar to the above, but using interactions of *eICU* with deciles of APACHE scores. Estimates from this exercise are plotted in Figure 1, Panel B, where I observe heterogeneous effects across patient health, with the healthiest (lowest) few deciles observing increases in mortality improvement in mortality comes from moderate patients. The sickest decile also sees a increase in mortality.

Figure 1



Panel A



Panel B

3.5.4 Mechanisms

Disentangling productivity drivers

The main result showed a substantial reduction in mortality for those patients treated by the eICU, however, it is yet unknown whether those outcomes improvements are a result of the use of software algorithms to direct patient care, or rather utilization of the remote physicians. My data includes information on the availability of on-site physicians during certain hours—depending on the time-of-day a patient was admitted to the ICU[3], initial care would have been directed either by an on-site physician or off-site physician. As both physician types use software algorithms, differences in outcomes can be attributed to remote versus on-site treatment. I estimate,

$$Mortality_{ijt} = \beta_0 + \beta_1 * eICU_{it} + \beta_2 * eICU * Remote_{it} + \beta_3 * APACHE_{it} + \chi_{it} + \gamma_{jt} + \nu_t + \varepsilon$$

where *eICU* is a patient-encounter indicator for treatment by the eICU, *APACHE* is the patient's APACHE health score, and fixed effects are the same as defined in 3.5.1.

Results from this estimation exercise can be found in Table 3, where we find that in the most saturated model (column 2), the gains from eICU usage are driven entirely by software algorithms, and not treatment using remote physicians.

[3] coverage does not follow a regular pattern within or across facilities

TABLE 3

Variables	(1) Mortality	(2) Mortality
Algorithms only	-0.0170*	-0.0145**
	(0.00628)	(0.00262)
Algorithms + Remote	0.00773	-0.00614
	(0.00697)	(0.00491)
Controls:		
Facility, Age, Gender	Yes	Yes
APACHE	Yes	Yes
Year and Month	Yes	Yes
Admission Hour	Yes	Yes
Diagnosis	No	Yes
Observations	99,673	99,673
Adjusted R-squared	0.247	0.280

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

Standard errors appear in parentheses and are clustered at the unit level. Observations are at the patient-encounter level and utilize the sample of ICU admissions from 2013 to 2018 for patients seen during the day shift. Controls are indicated in the table below and include unit fixed effects, age, gender, APACHE (an externally validated and medically relevant measure of severity), year, month, and diagnosis fixed effects.

Complementary resources

Section 3.2.6 highlighted the importance of complementary resources in ensuring the productivity improvements commonly associated with IT interventions. The focal hospital network is comprised of both teaching hospitals and smaller community hospitals with more limited resources (i.e. limited staffing and complementary technology). With this in mind, I can test whether the eICU improves outcomes for community hospitals that lack the depth of complementary resources available to more advanced hospitals by estimating the following model,

$$Mortality_{ijt} = \beta_0 + \beta_1 * eICU_{it} + \beta_2 * eICU * Community_{it} + \beta_3 * APACHE_{it} + \chi_{it} + \gamma_{jt} + \nu_t + \varepsilon$$

where *eICU* is a patient-encounter indicator for treatment by the eICU, *APACHE* is the patient's APACHE health score, and fixed effects are the same as defined in 3.5.1.

Results from this estimation exercise can be found in Table 4, where we find that in column 2, the most saturated model, the gains from the eICU are near zero for community hospitals (the coefficients should be interpreted additively, such that the main eICU effect is a reduction in hospital mortality of 2.4 percentage points, while the eICU effect for community hospitals is 1.9 percentage points *worse* than the main effect).

TABLE 4

Variables	(1) Mortality	(2) Mortality
eICU	-0.0286***	-0.0242***
	(0.00509)	(0.00411)
eICU*Community	0.0228***	0.0194***
	(0.00225)	(0.00352)
Controls:		
Facility, Age, Gender	Yes	Yes
APACHE	Yes	Yes
Year and Month	Yes	Yes
Diagnosis	No	Yes
Observations	99,673	99,673
Adjusted R-squared	0.245	0.277

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

Standard errors appear in parentheses and are clustered at the unit level. Observations are at the patient-encounter level and utilize the sample of ICU admissions from 2013 to 2018 for patients seen by facilities that maintained in-house critical care coverage during the day shift. eICU is an indicator variable that turns on for those patient-encounters that were treated via the eICU, while Night is an indicator variable for patients seen during the night shift. Controls are indicated in the table below and include unit fixed effects, age, gender, APACHE (an externally validated and medically relevant measure of severity), year, month, admission hour, and diagnosis fixed effects.

3.6 Discussion and Conclusions

In this study, I estimate the effects of a large-scale telemedical intervention at a major hospital network in which in-person provision of critical care by doctors is substituted by a combined hardware/software suite as well as a handful of remote workers. In this setting, I focus on critical care—a type of medical care characterized by high mortality, high costs, and high utilization, where small improvements in productivity or outcomes could have very large consequences. Further, my focus on intensive medical care provides me the opportunity to use standardized measures of *ex ante* and *ex post* patient health, APACHE scores and mortality, respectively, allowing for precise estimation.

Telemedicine is strongly outcomes enhancing, providing an overall 16% reduction in mortality on average. This average effect masks significant heterogeneity in the treatment effect across patients of varying health status. Both the absolute healthiest and sickest patients experience slight increases in mortality, while more moderate patients substantial mortality reductions. This heterogeneity appears driven by physician behavior—physicians respond to technology by reallocating treatment effort; pulling effort away from the sickest patients and redistributing it to patients that appear to benefit the most.

These findings have important implications for policy and healthcare. Taking estimates of the mortality burden associated with ICUs, a back-of-the-envelope extrapolation of the estimated reduction in average mortality (16%) would imply the potential for 80,000 lives to be spared annually in the United States alone. Clinical evaluations of the cost effectiveness of intensive care unites for the most common admissions (e.g. cardiac arrest and respiratory arrest) point to gains in quality-adjusted life-years of 5 to 11, respectively (Graf et al. [85], Linko et al. [114]). Using standard estimates of the value of a quality-adjusted life-year of $50,000 - $150,000 (Neumann & Cohen [131]) leads to anywhere from $20 billion to $135 billion in welfare gains.

This study highlights the importance of software algorithms in driving employee behavior (e.g.

time allocation) and ultimately, performance—the vast majority of gains attributable to the eICU stemmed from on-site physician use of software and technology, *not* remote work. Further, use of algorithms drove clear heterogeneity in employee time allocation and patient outcomes, and careful considering of algorithmic bias is needed before use of these technologies at scale.

In addition to the considerations previously mentioned, firms undertaking major investments in technology (particularly health care organizations) should consider the importance of complementary assets and the availability of reliable support infrastructure. With respect to those assets, community hospitals in this hospital network that did not likely maintain the same level of staffing and training as teaching hospitals did *not* benefit from eICU adoption. Further, while this hospital network was able to undertake major investments in ICU-specific infrastructure (e.g. dedicated broadband internet connections, redundant power systems), this might not be feasible in rural areas or developing regions—those areas are often the focus of arguments supporting the development and diffusion of telemedical (or other health care information) technology.

Additionally, telemedicine (or technology that shares the key features of telemedicine) is particularly beneficial when deployed in firms that face unpredictable or unbalanced service load across sites—by aggregating demand for services across facilities and responding through one channel (the remote eICU), the hospital network was able to reduce inefficient allocation of physicians. In fact, one of the most promising use cases is precisely this—telemedicine (or any remote work aggregation technology) reduces the need to have idle employees "on call" and allows unpredictable, dynamic service loads to be handled centrally, leveraging economies of scale and maximizing efficiency.

An important caveat to this study is that this is an analysis of the healthcare setting, which, while advantageous for a variety reasons mentioned above, is a highly risk-averse setting with continuous performance monitoring, feedback, and immediate outcomes consequences—providing a multitude of opportunities for learning and continuous process improvement as technology is implemented in organizational routines. Whether technology could provide these benefits in settings with slower

feedback loops is an area of potential continued research.

A

Chapter 1: The 510(k) Process

Appendix Exhibit 1: Extract from 510(k) Statement

510(K) SUMMARY JAN 11 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.

The information in this appendix is taken directly from the FDA's official description of the 510(k) (premarket notification) process[1]

[1] 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 807 Subpart 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 is not substantially equivalent, the applicant may:

- resubmit another 510(k) with new data,
- request a Class I or II designation through the de novo process
- file a reclassification 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. See Is 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 the Investigational Device Exemption (IDE) regu-

lation (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 have Preamendment 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 on Medical 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, 1976 and which 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 the owner of 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 marketed after May 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. See Preamendment 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 to Third Party Review Program Information page.

B

Chapter 1: The PMA Process

Appendix 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.

PMA P150029: FDA Summary of Safety and Effectiveness Data Page 7

The information in this appendix is taken directly from the FDA's official description of the

Premarket Approval process[1]

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.

---

[1] http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/PremarketSubmissions/PremarketApprovalPMA/default.htm

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 the Premarket 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. The list of medical devices reviewed by CBER are available on

the Internet. In addition to CDRH guidance on Premarket Approval, specific medical device guidance for devices reviewed by CBER is 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 with 21 CFR 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 the Product Classification database for your device. You may also seek input from the review branch via the Pre-Submission Program.

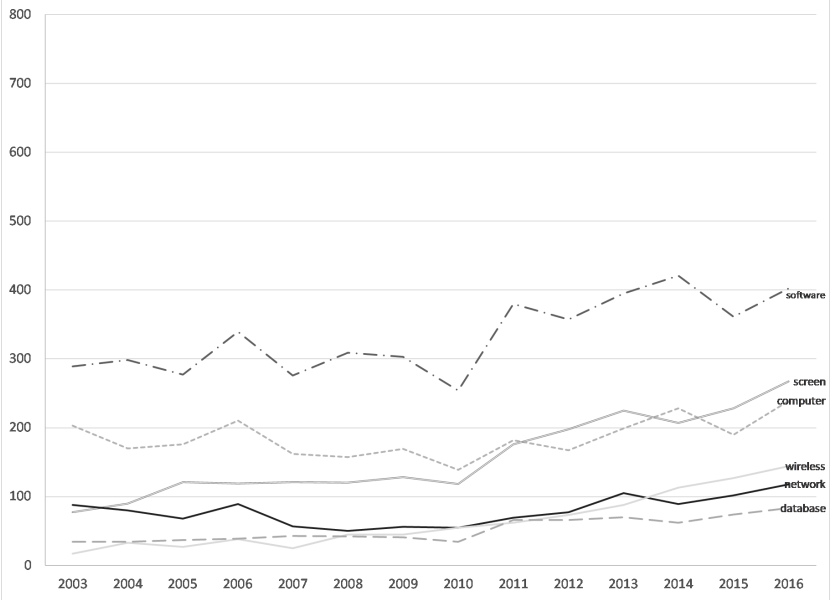
Clinical Investigations Section: Clinical investigations section includes study protocols, safety and effectiveness data, adverse reactions and complications, device failures and replacements, patient information, patient complaints, tabulations of data from all individual subjects, results of statistical analyses, and any other information from the clinical investigations. Any investigation conducted under an Investigational 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 FDA guidance documents that describe data requirements are available. Study protocols should include all applicable elements described in the device-specific guidance documents.

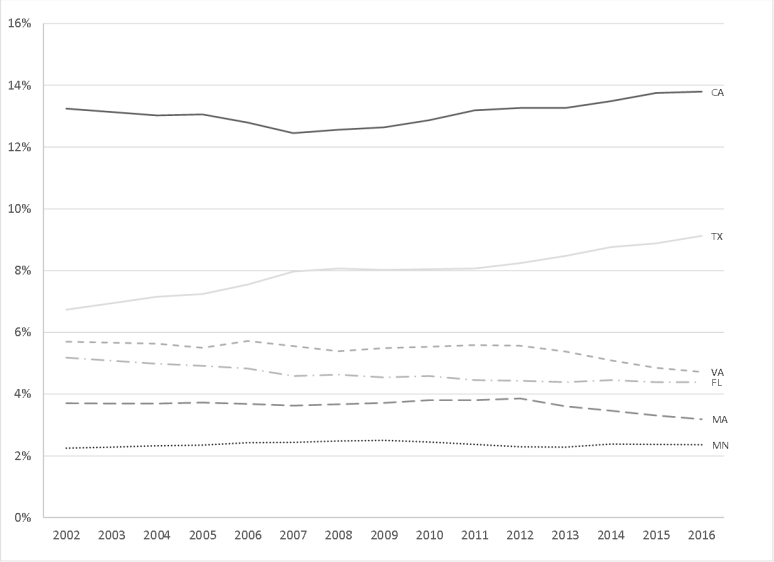
C

Chapter 1: Additional tables and results

Appendix Figure I: New digital devices

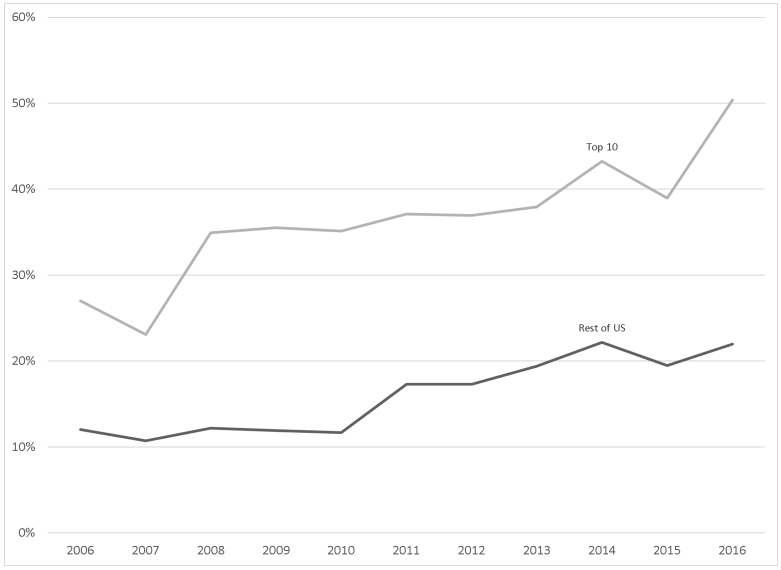


Appendix Figure II: Variation in state share of software engineers*

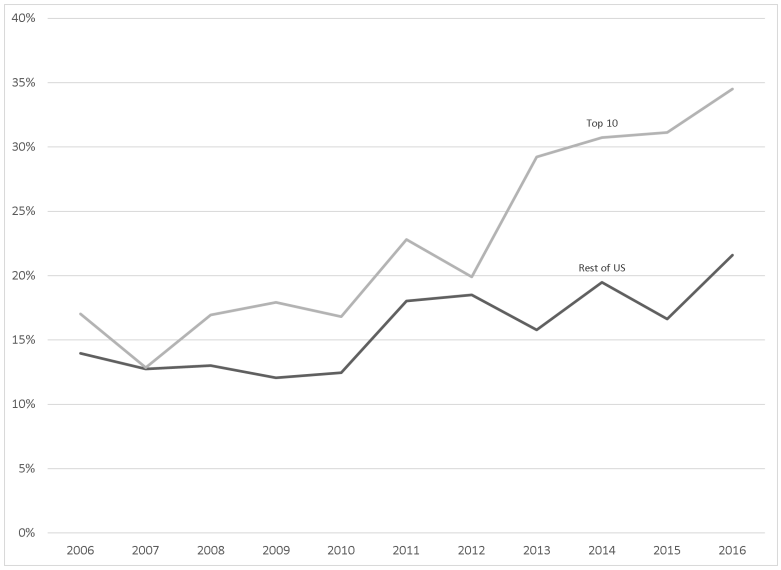


*linear imputation for years 2002-2004 and 2016

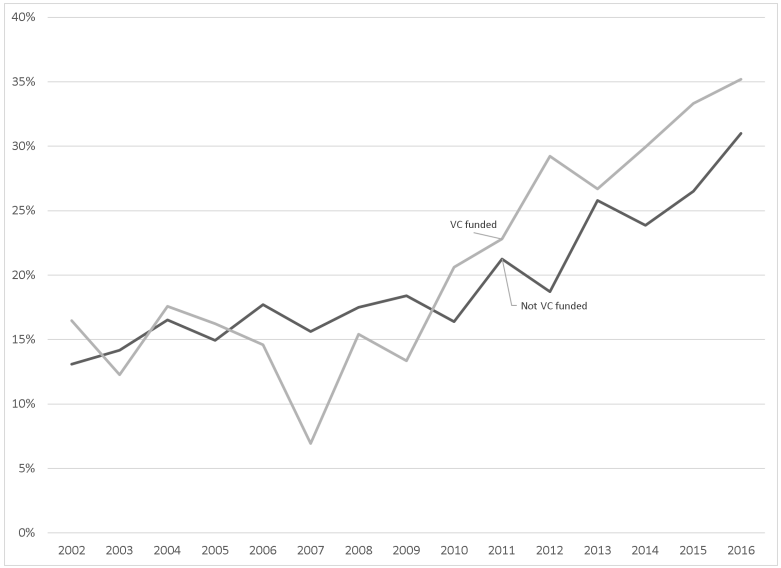
Appendix Figure III: Share of digital devices in *general* clusters vs. rest



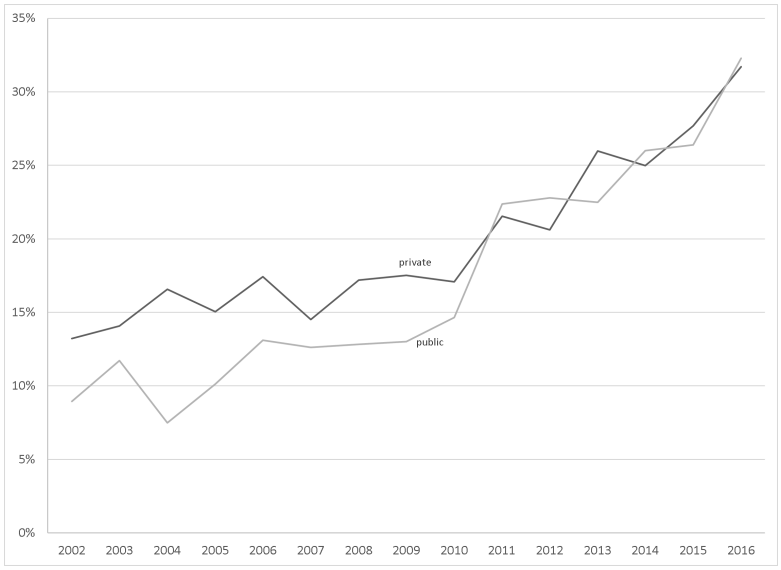
Appendix Figure IV: Share of digital devices in *class-specific* clusters vs. rest



Appendix Figure V: Share of digital devices: VC vs. non-VC-funded private firms



Appendix Figure VI: Share of digital devices: publicly-listed vs. private firms



Appendix Table I: Machine readable documents by sample year

Year	Readable Documents	Total Products	% Readable
2002	2,573	2,587	99.5
2003	2,565	2,579	99.5
2004	2,476	2,505	98.8
2005	2,338	2,364	98.9
2006	2,430	2,450	99.2
2007	2,245	2,318	96.9
2008	2,333	2,382	97.9
2009	2,287	2,333	98.0
2010	2,168	2,242	96.7
2011	2,405	2,452	98.1
2012	2,466	2,502	98.6
2013	2,404	2,428	99.0
2014	2,509	2,552	98.3
2015	2,334	2,408	96.9
2016	2,261	2,328	97.1
Total	35,794	36,496	98.1

Based on 8 most common medical specialty areas (classes).

Appendix Table II: 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%

This list includes all keywords found in >100 unique product descriptions.

Appendix Table III: Control variables: year and product class, MTI

Logit model: digital device commercialization		
	(1)	(2)
Cardiovascular	0.162***	0.162***
	(0.033)	(0.033)
Clinical Chemistry	0.099***	0.099***
	(0.028)	(0.029)
Dental	0.038*	0.038*
	(0.017)	(0.017)
Gastroenterology, Urology	0.062***	0.062***
	(0.014)	(0.014)
General Hospital	0.050*	0.050*
	(0.020)	(0.019)
General, Plastic Surgery	0.054***	0.053***
	(0.014)	(0.014)
Radiology	0.515***	0.516***
	(0.073)	(0.074)
N	22,291	22,291
Pseudo-R^2	0.2349	0.2339

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

Logit model results for years 2005-2016, inclusive. Column 1 includes year fixed effects; Column 2 includes a linear time trend. Omitted class = Orthopedic Devices; omitted year (Column 1) = 2005, marginal effects reported. Digital devices defined based on MTI method.

Appendix Table IV: Geographic and within-firm expertise, MTI

Logit model: digital device commercialization								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ln of digital device experience	0.080***		0.054***					0.051***
	(0.007)		(0.009)					(0.008)
Ln of same-class digital device experience		0.081***	0.033*					0.025*
		(0.009)	(0.013)					(0.011)
Ln of state software employment				0.006			0.009	0.005
				(0.004)			(0.005)	(0.004)
In digital device cluster (general)					0.056***		0.033***	0.012
					(0.010)		(0.007)	(0.007)
In digital device cluster (class-specific)						0.091***	0.085***	0.067***
						(0.009)	(0.009)	(0.007)
N	22,291	22,291	22,291	22,291	22,291	22,291	22,291	22,291

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

Firm experience and cluster variables are defined based on three most recent calendar years. All models control for volume of firm commercialization activity in past three years and state-level device clusters (representing all state-level medical device device commercialization). All models also include a full set of time and product class fixed effects. Marginal effects reported; standard errors are clustered at the product code level. Digital devices defined based on MTI method.

D

Chapter 3: Robustness

As described previously, several key features of this setting mitigate concerns of selection and allow for robust econometric identification of the causal impact of eICU on patient mortality. In order to validate those features, this section is dedicated to testing whether some facilities were selected in receiving technology before others, as well as empirically testing whether selection on patient health occurs on two margins: sorting of patients across traditional and electronic ICUs upon admission, and differential rates of patient transfers out across traditional and electronic ICUs.

Early versus late adoption

While interviews with directors and managers indicated that the timing of the technology rollout did not take into account hospital or facility characteristics, I am able to test this empirically by estimating whether patients treated at facilities that received technology early or late experienced differing outcomes.

More specifically, I estimate

$$Mortality_{ijt} = \beta_0 + \beta_1 * eICU_{it} + \beta_2 * eICU * Late_{it} + \beta_3 * APACHE_{it} + \chi_{it} + \gamma_{jt} + \nu_t + \varepsilon$$

where *eICU* is a patient-encounter indicator for treatment by the eICU, *APACHE* is the patient's APACHE health score, χ is a vector of patient controls, including patient age, gender, admission hour, and diagnosis, γ is a set of facility fixed-effects, and ν is a set of year and month-of-year fixed effects. Here, both β_1 and β_2 are the coefficients of interest, with the former measuring the outcomes effects for patients seen in facilities that received technology early hospitals, with the latter capturing the effect for patients seen at facilities receiving technology later relative to β_1. Standard errors are clustered at the facility level.

Results from this exercise can be found in Table D1, where columns 1 and 2 report hospital mortality outcomes, with column 2 being the more saturated model. The coefficient *eicu* $*$ *Late* is insignificant in both columns, indicating that patients treated at facilities that were late adopters of technology did not experience different outcomes.

TABLE D1

Variables	(1) Mortality	(2) Mortality
eICU	-0.0222***	-0.0175***
	(0.00577)	(0.00432)
eICU*Late	0.00391	0.00672
	(0.00422)	(0.00342)
Controls:		
Facility, Age, Gender	Yes	Yes
APACHE	Yes	Yes
Year and Month	Yes	Yes
Diagnosis	No	Yes
Observations	99,673	99,673
Adjusted R-squared	0.277	0.324

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

Standard errors appear in parentheses and are clustered at the unit level. Observations are at the patient-encounter level and utilize the sample of ICU admissions from 2013 to 2018 for patients seen by facilities that maintained in-house critical care coverage during the day shift. eICU is an indicator variable that turns on for those patient-encounters that were treated via the eICU, while Late is an indicator variable for patients seen in facilities that received the eICU later. Controls are indicated in the table below and include unit fixed effects, age, gender, APACHE (an externally validated and medically relevant measure of severity), year, month, and diagnosis fixed effects.

Transfers out of the eICU

In order to alleviate concerns of selection in which patients are transferred out of eICU facilities to facilities with traditional ICUs, I undertake a straightforward empirical test using the following model:

$$Transferred_{ijt} = \beta_0 + \beta_1 * eICU_{it} + \beta_2 * APACHE_{it} + \chi_{it} + \gamma_{jt} + \nu_t + \varepsilon$$

where controls are similar to those previously defined. Results from this exercise can be found in Table D2 column 2, where we observe that transfers do not increase for patients in the eICU, vali-

dating details of the setting described in section 3.2.4. In fact, transfers *decrease* by 0.63 percentage points off a mean of 2.16 percentage points, a near 30% reduction, pointing to increased treatment capacity within the IT-enabled ICUs.

TABLE D2

Variables	(1) Transfers	(2) Transfers
eICU	-0.00381*	-0.00630***
	(0.00216)	(0.00236)
Controls:		
Hospital, Gender, Age	Yes	Yes
APACHE	Yes	Yes
Diagnosis	Yes	Yes
Year and Month	No	Yes
Observations	93,357	93,357
Adjusted R-squared	0.016	0.016

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

Standard errors appear in parentheses and are clustered at the unit level. Observations are at the patient-encounter level and utilize the full sample of ICU admissions from 2013 to 2018. eICU is an indicator variable that turns on for those patient-encounters that were treated via the eICU. Controls are indicated in the table below and include unit fixed effects, age, APACHE (an externally validated and medically relevant measure of severity), and diagnosis fixed effects.

Selection on ex ante health

One area of concern regarding selection is whether healthier or sicker (*ex ante*) patients are sorted across traditional and electronic ICUs in a way that would bias the main result. More specifically, I test whether eICU patients appear *healthier* upon admission to the ICU compared to traditional ICU patients. I estimate

$$APACHE_{ijt} = \beta_0 + \beta_1 * eICU_{it} + \chi_{it} + \gamma_{jt} + \nu_t + \varepsilon$$

where controls are similar to those previously defined. Column 2 of Table D3 provides results from this estimation exercise, where we see that, if anything, APACHE scores are slightly *higher*, hence patients are marginally *sicker*, in the eICU versus the traditional ICU. Not only is this in a direction that might attenuate the main result (rather than bias it upward), it is potentially indicative of a relief of the ICU capacity constraint–the eICU allows for more patients, particularly sicker patients, as efficiency improvements have increased capacity.

TABLE D3

Variables	(1) APACHE	(2) APACHE	APACHE
eICU	3.352***	3.968***	2.647***
	(0.437)	(0.387)	(0.424)
Controls:			
Unit	Yes	Yes	Yes
Diagnosis	No	Yes	Yes
Year and Month, Gender	No	No	Yes
Observations	93,357	93,357	93,357
Adjusted R-squared	0.075	0.310	0.313

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

Standard errors appear in parentheses and are clustered at the unit level. Observations are at the patient-encounter level and utilize the full sample of ICU admissions from 2013 to 2018. eICU is an indicator variable that turns on for those patient-encounters that were treated via the eICU. Controls are indicated in the table below and include unit fixed effects, age, APACHE (an externally validated and medically relevant measure of severity), and diagnosis fixed effects.

Chapter 3: Separating Cardiovascular Events

Although the principal estimation exercises in the study utilize the full set of data and rely on disease fixed effects, interviews with staff members on site pointed to potential differences in outcomes for a certain set of patients: those suffering from acute heart failure. Prior to the introduction of the electronic ICU, heart failure patients were often (although not always) admitted to a specialized

"cardiac" intensive care unit (CICU) that were staffed with physicians that had particular experience in treating cardiac cases. Post-eICU, patients suffering from acute heart failure were seen by generalist critical care physicians in the eICU, not specialized physicians. This introduces potential bias into estimates of the true effect of telemedicine—this section is dedicated to separating cardiac cases in an effort to gain more clarity into those true effects.

Main effect

Using data on treatment via electronic versus traditional ICUs in conjunction with detailed medical records to control for health and a variety of fixed-effects to account for unobservables, I estimate the effect of eICU treatment on mortality using the following general specification at the patient-encounter level:

$$Mortality_{ijt} = \beta_0 + \beta_1 * eICU_{it} + \beta_2 * APACHE_{it} + \chi_{it} + \gamma_{jt} + \nu_t + \varepsilon$$

where *eICU* is a patient-encounter indicator for treatment by the eICU, *APACHE* is the patient's APACHE health score, χ is a vector of patient controls, including patient age, gender, and diagnosis, γ is a set of facility fixed-effects, and ν is a set of year and month-of-year fixed effects. Here, β_1 is the coefficient of interest, measuring the outcomes effect of eICU implementation. Standard errors are clustered at the facility level for all results to follow. In this section, cardiac cases are defined as patients suffering from acute myocardial infarctions or cardiac arrest.

The main estimation exercise seeks to identify the effect of eICU on patient-level mortality, separated for cardiac and non-cardiac cases. Results can be found in Appendix E Table 1, where column 1 presents results from a fully specified model for cardiac patients, while column 2 presents results from a fully specified model for all non-cardiac patients. Column 3 presents results from all patients (similar to section 3.5.2).

TABLE E1

Variables	(1) Mortality	(2) Mortality	(3) Mortality
eICU	0.00270	-0.0181***	-0.0172***
	(0.0247)	(0.00430)	(0.00447)
Controls:			
Facility, Gender, Age	Yes	Yes	Yes
APACHE	Yes	Yes	Yes
Diagnosis	Yes	Yes	Yes
Year and Month	Yes	Yes	Yes
Observations	7,922	91,751	99,673
Adjusted R-squared	0.539	0.223	0.277

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

Standard errors appear in parentheses and are clustered at the unit level. Observations are at the patient-encounter level and utilize the full sample of ICU admissions from 2013 to 2018. eICU is an indicator variable that turns on for those patient-encounters that were treated via the eICU. Controls are indicated in the table below and include unit fixed effects, age, APACHE (an externally validated and medically relevant measure of severity), and diagnosis fixed effects.

Focusing on column 1, we see that patients suffering from cardiac conditions appear to receive no benefit from telemedicine, while non cardiac patients (in column 2) realize greater mortality gains than the average patient from the full sample, approximately a 20% reduction in mortality.

Treatment Effort

The detailed patient medical records indicate whether a patient was actively treated or simply monitored by the eICU, and so I can test whether there are significant differences in treatment rates (i.e., whether a physician directed or otherwise provided treatment for a patient versus a baseline of monitoring) for eICU patients versus traditional ICU patients using the following specification:

$$Treated_{ijt} = \beta_0 + \beta_1 * eICU_{it} + \beta_2 * APACHE_{it} + \chi_{it} + \gamma_{jt} + \nu_t + \varepsilon$$

Again, the coefficient of interest is β_1. Appendix E, Table 2 provides results from this exercise,

where columns one and two provide results from the most specified model for cardiac and non-cardiac patients, respectively. Column 3 provides results from the full sample for comparison. Focusing on Column 1, we see that patients suffering from cardiac conditions appear to receive nearly 7% less treatment on average than their non-cardiac condition peers in column 2.

TABLE E2

Variables	(1) Active Treatment	(2) Active Treatment	(3) Active Treatment
eICU	-0.0676**	0.0121	0.0176
	(0.0267)	(0.0260)	(0.0248)
Controls:			
Facility, Gender, Age	Yes	Yes	Yes
APACHE	Yes	Yes	Yes
Diagnosis	Yes	Yes	Yes
Year and Month	Yes	Yes	Yes
Observations	7,922	91,751	99,673
Adjusted R-squared	0.459	0.315	0.324

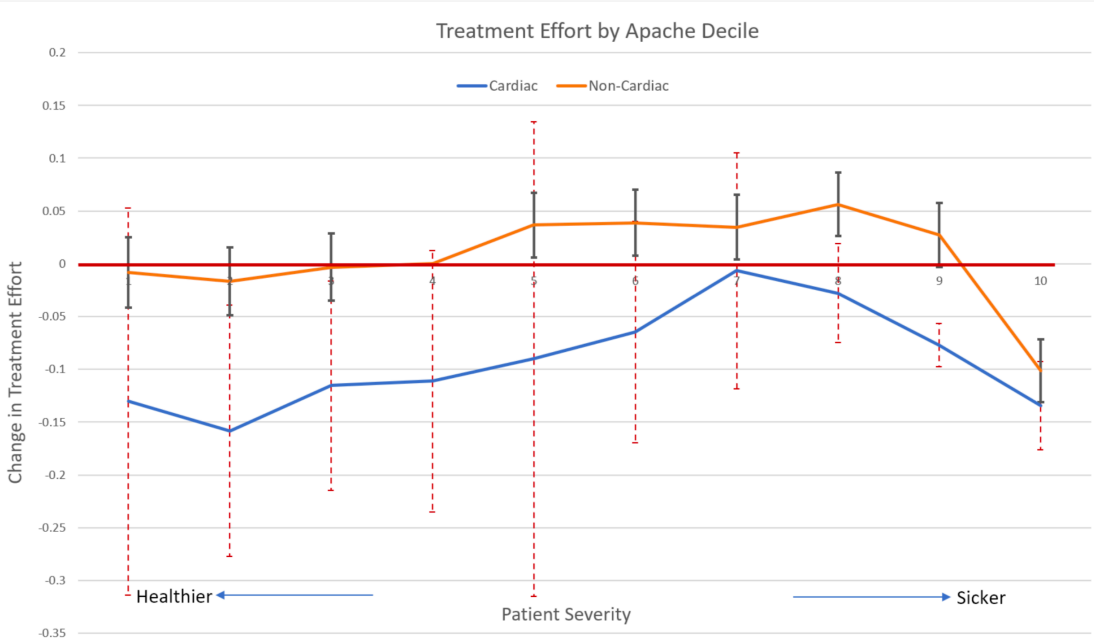
* p<0.10, ** p<0.05, *** p<0.01

Standard errors appear in parentheses and are clustered at the unit level. Observations are at the patient-encounter level and utilize the full sample of ICU admissions from 2013 to 2018. eICU is an indicator variable that turns on for those patient-encounters that were treated via the eICU. Controls are indicated in the table below and include unit fixed effects, age, APACHE (an externally validated and medically relevant measure of severity), and diagnosis fixed effects.

Heterogeneity

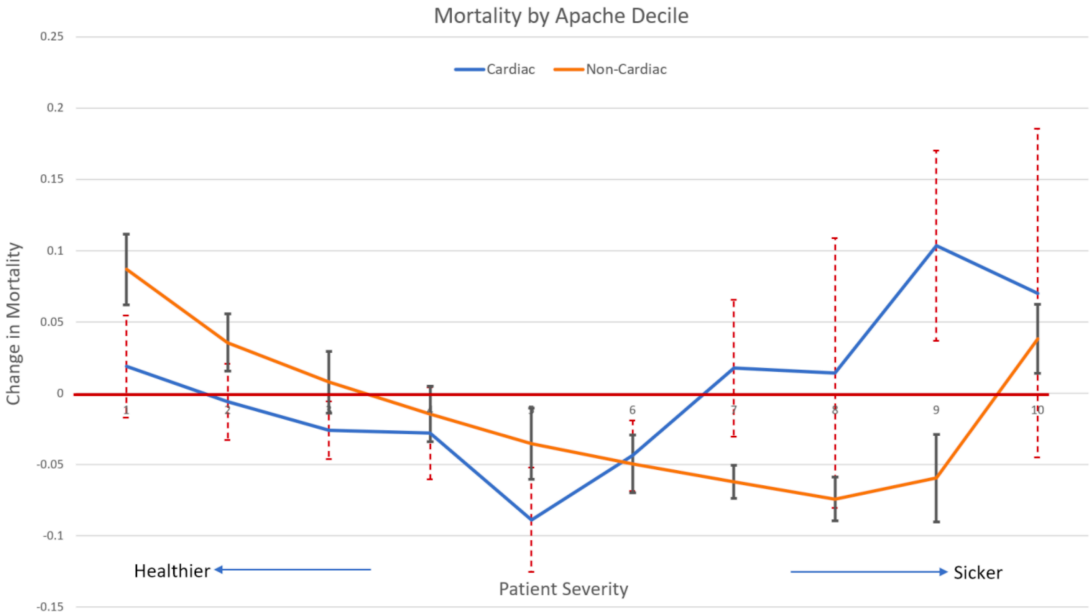
Similar to Section 3.5.3, I seek to uncover potential treatment effect and effort heterogeneity underlying the average effects estimated in the cardiac condition population above. To test for this, I estimate a model in which a dummy for treatment was regressed on deciles of APACHE scores as a proxy for patient severity. Results of this exercise are visualized in Appendix E, Figure 1, where we see that treatment rates differ by patient health for cardiac patient

Figure 1



In order to further test the consequences of these reallocation decisions, I estimate a model similar to the above, but using interactions of *eICU* with deciles of APACHE scores. Estimates from this exercise are plotted in Appendix E, Figure 2 where I observe heterogeneous effects across patient health, with middle of the severity distribution observing a slight improvement in mortality, while the upper deciles see much larger increases in mortality compared to non-cardiac patients.

Figure 2


Mortality by Apache Decile
Cardiac
Non-Cardiac
Change in Mortality
0.25
0.2
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
1
2
3
4
5
6
7
8
9
10
Healthier
Sicker
Patient Severity

References

[1] (1999). *Guidance for Industry, FDA Reviewers and Compliance on Off-The-Shelf Software Use in Medical Devices*. Technical report, Food and Drug Administration: Center for Devices and Radiological Health.

[2] (2005a). *Guidance for Industry and FDA Staff on the Content of Premarket Submissions for Software Contained in Medical Devices*. Technical report, Food and Drug Administration: Center for Devices and Radiological Health.

[3] (2005b). *Guidance for Industry: Cybersecurity for Networked Medical Devices Containing Off the-Shelf (OTS) Software*. Technical report, Food and Drug Administration: Center for Devices and Radiological Health.

[4] (2010). *Work life balance and the economics of workplace flexibility*. Technical report, Executive Office of the President, Council of Economic Advisors.

[5] (2013). *U.S. Patent and Trademark Office Ranks 1 in Best Places to Work in the Federal Government*. Technical report, United States Patent and Trademark Office - An Agency of the Department of Commerce.

[6] (2014). *Guidance for and FDA Staff on the Content of Premarket Submissions for Management of Cybersecurity in Medical Devices*. Technical report, Food and Drug Administration: Center for Devices and Radiological Health.

[7] (2015). *2015 Telework Annual Report*. Technical report, United States Patent and Trademark Office.

[8] (2016). *Guidance for Industry and Food and Drug Administration Staff on Postmarket Management of Cybersecurity in Medical Devices*. Technical report, Food and Drug Administration: Center for Devices and Radiological Health.

[9] (2016). *National Health Expenditure Accounts Highlights*. Technical report, Centers for Medicare and Medicaid Services.

[10] (2017). *State of the American Workplace 2017*. Technical report, Gallup, Inc.

[11] (2018). *Work From Anywhere – NASA's Telework Program*. Technical report, NASA.

[12] (2019). *AHA Fact Sheet: Telehealth*. Technical report, American Hospital Association.

[13] (2019). *The Dartmouth Atlas of Health Care*.

[14] Acemoglu, D. & Restrepo, P. (2018). Modeling automation. *AEA Papers and Proceedings*, 108, 48–53.

[15] Acemoglu, D. & Restrepo, P. (2020). Robots and jobs: Evidence from us labor markets. *Journal of Political Economy*, 128(6), 000–000.

[16] Acharya, V. & Xu, Z. (2017). Financial dependence and innovation: The case of public versus private firms. *Journal of Financial Economics*, 124(2), 223 – 243.

[17] Adler-Milstein, J. & Jha, A. K. (2017). Hitech act drove large gains in hospital electronic health record adoption. *Health Affairs*, 36(8), 1416–1422.

[18] Adler-Milstein, J., Kvedar, J., & Bates, D. W. (2014). Telehealth among us hospitals: several factors, including state reimbursement and licensure policies, influence adoption. *Health Affairs*, 33(2), 207–215.

[19] Agarwal, R. & Ohyama, A. (2013). Industry or academia, basic or applied? career choices and earnings trajectories of scientists. *Management Science*, 59(4), 950–970.

[20] Agha, L. (2014). The effects of health information technology on the costs and quality of medical care. *Journal of Health Economics*, 34, 19 – 30.

[21] Akerlof, G. A. & Kranton, R. E. (2005). Identity and the economics of organizations. *Journal of Economic perspectives*, 19(1), 9–32.

[22] Angus, D. C., Barnato, A. E., Linde-Zwirble, W. T., Weissfeld, L. A., Watson, R. S., Rickert, T., Rubenfeld, G. D., Group, R. W. J. F. I. E.-O.-L. P., et al. (2004). Use of intensive care at the end of life in the united states: an epidemiologic study. *Critical care medicine*, 32(3), 638–643.

[23] Argote, L. & Miron-Spektor, E. (2011). Organizational learning: From experience to knowledge. *Organization science*, 22(5), 1123–1137.

[24] Arora, A., Branstetter, L. G., & Drev, M. (2013). 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 Statistics*, 95(3), 757–775.

[25] Athey, S. & Stern, S. (2002). The impact of information technology on emergency health care outcomes. *The RAND Journal of Economics*, 33(3), 399–432.

[26] Autor, D. & Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the us labor market. *American Economic Review*, 103(5), 1553–97.

[27] Autor, D. H., Katz, L. F., & Kearney, M. S. (2006). The polarization of the us labor market. *The American economic review*, 96(2), 189–194.

[28] Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly journal of economics*, 118(4), 1279–1333.

[29] Azoulay, P., Graff Zivin, J. S., & Manso, G. (2011). Incentives and creativity: evidence from the academic life sciences. *The RAND Journal of Economics*, 42(3), 527–554.

[30] Azoulay, P. & Lerner, J. (2012). Technological innovation and organizations. In R. Gibbons & J. Roberts (Eds.), *The Handbook of Organizational Economics*. Princeton University Press.

[31] Babiarz, J. C. & Pisano, D. J. (2008). Overview of fda and drug development. In *FDA Regulatory Affairs* (pp. 11–42). CRC Press.

[32] Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4), 1593–1636.

[33] Bayus, B. L. & Agarwal, R. (2007). The role of pre-entry experience, entry timing, and product technology strategies in explaining firm survival. *Management Science*, 53(12), 1887–1902.

[34] Bélanger, F. (1999). Workers' propensity to telecommute: An empirical study. *Information & Management*, 35(3), 139–153.

[35] Bessen, J. (2008). The value of us patents by owner and patent characteristics. *Research Policy*, 37(5), 932–945.

[36] Bloom, N., Liang, J., Roberts, J., & Ying, Z. J. (2015). Does working from home work? evidence from a chinese experiment. *The Quarterly Journal of Economics*, 130(1), 165–218.

[37] Bloom, N., Sadun, R., & Van Reenen, J. (2012). Americans do it better: Us multinationals and the productivity miracle. *American Economic Review*, 102(1), 167–201.

[38] Booth, B. L. & Salehizadeh, B. (2011). In defense of life sciences venture investing. *Nature Biotechnology*, 29, 579 EP –.

[39] Branstetter, L. G., Drev, M., & Kwon, N. (2018). Get with the program: Software-driven innovation in traditional manufacturing. *Management Science*.

[40] Bresnahan, T., Greenstein, S., Brownstone, D., & Flamm, K. (1996). Technical progress and co-invention in computing and in the uses of computers. *Brookings Papers on Economic Activity. Microeconomics*, 1996, 1–83.

[41] Bresnahan, T. F., Brynjolfsson, E., & Hitt, L. M. (2002). Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *The quarterly journal of economics*, 117(1), 339–376.

[42] Briscoe, F. (2007). From iron cage to iron shield? how bureaucracy enables temporal flexibility for professional service workers. *Organization Science*, 18(2), 297–314.

[43] Brynjolfsson, E. & McElheran, K. (2016). The rapid adoption of data-driven decision-making. *American Economic Review*, 106(5), 133–39.

[44] Brynjolfsson, E., Mitchell, T., & Rock, D. (2018). What can machines learn, and what does it mean for occupations and the economy? In *AEA Papers and Proceedings*, volume 108 (pp. 43–47).

[45] Cable, D. M. & Turban, D. B. (2003). The value of organizational reputation in the recruitment context: A brand-equity perspective. *Journal of Applied Social Psychology*, 33(11), 2244–2266.

[46] Campbell, B. A., Coff, R., & Kryscynski, D. (2012). Rethinking sustained competitive advantage from human capital. *Academy of Management Review*, 37(3), 376–395.

[47] Carnahan, S., Agarwal, R., & Campbell, B. A. (2012). Heterogeneity in turnover: The effect of relative compensation dispersion of firms on the mobility and entrepreneurship of extreme performers. *Strategic Management Journal*, 33(12), 1411–1430.

[48] Catalyst, N. (2018). What is telehealth? *NEJM Catalyst*, 4(1).

[49] Chatterji, A. K. (2009). Spawned with a silver spoon? entrepreneurial performance and innovation in the medical device industry. *Strategic Management Journal*, 30(2), 185–206.

[50] Choudhury, P., Foroughi, C., & Larson, B. (2019). Work-from-anywhere: The productivity effects of geographic flexibility. *Harvard Business School Technology & Operations Mgt. Unit Working Paper*, (19-054).

[51] Choudhury, P., Khanna, T., & Mehta, S. (2017). The future of patent examination at the uspto.

[52] Coff, R. & Kryscynski, D. (2011). Invited editorial: Drilling for micro-foundations of human capital–based competitive advantages. *Journal of management*, 37(5), 1429–1443.

[53] Cohen, W. M. (2010). Chapter 4 - fifty years of empirical studies of innovative activity and performance. In B. H. Hall & N. Rosenberg (Eds.), *Handbook of The Economics of Innovation, Vol. 1*, volume 1 of *Handbook of the Economics of Innovation* (pp. 129 – 213). North-Holland.

[54] Cohen, W. M. & Klepper, S. (1996). A reprise of size and r&d. *The Economic Journal*, 106(437), 925–951.

[55] Cooper, C. D. & Kurland, N. B. (2002). Telecommuting, professional isolation, and employee development in public and private organizations. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, 23(4), 511–532.

[56] Cramton, C. D. (2001). The mutual knowledge problem and its consequences for dispersed collaboration. *Organization science*, 12(3), 346–371.

[57] Dahl, M. S. & Sorenson, O. (2010a). The migration of technical workers. *Journal of Urban Economics*, 67(1), 33–45.

[58] Dahl, M. S. & Sorenson, O. (2010b). The social attachment to place. *Social Forces*, 89(2), 633–658.

[59] Delgado, M., Porter, M. E., & Stern, S. (2014). Clusters, convergence, and economic performance. *Research Policy*, 43(10), 1785–1799.

[60] Deming, D. & Kahn, L. B. (2018). Skill requirements across firms and labor markets: Evidence from job postings for professionals. *Journal of Labor Economics*, 36(S1), S337–S369.

[61] Dewett, T. & Jones, G. R. (2001). The role of information technology in the organization: a review, model, and assessment. *Journal of management*, 27(3), 313–346.

[62] Dorsey, E. R. & Topol, E. J. (2016). State of telehealth. *New England Journal of m=Medicine*, 375(2), 154–161.

[63] Dranove, D., Forman, C., Goldfarb, A., & Greenstein, S. (2014). The trillion dollar conundrum: Complementarities and health information technology. *American Economic Journal: Economic Policy*, 6(4), 239–70.

[64] Ederer, F. & Manso, G. (2013). Is pay for performance detrimental to innovation? *Management Science*, 59(7), 1496–1513.

[65] Edmondson, A. C. & McManus, S. E. (2007). Methodological fit in management field research. *Academy of management review*, 32(4), 1246–1264.

[66] Evans, J. A., Kunda, G., & Barley, S. R. (2004). Beach time, bridge time, and billable hours: The temporal structure of technical contracting. *Administrative Science Quarterly*, 49(1), 1–38.

[67] Fatherly (2017). 30. akamai: The best places for new dads to work in 2016.

[68] Feeley, T. W. (2016). The value of icu care at the end of life. *NEJM Catalyst*, 2(5).

[69] Fehr, E., Herz, H., & Wilkening, T. (2013). The lure of authority: Motivation and incentive effects of power. *American Economic Review*, 103(4), 1325–59.

[70] Forman, C., Goldfarb, A., & Greenstein, S. (2016). Agglomeration of invention in the bay area: Not just ict. *American Economic Review*, 106(5), 146–51.

[71] France, D. J., Levin, S., Hemphill, R., Chen, K., Rickard, D., Makowski, R., Jones, I., & Aronsky, D. (2005). Emergency physicians' behaviors and workload in the presence of an electronic whiteboard. *International journal of medical informatics*, 74(10), 827–837.

[72] Gajendran, R. S. & Harrison, D. A. (2007). The good, the bad, and the unknown about telecommuting: Meta-analysis of psychological mediators and individual consequences. *Journal of applied psychology*, 92(6), 1524.

[73] Gambardella, A., Panico, C., & Valentini, G. (2015). Strategic incentives to human capital. *Strategic management journal*, 36(1), 37–52.

[74] Ganco, M., Ziedonis, R. H., & Agarwal, R. (2015). More stars stay, but the brightest ones still leave: Job hopping in the shadow of patent enforcement. *Strategic Management Journal*, 36(5), 659–685.

[75] Gans, J. S. & Stern, S. (2003). The product market and the market for "ideas": commercialization strategies for technology entrepreneurs. *Research Policy*, 32(2), 333–350.

[76] Gentzkow, M., Kelly, B. T., & Taddy, M. (2017). *Text as Data*. Working Paper 23276, National Bureau of Economic Research.

[77] Gibbons, R. & Waldman, M. (2004). Task-specific human capital. *American Economic Review*, 94(2), 203–207.

[78] Gibson, C. B. & Gibbs, J. L. (2006). Unpacking the concept of virtuality: The effects of geographic dispersion, electronic dependence, dynamic structure, and national diversity on team innovation. *Administrative science quarterly*, 51(3), 451–495.

[79] Golden, T. D., Veiga, J. F., & Dino, R. N. (2008). The impact of professional isolation on teleworker job performance and turnover intentions: Does time spent teleworking, interacting face-to-face, or having access to communication-enhancing technology matter? *Journal of Applied Psychology*, 93(6), 1412.

[80] Golden, T. D., Veiga, J. F., & Simsek, Z. (2006). Telecommuting's differential impact on work-family conflict: Is there no place like home? *Journal of applied psychology*, 91(6), 1340.

[81] Goldfarb, A., Greenstein, S. M., & Tucker, C. E. (2015). *Economic analysis of the digital economy*. University of Chicago Press.

[82] Goos, M. & Manning, A. (2007). Lousy and lovely jobs: The rising polarization of work in britain. *The review of economics and statistics*, 89(1), 118–133.

[83] Gopalan, P. D. & Pershad, S. (2019). Decision-making in icu–a systematic review of factors considered important by icu clinician decision makers with regard to icu triage decisions. *Journal of critical care*, 50, 99–110.

[84] Görges, M., Kück, K., Koch, S. H., Agutter, J., & Westenskow, D. R. (2011). A far-view intensive care unit monitoring display enables faster triage. *Dimensions of Critical Care Nursing*, 30(4), 206–217.

[85] Graf, J., Mühlhoff, C., Doig, G. S., Reinartz, S., Bode, K., Dujardin, R., Koch, K.-C., Roeb, E., & Janssens, U. (2008). Health care costs, long-term survival, and quality of life following intensive care unit admission after cardiac arrest. *Critical Care*, 12(4), R92.

[86] Grant, R. M. (1996). Toward a knowledge-based theory of the firm. *Strategic management journal*, 17(S2), 109–122.

[87] Greenstein, S. (2000). Building and delivering the virtual world: Commercializing services for internet access. *The Journal of Industrial Economics*, 48(4), 391–411.

[88] Greenwood, M. J. (1975). Research on internal migration in the united states: a survey. *Journal of Economic Literature*, (pp. 397–433).

[89] Gruenberg, D. A., Shelton, W., Rose, S. L., Rutter, A. E., Socaris, S., & McGee, G. (2006). Factors influencing length of stay in the intensive care unit. *American Journal of critical care*, 15(5), 502–509.

[90] Hall, B. H. & Lerner, J. (2010). The financing of r&d and innovation. In *Handbook of the Economics of Innovation*, volume 1 (pp. 609–639). Elsevier.

[91] Halpern, N. A., Pastores, S. M., Oropello, J. M., & Kvetan, V. (2013). Critical care medicine in the united states: addressing the intensivist shortage and image of the specialty. *Critical care medicine*, 41(12), 2754–2761.

[92] Hao, K. Y. & Jaffe, A. B. (1993). Effect of liquidity on firms' r&d spending. *Economics of Innovation and New Technology*, 2(4), 275–282.

[93] Harris, R. S., Jenkinson, T., & Kaplan, S. N. (2014). Private equity performance: What do we know? *The Journal of Finance*, 69(5), 1851–1882.

[94] Helfat, C. E. (1997). Know-how and asset complementarity and dynamic capability accumulation: the case of r&d. *Strategic Management Journal*, 18(5), 339–360.

[95] Henderson, R. & Cockburn, I. (1996). Scale, scope, and spillovers: The determinants of research productivity in drug discovery. *RAND Journal of Economics*, 27(1), 32–59.

[96] Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics*, 108(3), 577–598.

[97] Katila, R. & Ahuja, G. (2002). Something old, something new: A longitudinal study of search behavior and new product introduction. *Academy of management journal*, 45(6), 1183–1194.

[98] King, A. A. & Tucci, C. L. (2002). Incumbent entry into new market niches: The role of experience and managerial choice in the creation of dynamic capabilities. *Management Science*, 48(2), 171–186.

[99] Kochanek, K. D., Murphy, S. L., Xu, J., & Arias, E. (2019). Deaths: final data for 2017. *National vital statisics reports*.

[100] Kolstad, J. T. (2013). Information and quality when motivation is intrinsic: Evidence from surgeon report cards. *American Economic Review*, 103(7), 2875–2910.

[101] Korteweg, A. & Nagel, S. (2016). Risk-adjusting the returns to venture capital. *The Journal of Finance*, 71(3), 1437–1470.

[102] Kortum, S. & Lerner, J. (2001). *Does venture capital spur innovation?*

[103] Kossek, E. E., Lautsch, B. A., & Eaton, S. C. (2006). Telecommuting, control, and boundary management: Correlates of policy use and practice, job control, and work–family effectiveness. *Journal of Vocational Behavior*, 68(2), 347–367.

[104] Kramer, D. B., Xu, S., & Kesselheim, A. S. (2012). Regulation of medical devices in the united states and european union.

[105] Kryscynski, D. (2011). The strategic implications of firm-specific incentives-dissertation executive summary.

[106] Lazear, E. P. & Shaw, K. L. (2007). Personnel economics: The economist's view of human resources. *Journal of economic perspectives*, 21(4), 91–114.

[107] Leana, C. R. & Meuris, J. (2015). Living to work and working to live: Income as a driver of organizational behavior. *The Academy of Management Annals*, 9(1), 55–95.

[108] Lee, J., Kuo, Y.-F., & Goodwin, J. S. (2013). The effect of electronic medical record adoption on outcomes in us hospitals. *BMC Health Services Research*, 13(1), 39.

[109] Lehoux, P., Sicotte, C., Denis, J.-L., Berg, M., & Lacroix, A. (2002). The theory of use behind telemedicine:: how compatible with physicians' clinical routines? *Social science & medicine*, 54(6), 889–904.

[110] Lemley, M. A. & Sampat, B. (2012). Examiner characteristics and patent office outcomes. *Review of economics and statistics*, 94(3), 817–827.

[111] Leonardi, P. & Neeley, T. (2017). What managers need to know about social tools. *Harvard Business Review*, 95(6), 118–126.

[112] Lilly, C. M., Swami, S., Liu, X., Riker, R. R., & Badawi, O. (2017). Five-year trends of critical care practice and outcomes. *Chest*, 152(4), 723–735.

[113] Lin, S. C., Jha, A. K., & Adler-Milstein, J. (2018). Electronic health records associated with lower hospital mortality after systems have time to mature. *Health Affairs*, 37(7), 1128–1135.

[114] Linko, R., Suojaranta-Ylinen, R., Karlsson, S., Ruokonen, E., Varpula, T., Pettilä, V., study investigators, F., et al. (2010). One-year mortality, quality of life and predicted life-time cost-utility in critically ill patients with acute respiratory failure. *Critical Care*, 14(2), R60.

[115] Maisel, W. H. (2004). Medical device regulation: an introduction for the practicing physician. *Annals of Internal Medicine*, 140(4), 296–302.

[116] Makarius, E. E. & Larson, B. Z. (2017). Changing the perspective of virtual work: Building virtual intelligence at the individual level. *Academy of Management Perspectives*, 31(2), 159–178.

[117] Makower, J. (2010). Public health effectiveness of the fda 510(k) clearance process: Balancing patient safety and innovation. In *The Structure of the MedTech Innovation Ecosystem*.

[118] Malhotra, A. & Majchrzak, A. (2014). Enhancing performance of geographically distributed teams through targeted use of information and communication technologies. *Human Relations*, 67(4), 389–411.

[119] Mariani, M. (2004). What determines technological hits?: Geography versus firm competencies. *Research Policy*, 33(10), 1565 – 1582.

[120] Marshall, J. C., Bosco, L., Adhikari, N. K., Connolly, B., Diaz, J. V., Dorman, T., Fowler, R. A., Meyfroidt, G., Nakagawa, S., Pelosi, P., et al. (2017). What is an intensive care unit? a report of the task force of the world federation of societies of intensive and critical care medicine. *Journal of critical care*, 37, 270–276.

[121] Martin, A. B., Hartman, M., Washington, B., Catlin, A., Team, N. H. E. A., et al. (2019). National health care spending in 2017: growth slows to post–great recession rates; share of gdp stabilizes. *Health Affairs*, (pp. 10–1377).

[122] Mas, A. & Pallais, A. (2017). Valuing alternative work arrangements. *American Economic Review*, 107(12), 3722–59.

[123] Mayer, K. (2017). Benefits professional of the year creates a balance for akamai employees.

[124] McCullough, J. S., Casey, M., Moscovice, I., & Prasad, S. (2010). The effect of health information technology on quality in us hospitals. *Health Affairs*, 29(4), 647–654.

[125] Miller, A. R. & Tucker, C. (2017). Privacy protection, personalized medicine, and genetic testing. *Management Science*, 64(10), 4648–4668.

[126] Moore, E. (2018). 7 companies with work-from-anywhere jobs.

[127] Moretti, E. (2013). Real wage inequality. *American Economic Journal: Applied Economics*, 5(1), 65–103.

[128] Morton, F. M. S. (1999). Entry decisions in the generic pharmaceutical industry. *The RAND Journal of Economics*, 30(3), 421–440.

[129] Moser, P., Ohmstedt, J., & Rhode, P. W. (2017). Patent citations—an analysis of quality differences and citing practices in hybrid corn. *Management Science*, 64(4), 1926–1940.

[130] Nerkar, A. & Roberts, P. W. (2004). Technological and product-market experience and the success of new product introductions in the pharmaceutical industry. *Strategic Management Journal*, 25(8-9), 779–799.

[131] Neumann, P. J. & Cohen, J. T. (2018). Qalys in 2018—advantages and concerns. *Jama*, 319(24), 2473–2474.

[132] Newman, L. H. (2017). Medical devices are the next security nightmare. *WIRED*.

[133] Panzar, J. C. & Willig, R. D. (1981). Economies of scope. *The American Economic Review*, 71(2), 268–272.

[134] Parente, S. T. & McCullough, J. S. (2009). Health information technology and patient safety: evidence from panel data. *Health Affairs*, 28(2), 357–360.

[135] Pfeffer, J. & Langton, N. (1993). The effect of wage dispersion on satisfaction, productivity, and working collaboratively: Evidence from college and university faculty. *Administrative Science Quarterly*, (pp. 382–407).

[136] Porterfield, M. (2016). Work from anywhere: How to land that bigger office.

[137] Raikhelkar, J. & Raikhelkar, J. K. (2015). The impact of telemedicine in cardiac critical care. *Critical care clinics*, 31(2), 305–317.

[138] Raphael, R. (2019). Digital health hits record funding in 2018, totaling $14b: Is it enough?

[139] Rein, L. (2014). Patent office filters out worst telework abuses in report to its watchdog.

[140] Reynolds, B. W. (2014). 26 virtual companies that thrive on remote work.

[141] Reynolds, B. W. (2019). Top 20 companies hiring for work-from-anywhere remote jobs.

[142] Richard, H. J. & Oldham, G. (1976). Motivation through the design of work: Test of a theory. *Organizational behavior and human performance*, 16(2)), 250–79.

[143] Roberts, K. H. & Grabowski, M. (1999). Organizations, technology and structuring. *Managing organizations: Current issues*, 159.

[144] Rosenkopf, L. & McGrath, P. (2011). Advancing the conceptualization and operationalization of novelty in organizational research. *Organization Science*, 22(5), 1297–1311.

[145] Ryan, R. M. & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, 25(1), 54–67.

[146] Sackett, D. L., Rosenberg, W. M., & Gray, J. (1996). Evidence based medicine: what it is and what it isn 't: It's about integrating individual clinical expertise and the best external evidence. *British Medical Journal*, 312(7023), 71–72.

[147] Saffle, J. R., Edelman, L., Theurer, L., Morris, S. E., & Cochran, A. (2009). Telemedicine evaluation of acute burns is accurate and cost-effective. *Journal of Trauma and Acute Care Surgery*, 67(2), 358–365.

[148] Sauermann, H. & Cohen, W. M. (2010). What makes them tick? employee motives and firm innovation. *Management science*, 56(12), 2134–2153.

[149] Sauermann, H. & Stephan, P. (2013). Conflicting logics? a multidimensional view of industrial and academic science. *Organization science*, 24(3), 889–909.

[150] Schwartz, A. (1973). Interpreting the effect of distance on migration. *Journal of political economy*, 81(5), 1153–1169.

[151] Sheen, A. (2014). The real product market impact of mergers. *The Journal of Finance*, 69(6), 2651–2688.

[152] Simons, J. (2017). Ibm, a pioneer of remote work, calls workers back to the office.

[153] Sjaastad, L. A. (1962). The costs and returns of human migration. *Journal of political Economy*, 70(5, Part 2), 80–93.

[154] Srikanth, K. & Puranam, P. (2014). The firm as a coordination system: Evidence from software services offshoring. *Organization Science*, 25(4), 1253–1271.

[155] Starr, E., Ganco, M., & Campbell, B. A. (2018). Strategic human capital management in the context of cross-industry and within-industry mobility frictions. *Strategic Management Journal*, 39(8), 2226–2254.

[156] Stern, A. D. (2017). Innovation under regulatory uncertainty: evidence from medical technology. *Journal of public economics*, 145, 181–200.

[157] Stern, S. (2004). Do scientists pay to be scientists? *Management science*, 50(6), 835–853.

[158] Susskind, R. & Susskind, D. (2016). Technology will replace many doctors, lawyers, and other professionals. *Harvard Business Review*, 11.

[159] Swisher, K. (2013). "physically together": Here's the internal yahoo no-work-from-home memo for remote workers and maybe more.

[160] Team, G. (2019). 11 companies that let you work remotely.

[161] Tecco, H. (2016). 2016 year end funding report: a reality check for digital health. Online.

[162] Tecco, H. & Zweig, M. (2017). Digital health funding 2017 midyear review: A record breaking first half. Online.

[163] Teece, D. J. (1982). Towards an economic theory of the multiproduct firm. *Journal of Economic Behavior & Organization*, 3(1), 39–63.

[164] Teece, D. J. (1986). Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research policy*, 15(6), 285–305.

[165] Thompson, J. (1967). 1967 organizations in action: Social science bases of administrative theory. new york: Mcgraw-hill.

[166] Truog, R. D. (1992). Triage in the icu. *Hastings Center Report*, 22(3), 13–17.

[167] Tuckson, R. V., Edmunds, M., & Hodgkins, M. L. (2017). Telehealth. *New England Journal of Medicine*, 377(16), 1585–1592.

[168] Wachter, R. M. & Howell, M. D. (2018). Resolving the Productivity Paradox of Health Information Technology: A Time for Optimism. *JAMA*, 320(1), 25–26.

[169] Wagner, J. & Watch, D. (2017). *Innovation Spaces: The New Design of Work*.

[170] Wiedner, R. & Mantere, S. (2019). Cutting the cord: Mutual respect, organizational autonomy, and independence in organizational separation processes. *Administrative Science Quarterly*, 64(3), 659–693.

[171] Wizemann, T. et al. (2010). *Public health effectiveness of the FDA 510 (k) clearance process: balancing patient safety and innovation: workshop report*. National Academies Press.

[172] Young, L. B., Chan, P. S., Lu, X., Nallamothu, B. K., Sasson, C., & Cram, P. M. (2011). Impact of telemedicine intensive care unit coverage on patient outcomes: a systematic review and meta-analysis. *Archives of internal medicine*, 171(6), 498–506.