UNIVERSITY OF CALIFORNIA SAN DIEGO

Regulating Health Care Markets: Impacts on Innovation, Market Structure, and Product Quality

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Economics

by

Parker Rogers

Committee in charge:

Professor Jeffrey Clemens, Co-Chair
Professor Joshua Graff Zivin, Co-Chair
Professor Gordon Dahl
Professor Craig McIntosh
Professor Paul Niehaus

2023

The Dissertation of Parker Rogers is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2023

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am deeply grateful to Professor Jeffrey Clemens for his unwavering mentorship and support throughout my graduate journey. His willingness to embark on a project with me as a first-year graduate student, his patient guidance in teaching me countless academic soft skills, and his introduction to the exciting field of innovation economics have been invaluable. I would also like to thank Professor Paul Niehaus for his decisive guidance on writing, idea generation, and personal process, Professor Craig McIntosh for his creative insights and thoughtful feedback that helped steer my research in rewarding directions, and Professors Joshua Graff Zivin and Gordon Dahl for their valuable guidance and help throughout the dissertation writing process and the job market.

Chapter 1, in full, is currently being prepared for submission for publication of the material. The dissertation author was the sole investigator and author of this paper.

Chapter 2, in full, is a conditionally accepted publication at the *Review of Economics and Statistics*. This project was co-authored with Jeffrey Clemens. The dissertation author was a primary investigator and author of this paper.

Chapter 3 is currently being prepared for submission for publication of the material. This project is co-authored with Yunan Ji. The dissertation author was a primary investigator and author of this paper.

VITA

2017	Bachelor of Sciences, Brigham Young University
2018-2020	Teaching Assistant, Department of Economics University of California San Diego
2018-2019	Research Assistant, Dr. Jeffrey Clemens, University of California San Diego
2020-2022	Research Assistant, Dr. Shamim Nemati, University of California San Diego
2022-2023	Lead Data Scientist, Center for Health Innovation, University of California San Diego
2023	Doctor of Philosophy, University of California San Diego

PUBLICATIONS

"Optimizing the Implementation of Clinical Predictive Models to Minimize National Costs: Sepsis Case Study" (with Aaron E. Boussina, Supreeth P. Shashikumar, Gabriel Wardi, Christopher A. Longhurst, and Shamim Nemati), *Journal of Medical Internet Research*, 25:e43486, 2023.

FIELDS OF STUDY

Major Field: Economics

ABSTRACT OF THE DISSERTATION

Regulating Health Care Markets: Impacts on Innovation, Market Structure, and Product Quality

by

Parker Rogers

Doctor of Philosophy in Economics

University of California San Diego, 2023

Professor Jeffrey Clemens, Co-Chair
Professor Joshua Graff Zivin, Co-Chair

Chapter one examines how FDA regulation affects innovation and market concentration. I examine this question by exploiting FDA deregulation events that affected certain medical device types but not others. I collect comprehensive data on medical device innovation, device safety, firm entry, prices, and regulatory changes and enhance these data using text analysis methods. My analysis of these data reveals three key findings. First, deregulation events significantly increased the quantity and quality of new technologies in affected medical device types relative to controls. These increases are particularly strong among small and inexperienced firms. Second, these events increased firm entry and reduced prices for medical procedures that

utilize affected medical device types. Finally, rates of serious injuries and deaths attributable to defective devices did not significantly increase following these events. Interestingly, deregulating certain device types was associated with reduced adverse event rates, possibly due to firms increasing their emphasis on product safety in response to increased litigation risk.

In **chapter two,** we analyze wartime prosthetic device patents to investigate how demand shocks and procurement environments can shape medical innovation. We use machine learning tools to develop new data describing the aspects of medical and mechanical innovations that are emphasized in patent documents. Our analysis of historical patents yields three primary facts. First, we find that the U.S. Civil War and World War I led to substantial increases in the quantity of prosthetic device patenting relative to patenting in other medical and mechanical technology classes. Second, we find that the Civil War led inventors to increase their focus on reducing cost, while World War I did not. The Civil War era emphasis on cost is consistent with a role for that period's cost-conscious procurement model. Third, we find that inventors emphasized dimensions of product quality (e.g., a prosthetic limb's comfort or facilitation of employment) that aligned with differences in buyers' preferences across wars. We conclude that procurement environments can significantly shape the dimensions of the technical frontier with which inventors engage.

In **chapter three**, we study how government price reforms affect innovation, market structure, and product quality within the health care sector. We exploit a Medicare payment reform that reduced expenditures on certain types of durable medical equipment (DME) by 66% while leaving other types unaffected. We find that manufacturers filed 29% fewer patents and introduced 22% fewer new models in DME types affected by the price reform relative to those that were unaffected. Additionally, patents filed after the price reform increasingly focused on "process" rather than "product" innovation, consistent with increased market demand for lower-cost products. The market structure was also affected, with 25% fewer manufacturers entering affected product markets and a 65% increase in outsourcing to foreign companies. The shift towards cost-cutting, both in patenting and supply chain restructuring, was associated with

increased device repair rates among Medicare beneficiaries and reported adverse events. Firms that outsourced to foreign manufacturers experienced the highest increase in adverse events. While the Medicare price reform generated substantial savings, these gains were dampened by the adverse effects on innovation, market structure, and product quality in the long run. Our findings highlight the importance of considering long-run impacts when designing policy reforms.

Chapter 1

Regulating the Innovators: Approval Costs and Innovation in Medical Technologies

While new technologies can improve consumer well-being, they can also cause harm. One way to mitigate harm is through regulation that requires innovators to demonstrate the safety of their products before commercialization, an ex-ante approach taken by the U.S. Food & Drug Administration (FDA). Another strategy relies on the threat of ex-post litigation to deter harm. A decades-long debate considers these alternatives. Critics of regulation claim that it chills innovation and market competition by raising entry costs (Peltzman 1973) and that litigation is more efficient (Coase 1960). Proponents counter that regulation increases public confidence in products marketed by lesser-known firms, encouraging entry and innovation (Carpenter et al. 2010). Clear evidence testing these claims is important given the $2.8 trillion market size of FDA-regulated products alone (FDA 2020b).

I advance this debate by measuring the impact of FDA regulation on innovation and market structure. To study this relationship, I first consider a *less stringent* regime by examining deregulation events that moved, or "down-classified," certain higher-risk medical device types, like spinal implants, from stringent (Class III) to moderate (Class II) testing requirements. Second, I consider the *litigation* alternative by analyzing events that moved lower-risk device types, like ventilator tubing, from moderate (Class II) to no testing requirements (Class I),

exposing innovators to more litigation. Examining both of these types of events is valuable for two reasons: First, it allows me to measure the impact of strict FDA regulation (i.e., clinical trials) relative to the existing alternative policies. Second, it enables me to estimate local average treatment effects among deregulated devices at different points in the distribution of safety risk (low-to-moderate risk versus moderate-to-high risk devices).

I infer the causal effect of these events by comparing affected device types to a variety of possible control groups. These groups include device types matched on pre-event means, intuitively similar devices, later-deregulated devices, and a broad set of unaffected devices. I find that my results are stable across these control groups. Further, comparing deregulated device types to control groups reveals no divergent pre-existing trends in the outcomes of interest, consistent with the ”unpredictable” characterization of these events by device manufacturers (Makower et al. 2010, Powell 2018).

An important contribution of this paper is the assembly of novel data on the tradeoffs of FDA regulation. Regulation affects many factors, and data on these factors are siloed, unorganized, and unconnected to medical device types, limiting research on this topic. I use a combination of programmatic online text extraction, text analysis algorithms, and hand linkages to create, merge, and harmonize the required data. When unified, these data comprehensively detail the effects of medical device regulation by device type. These data include all FDA device type regulation changes over the last 40 years and multiple corroborative measures of device innovation, innovator characteristics, innovation quality, market structure, prices, and device safety.

My analysis of these data shows that down-classification events increase the quantity and quality of new technologies. After moving from Class III (high regulation) to II (moderate), device types exhibited a 200% increase in patenting and FDA submission rates relative to control groups. Patents filed after these events were also of significantly higher quality, as measured by a 200% increase in received citations and market valuations. These effects do not

spill over into similar device types.[1] For Class II to I deregulations, the rate of patent filings increased by 50%, though insignificantly, and the quality of patent filings exhibited a significant 10-fold improvement, suggesting that litigation better promotes innovation. There is substantial heterogeneity in how firms respond to deregulation as increases in innovation are strongest among smaller firms and those with the least regulatory experience—the same subset of firms found most likely to produce groundbreaking innovation (Wu et al. 2019).

Second, these events led to significant changes in market structure. Class III to II events generated a ten-fold increase in new entry (i.e., firms with no approved devices) and a four-fold increase in incumbent entry (i.e., firms with approved devices of another type) into treated device types. Increased competition impacted health care prices: Using claims data from a university hospital system, I find that these events were associated with a significant 40% drop in the prices of medical procedures that use deregulated device types relative to controls.[2] Class II to I events led to a significant 200% increase in new entry into treated device types, with no effect on incumbent entry, suggesting that litigation obstructs new firm entry less than regulation.

Down-classification yields considerable benefits, as the proponents of deregulation would predict, but what of product safety? Perhaps counterintuitively, I find that deregulation can improve product safety by exposing firms to more litigation. Despite some adverse event rates increasing after Class III to II events (albeit insignificantly), Class II to I events are associated with significantly *lower* adverse event rates.[3] My analysis of patent texts also reveals that inventors focus more on product safety after deregulation. These results suggest that litigation encourages product safety more than regulation: Instead of meeting Class II requirements, which the National Institute of Medicine deems as insufficient for product safety (IOM, 2011), inventors must decrease the likelihood that their products injure consumers to prevent litigation. I identify

[1] These localized effects could be explained by extreme specialization: many device inventions originate from practicing physicians or researchers acting within their medical specialty (NIM, 2010).

[2] This price drop could even be mechanically driven by the 68% reduction in testing costs from these events (Makower et al. 2010), which may reduce markups intended to recover regulatory costs.

[3] "Rates" are counts per device type-year. I do not normalize by utilization, but I show that this normalization would likely strengthen my adverse event estimates as deregulation increases utilization.

litigation as a mechanism using variation in firms' exposure to litigation after deregulation: Smaller firms expect less liability as they can use bankruptcy to avoid liability that exceeds their assets (Shavell 1986). I find that safety improvements are strongest at larger firms for which a larger share of liability is unavoidable.

A back-of-the-envelope calculation suggests that the benefits of these events outweigh the costs. Accounting for the cost of adverse events and the value of increased innovation and decreased health care prices, the unmeasured costs of Class III to II events would need to be larger than the measured costs to justify Class III regulation. For Class II to I events, there are virtually no measurable costs of down-classifications as adverse events *decline*. By contrast, the benefit of these events amounts to more than $22 million a year per device type. Although these benefits are based on local average treatment effects among deregulated device types, I find evidence that these benefits may generalize to current Class II device types: More dangerous, marginal deregulated device types (according to the FDA's decision rule) exhibit the largest decreases in adverse events. If this relationship holds, the yearly forgone benefits could amount to as much as $55 billion across 2,500 current Class II device types, or nearly 32% of the annual value of medical devices consumed.

I build a model that illustrates the range of possible consequences of deregulation. The model incorporates the central concerns of medical device innovators. First, regulation imposes approval delays, but firms shorten delays as they gain more experience navigating approval requirements through "learning by doing" (Arrow 1971). Firms also face financing costs if approval costs exceed their assets (Buera and Shin 2013, Moll 2014). Lastly, when regulations are lifted (Class I), firms are exposed to more litigation from product design flaws, but small firms are exposed to less liability due to bankruptcy. This characterization of the firm's decision shapes the effects of deregulation: Deregulation can improve product safety and disproportionately benefit small firms and those with less regulatory experience.

My findings contribute to several literatures. First, I add to the growing literature on

the effects of public policy on medical innovation.[4] Despite the significant size of the medical device market, valued at around $500 billion and projected to reach nearly $1 trillion by 2030 (Stewart 2022), there is a lack of evidence on the impact of regulation on innovation in this sector. Previous studies by Stern (2017) and Grennan and Town (2020a) use cross-group comparisons to suggest that regulations affect investments in Class III cardiovascular technologies, but they do not address the broader impact of FDA regulation on innovation. My research fills this gap by examining the relationship between regulation and innovation using quasi-exogenous regulatory shocks across a range of device types and at several levels of regulatory stringency. Additionally, my study evaluates the safety benefits of device regulation, which has received little attention.

I also add to a longstanding literature on the tradeoffs between regulation and litigation.[5] Regulation, a preventive strategy, sets a lower bar on product safety, whereas litigation, a deterrence strategy, punishes those who violate standards through the courts (Kessler 2010). A study by Philipson et al. (2010) finds that regulation and litigation together are less efficient than regulation alone, but did not examine which approach is more efficient on its own. I find that litigation can more effectively prevent adverse events while promoting innovation.

Lastly, my findings relate to the literature on endogenous growth (Romer 1990). Recent work shows that labor regulations can influence innovation, the key determinant of economic growth (Acharya et al. 2014; 2013, Aghion et al. 2019). Other work shows that regulation can reduce market competition, creating long-run inefficiencies (Buettner 2006, Aghion et al. 2009; 2005, Djankov et al. 2006, Hahn and Hird 1991). I add to this literature by showing that product regulation reduces innovation and market competition. My findings, however, depart from the common presupposition that regulatory knowledge flows smoothly across firms: Deregulation disproportionately benefits firms with less regulatory experience, suggesting that regulatory proficiency stays with the firms that acquire it (akin to Azoulay et al. (2011)). These frictions amplify the costs of regulation and may advantage experienced multiproduct firms across a wide

[4] See Mulligan (2021), Grennan and Town (2020a), Clemens and Rogers (2020), Stern (2017), Budish et al. (2015), Acemoglu and Linn (2004b), Finkelstein (2004b).

[5] See Coase (1960), Ehrlich and Posner (1974), Kolstad et al. (1990), Glaeser et al. (2001), Shavell (1986; 2018).

range of regulated products.

The tension between regulation and litigation affects a variety of everyday products. These products range from those regulated similarly to Class III or II medical devices, like pharmaceuticals and genetically modified foods, to broader categories like aircraft, automobiles, pesticides, and over 15,000 consumer products regulated by the Consumer Product Safety Commission (Schwartz and Appel 2020, Schauzu 2000, Pisani 2011). After regulatory compliance, these products receive at least some protection from litigation, making my findings particularly relevant.

This paper is organized as follows. Section 1.1 provides background on the FDA regulatory process, section 1.2 provides the conceptual framework, section 3.2 discusses my data, section 3.3 describes my empirical strategy, section 1.5 presents my empirical results, section 1.6 presents a back-of-the-envelope welfare calculation, and section 3.5 concludes.

1.1 Background

This section describes the structure and legal consequences of FDA medical device regulations. Medical devices include products like COVID-19 tests, pacemakers, X-ray machines, and spinal implants.

1.1.1 Enactment of Medical Device Regulations

In 1976, the Medical Device Amendments (MDA) expanded the FDA's oversight to include medical devices. According to these new laws, medical devices were grouped into generic types to allow targeted regulation. "Daily-wear soft contact lenses," for example, is a device type regulated differently than "extended-wear soft contact lenses." The policy variation I study occurs at the level of these generic device types, and I refer to them as "device types."

Device types are organized into a three-tier risk classification system. Manufacturers of Class I low-risk devices must register their facility with the FDA, which carries a small fee and takes less than one month to process. The FDA requires Class II, moderate risk device

manufacturers to file a "510(k)" to prove their device is similar to an already marketed device.[6] This process of proving "substantial equivalence" has been criticized by many, including the National Institute of Medicine, as being insufficient for establishing safety (IOM, 2011) while imposing substantial costs. The 510(k) process, on average, costs firms $24 million (Makower et al. 2010) and delays commercialization by ten months. Class III, high-risk device manufacturers must conduct clinical trials via the "premarket approval" (PMA) process to ensure their new device is safe and effective before commercialization. The PMA process is much longer than the 510(k) process and costs, on average, $75 million (Makower et al. 2010). The average costs of these different levels of regulation are shown in figure 1.1. Appendix A5.3 provides more details.

1.1.2 Deregulation of Medical Device Types

The FDA can lower the class of a medical device type after observing the safety outcomes of marketed devices. Without any safety information, the FDA regulates new, markedly novel devices in Class III to ensure safety in the presence of unknown risks.[7] Surveillance data from marketed devices clarify these risks and inform the FDA's choice to move a device type into Class II, or "down-classify" (see figure 1.1).[8] These events are described by manufacturers as "unpredictable," suggesting the difficulty of anticipating such policy changes (Powell 2018). My empirical analysis supports this assessment as I do not find evidence of divergent pre-existing trends when comparing down-classified device types to control groups.

By contrast, the Class II to I down-classifications I study are systematic. In 1995, the FDA scored all Class II devices based on average yearly adverse event counts and down-classified those that fell below a previously unknown threshold (FDA 1995). Although this policy change

[6] Manufacturers must also follow best-practice protocols (called "special controls").

[7] In 1997, the FDA began allowing manufacturers of markedly novel devices to petition for a direct Class II or I classification under the "De Novo" process by showing that best practices assure the safety and efficacy of their device. However, all the device types I consider existed before 1997 and thus were either automatically or intentionally classified into Class III.

[8] Additionally, manufacturers can file a petition for down-classification, bringing the FDA's attention to particular device types for further investigation. My analysis, however, focuses on down-classification events explicitly enacted by the FDA's initiative (rather than a petition).

appears to justify using a regression discontinuity design, the sparseness of device types at the threshold does not permit this approach. Instead, a series of unaffected Class I device types that would have received similar scores as treated device types serve as appropriate controls. These types include previously deregulated and always Class I device types. Importantly, scores were not contingent on potential changes in adverse events or trends.[9] My event-study results reaffirm these assessments.

It is worth noting that deregulation only occurs in established medical device types. Thus, rather than measuring the effect of regulation on radical innovation, this paper measures how regulation affects the development and improvement of existing medical device types. Improving medical devices may require fundamental scientific advances and bring substantial health benefits through increased efficacy or reduced side effects and adverse events.

1.1.3 Regulation versus Litigation: Federal Preemption

In the US, medical device firms incur damages from tort claims amounting to as much as 3.8% of annual revenues (Fuhr et al. 2018). Galasso and Luo (2018) show that this liability risk chills innovation and can bankrupt smaller firms. Compared to Europe, the US is particularly litigious, with class-action lawsuits, high punitive damage payouts, and few damage caps (Guendling 2016). These conditions make liability risk a powerful incentive for ensuring the safety of products marketed in the US.

However, FDA approval shields medical device manufacturers from product liability, creating a stark tradeoff between regulation and litigation. This protection, called "federal preemption," is upheld by Riegel v. Medtronic Inc. (2008), a supreme court case establishing that Class III device approvals bar legal claims against device manufacturers. The Class II devices I analyze are also often protected from litigation as they are FDA-approved and subject to "special controls" requirements that ensure safety and efficacy (Costello and Pham 2016).[10] Class I

[9] See appendix A5.1 for more details and for an example of Class III to II events.

[10] The recent court case Kelsey v. Alcon Laboratories Inc. (2019) offers an example of a Class II approval barring legal claims through preemption. In this case, the plaintiff claimed that Alcon's contact lens disinfectant did not

devices are not FDA-approved, exposing manufacturers to litigation.

1.2 Conceptual Framework

In this section, I model R&D as a two-stage process: development and commercialization. First, firms invent and patent a new product, improve its safety profile, and raise capital to cover commercialization. Second, firms bring their products to market by attaining regulatory approval, forming distribution networks, etc. The model builds on that of Budish et al. (2015), who formalize the impacts of commercialization lags on innovation. For comparability, I follow their notation closely wherever possible. I introduce into their framework two alternative policy regimes (i.e., regulation and litigation), which include differences in commercialization lags, liability risk, and financing costs.

The model's purpose is to illustrate the range of possible consequences of deregulation, to connect these to underlying fundamentals, and, in particular, to relate these effects to firm traits. In turn, the insights from this model will be helpful for interpreting my empirical results. My model considers the medical device industry, though its implications may apply to other regulated products.

1.2.1 Model Preliminaries

Undirected R&D yields stochastic inventions to a representative, profit-maximizing firm. Upon realizing the new technology, the firm decides if it will allocate capital for directed R&D to (i) improve the product's safety profile during the development phase and (ii) commercialize the invention. The firm makes this decision in one of two environments: regulation "R" or litigation "L." The model is characterized by the following parameters:

prevent a severe eye infection due to a product flaw. However, the disinfectant was approved as a Class II regulated device and was subject to special controls. The district court handling the case deemed that the FDA's approval adequately tested the product's safety, preventing legal liability. This is just one of many recent instances where Class II medical devices have been protected from design defect claims through preemption. Other examples include cases involving latex gloves, contact lenses, tampons, condoms, angioplasty catheters, wound dressing, tissue adhesive with wound closure device, a hemorrhoid prevention pressure wedge, and electrical stimulation devices (Munford 2018).

Timing Parameters.—The year a firm realizes and develops an invention is given by t_{invent}, which I normalize to zero. The years it takes to commercialize the product is $t_{comm,f}$. In the medical device industry, FDA approval plays a key role in delaying commercialization (Makower et al. 2010, Pietzsch et al. 2012).[11] Thus, for concreteness, think of $t_{comm,f}$ as the approval delay. Under litigation L, there are no approval delays (i.e., $t_{comm,f,L} = 0$). In the regulated environment, approval delays are positive but decrease with regulatory experience (Olson 1997, Carpenter 2004b, Makower et al. 2010).[12] Following Arrow (1971), I model this relationship by equating the present delay $t_{comm,f}$ to the learning curve $\beta\ T_f^{-\gamma}$, where T_f is prior experience, β is the delay with no prior experience (i.e., $T = 1$) and $\gamma > 0$. Delay costs are given by $\chi t_{comm,f}$, where χ is the yearly cost of approval delays.[13]

Financing Costs.—Smaller firms must raise external capital to cover the costs of development and commercialization at time t_{invent}.[14] Fundraising can be difficult: 56% of small medical device firms claim funding as a central challenge (Emergo 2019). Following Stein (2003), I capture these financing frictions by assuming deadweight costs given by $C(e_f)$, where $C(\ )$ is an increasing convex function of external funds e_f (similar to the R&D model of Stern (2017)). External funds e_f are equal to the difference between the non-financing costs and internal capital K_f. I omit other costs of commercialization for simplicity.

Regulated and Deregulated Effective Lives.—A successfully commercialized product becomes less relevant over time. For expositional ease, I describe the neoclassical risk-adjusted discount factor of the R&D project as δ, which includes obsolescence and commercialization

[11] Approval delays in other areas of health care, like delays in securing medical procedure reimbursement codes, have also been shown to play a key role in innovation (Dranove et al. 2022).

[12] Two factors may explain this pattern, both of which are driven by the complexity of the regulatory process. First, inexperienced firms report difficulty benefiting from hired regulatory experts and must instead learn the process independently (Y Combinator 2016). From the regulator's perspective, having prior experience with a firm reduces the uncertainty about the quality of its products, which may merit shorter review times (Olson 1997, Carpenter 2004b).

[13] Makower et al. (2010) find an average monthly cost of $1.3 million for Class III approval delays (e.g., clinical trial costs, etc.). I assume $t_{comm,f}$ and several other parameters below are deterministic for simplicity.

[14] For simplicity, I assume firms finance their project instantaneously. Although fundraising could prolong commercialization delays, removing this assumption does not change my theoretical results.

risk.[15,16] Firms enjoy longer or shorter effective product lives depending on the regulatory environment. Under regulation, I define an invention's *Regulated Effective Life (REL)* as the expected years it will be commercialized and non-obsolete in present value terms as discounted by the regulated firm. The effective life of the regulated product begins at time $t_{comm,f}$, yielding an effective life of $REL_f = \sum_{t_{comm,f}}^{\infty} \delta^t = \delta^{t_{comm,f}}/(1-\delta)$. By contrast, in a deregulated environment N, I define an invention's *Effective Life (EL)* similar to *REL*, except the lifespan of the product starts at t_{invent}, given by $EL = \sum_{t=0}^{\infty} \delta^t = 1/(1-\delta)$. Notice that $REL_f < EL$ by definition, as regulated profit flows are delayed.

Expected Damages and Safety Effort Costs.—Borrowing from Shavell (1986) and Boomhower (2019), if a firm chooses to commercialize its product, it exerts x_f effort to improve product safety, costing ψ per unit, at t_{invent}.[17] Under litigation L, a commercialized product generates stochastic adverse events that yield $\phi(x_f;\vec{Z})$ legal damages per year, a random variable with expected value $D(x_f;\vec{Z})$ and vector $\vec{Z}$ containing other factors that influence damages in expectation (e.g., firm seizable assets K_f, the litigation environment, damage caps). The expected damages function $D(\,)$ is a positive decreasing convex function of safety effort x_f. The firm exerts effort to maximize the returns to commercialization by equating the marginal cost of effort $\psi + C_x(\psi x_f^* - K_f)$ to the present value of its marginal benefits $-EL \cdot D'(x_f^*;\vec{Z})$ (i.e., marginal abatement of expected damages). By contrast, under regulation R, the firm is exposed to no legal damages due to federal preemption. Thus, firms exert the mandated level of safety effort $\underline{x}$, as any further effort yields no return.

Profits.—If the product is successfully commercialized and non-obsolete, it generates

[15] A product may also face a probability of successful commercialization p, which may be appropriately modeled as a function of safety effort; however, the FDA approves 80%–90% of all medical device submissions (GAO, 2009). Thus, for simplicity, I assume that approval is certain given a firm achieves the mandated safety effort, and I abstract away from other non-approval-related commercialization uncertainty. Including product denial and commercialization risks does not meaningfully change my theoretical insights.

[16] Although obsolescence risk is more appropriately modeled as endogenous to R&D investments, I follow the patent literature and take it as exogenous (Budish et al. 2015).

[17] For simplicity, I assume firms exert safety effort instantaneously. Alternatively, safety efforts could prolong commercialization delays. Modeling such delays, however, would not change the model implications.

profits π per year for the innovating firm. Although regulation can affect profits by altering market structure, I do not model this relationship, focusing instead on motivating my firm composition and product safety results. Thus, for simplicity, I assume that deregulation increases the aggregate level of R&D, consistent with my empirical findings, which implies that deregulation does not cut profits enough to outweigh declines in commercialization costs.[18] I assume only expert regulators can perceive safety effort (i.e., asymmetric information); hence, safety effort does not affect profits once a product is approved.

1.2.2 Characterization of the Investment Decision

In the regulated environment R, firm f expects to receive profits from commercializing a device for REL_f years. The firm will develop and commercialize its invention if and only if these expected profits exceed the combined delay, safety effort, and financing costs:[19]

$$\text{Regulated Firm Invests} \iff \underbrace{REL_f}_{\substack{\text{Regulated}\\ \text{effective life}}} \cdot \underbrace{\pi_R}_{\text{Profits}} \geq \underbrace{\chi t_{comm,f}}_{\text{Delay costs}} + \underbrace{\psi \underline{x}}_{\substack{\text{Mandated safety}\\ \text{effort costs}}} + \underbrace{C(e_{f,R})}_{\substack{\text{Financing}\\ \text{costs}}}. \quad (1.2.1)$$

The amount of external capital $e_{f,R}$ needed to finance the project is given by the difference between the non-financing commercialization costs and the firm's internal capital K_f (i.e., $e_{f,R} = \chi t_{comm,f} + \psi \underline{x} - K_f$ if $e_{f,R} \geq 0$, and 0 otherwise).

In the litigation environment L, firm f will choose to commercialize if and only if the net expected profits (less expected damages) are greater than the combined safety effort and

[18] Note that this assumption also places an upper bound on the value of legal damages and safety effort costs after deregulation.

[19] Notice the implicit assumption that firms do not consider the future benefits of regulatory experience (i.e., learning by doing) in their investment decisions. This assumption is consistent with a large literature documenting that managers maximize short-term rather than long-term firm value (Budish et al. 2015).

financing costs:[20]

$$\text{Deregulated Firm Invests} \iff \underbrace{EL}_{\substack{\text{Effective}\\ \text{life}}} \cdot [\underbrace{\pi_L}_{\text{Profits}} - \underbrace{D(x_f^*;\vec{Z})}_{\substack{\text{Expected}\\ \text{damages}}}] \geq \underbrace{\psi x_f^*}_{\substack{\text{Optimal safety}\\ \text{effort costs}}} + \underbrace{C(e_{f,L})}_{\substack{\text{Financing}\\ \text{costs}}}. \quad (1.2.2)$$

The amount of external capital $e_{f,L}$ needed to finance the project is given by the difference between safety effort costs $\psi x_f^*$and the firm's internal capital K_f.

Notice the key differences between the investment incentives in environments R and L: firms that commercialize in L (i) expect legal damages, (ii) choose and pay for an optimal level of safety effort, (iii) enjoy a longer effective life of their products, and (iv) do not incur delay costs.[21]

1.2.3 Distortions from Regulation

I focus on model implications related to distortions in firm participation and safety efforts resulting from regulation. Throughout, I assume that deregulation increases the level of R&D activity. This assumption is supported by my empirical results and allows me to more clearly motivate the less intuitive results I find in my analysis.

First, I explore how deregulation can improve product safety. If mandated levels of safety effort are low enough, deregulation can improve safety by increasing the net incentives for safety improvements. I state this formally as follows:

Proposition 1 *(Deregulation can increase firm safety efforts) If the marginal cost of regulated effort is less than the ex-post marginal benefit of that effort (i.e., $\psi + C_x(\underline{x}) < -EL \cdot D'(\underline{x})$), then deregulation will increase firm safety effort.*

Figure 1.2 helps clarify the necessary conditions for proposition 1. The figure shows that the ex-ante-mandated safety effort is sufficiently low, leading the deregulated firm to exert more

[20]Note that financing frictions do not affect the payment of damages since they can be financed with profits (i.e., in expectation, damages will always be less than profits if a firm chooses to commercialize).

[21]Profits and financing costs also differ across these environments; however, the direction of the difference is ambiguous (e.g., if expected damages are large, safety effort costs could increase financing costs).

effort. This proposition implies that ineffective regulations could make products less safe. I show in section 1.5 that Class II regulations may lead to such an outcome. These insights, however, may be specific to the litigious US environment. For example, if a country aggressively caps damages (represented in $\vec{Z}$), firms would face lower expected damages, and safety effort could drop relative to regulated levels.

Another factor influencing a firm's expected damages is the value of its seizable assets. Following insights on the "judgment proof problem" (Shavell 1986), when damages exceed the value of a firm's seizable assets, the difference can be discharged through bankruptcy. This option protects small firms from worst-case damages, lowering expected damages and the marginal benefit of exerting safety effort. Thus, if deregulation increases safety efforts, it will do so most for large firms. I state this as follows (and more formally in appendix A1):

Proposition 2 *(Deregulation introduces bankruptcy distortion) Assume firm A has fewer assets than firm B (i.e., $K_A < K_B$) and has too few assets to cover its worst-case damages. Firms A and B are otherwise identical. If deregulation increases firms' safety effort (see Proposition 1), then firm B will increase its safety efforts the most.*

The next distortion I detail arises from regulatory complexity (i.e., the delays from complex regulatory requirements). Complexity distorts the composition of firms that commercialize as inexperienced firms reap lower returns from commercialization. Deregulation removes these distortions and disproportionately increases the returns to commercialization for inexperienced firms. To formalize this claim, I present the following proposition:[22]

Proposition 3 *(Deregulation disproportionately benefits inexperienced firms) If firm A has less regulatory experience than firm B (i.e., $T_A < T_B$; all else equal), then deregulation increases the returns to commercialization most for firm A.*

An example helps illustrate the potentially dramatic implications of proposition 3. Consider firm A has no prior experience, and firm B has one previously commercialized project that

[22] Proofs are presented in appendix A2.

was delayed for two years. Consistent with the values of the learning curve parameters γ and β estimated in section A3.1, firm A must wait out a two-year delay. By contrast, firm B waits out a one-year delay, incurring 50% lower delay costs than firm A and enjoying a longer effective life of its product. Although deregulation removes delay-related costs for both firms, the increase in returns to commercialization is at least twice as large for firm A.

Lastly, I discuss distortions that arise from financing frictions and regulation. Small firms incur deadweight costs when raising capital to commercialize their products (Gaglani 2014, Emergo 2019). Deregulation can decrease commercialization costs and financing costs, especially for small firms. I state this claim formally as follows:

Proposition 4 *(Deregulation can disproportionately benefit smaller firms) Assume firm A is smaller than firm B and has non-zero financing costs when regulated (i.e., $K_A < K_B$ and $K_A < \chi t_{comm,A} + \psi \underline{x}$). Firms A and B are otherwise identical. If deregulation does not increase financing costs for firm A (i.e., $\psi x_A^* < \chi t_{comm,A} + \psi \underline{x}$), then deregulation increases commercialization returns most for firm A.*

However, deregulation could lead to lower returns to commercialization for small firms if financing costs increase after deregulation. For example, if deregulation induces enough additional safety effort costs to outweigh the decrease in approval delay costs, financing costs could increase for smaller firms. By contrast, if the assumptions hold, Proposition 2 will amplify Proposition 4 as small firms face lower expected damages and lower safety effort costs after deregulation and, thus, even lower financing costs.

1.3 Data

To conduct my empirical analysis, I compile data from eight sources to provide an expansive view of the costs and benefits of medical device regulations. Summary statistics for these data are provided in table 1.1 and a data catalog is presented in figure A1.

FDA Device Submissions (PMA and 510(k) Databases). The primary dataset used in this

study is derived from FDA administrative data on the universe of medical devices submitted for FDA approval. These data combine the FDA's PMA and 510(k) databases to cover both Class III and II devices. Submissions include the submitting company name, device brand name, medical device type, and submission and approval dates. I use fuzzy matching to form three measures of market dynamics and innovation. First, I measure "new entry" by identifying firms submitting approval documents for the first time. Second, I also form a measure of "incumbent entry," by locating firms that have filed prior approval documents but are starting to submit for approval in a given device type. Third, I isolate the first occurrence of unique device brand names within a device type to form the "unique devices submitted" measure. These variables are aggregated to the device type-year level. To measure each firm's regulatory proficiency, I calculate the total approval delays (in days) the submitting firm has experienced up to the given point in time.

FDA Deregulation Events. To provide a comprehensive analysis of FDA deregulation events, I collect all down-classifications from 1980 to 2015. For Class III to II events, I also indicate whether the event was motivated by the FDA's "own initiative" or by an industry petition. This distinction is empirically important. Figure A2 shows that device types that experience a petitioned down-classification exhibit divergent pre-trends in patenting rates in the five years before the event. The Class III to II events I consider are those enacted by the FDA's own initiative and for which down-classified device types experienced at least one PMA document submission beforehand.[23] For Class II to I events, I consider affected device types that experienced at least one 510(k) document submission beforehand.

FDA Adverse Event Reports (MAUDE). The FDA's Manufacturer and User Facility Device Experience (MAUDE) database contains adverse event reports related to medical devices. Using this data, I create measures of device safety using reported deaths, hospitalizations, and life-threatening events for each device type from 1992–2019. I follow Ensign and Cohen (2017) to account for data and coding idiosyncrasies in the MAUDE data. Adverse events are aggregated to the device-type-year level. Adverse event rates (e.g., deaths per year) of down-classified

[23]Many Class III "preamendment" devices were never officially required to submit PMA documentation.

device types are similar to those of device types in the prospective class (see figure A3). For the top 300 manufacturers by adverse event volume, I hand-linked firm names listed on adverse event reports to data on firm assets. Asset totals are derived for public firms using data from CRSP/Compustat. This linkage allows heterogeneity analyses of device safety by firm size.

USPTO Patent Grants Extract. Patents offer an additional measure of innovation to support my "unique devices approved" measure. However, there is no standard dataset linking medical devices with their associated patents (similar to the "Orange Book" data for drugs). To address this, I follow a three-step procedure to create a patent-based measure of innovation for each device type. First, I compile a list of keywords from each FDA device type description. Second, I use a computer program to collect all patents granted by the USPTO that contain those keywords in their text. Third, I calculate the annual number of patents filed within each device type based on the date the patent was first filed. The resulting dataset is a panel of yearly patent counts across 5,000 FDA-defined medical device types from 1976 to 2019. Patents are a useful complement to FDA device data for several reasons. First, patents allow me to analyze how Class II to I events affect innovation, as I only observe my "unique devices approved" measure for Class III and II devices. For this same reason, patents also enable comparisons of effect sizes across down-classification types. Lastly, an analysis of two different measures of innovation provides corroborative evidence. In section 1.5, I show that the estimates of changes in patent filing rates and device submission rates are quite similar for Class III to II events. Appendix A4 provides more details on the patent collection process.

Patent and Patent Applicant Characteristics. I enrich the patent data with measures of innovation quality and applicant characteristics. A patent's quality is measured using the number of citations it received from other patents and its market value.[24] Patent market values (in millions USD) are derived from Kogan et al. (2017). These values are based on the increase in the patent assignee's stock price resulting from a USPTO announcement of patent issuance

[24] I omit examiner citations and set patent citations and market values to zero when no patents were filed in a given device-type-year.

and are only available for publicly traded firms. I also generate a quality-related measure of device safety using patent texts. Following a procedure used in Clemens and Rogers (2020), I calculate the annual share of patents within a device type that mention keywords related to safety.[25] This variable allows me to directly analyze how deregulation affects inventors' emphases on improving device safety, corroborating adverse event analyses. Lastly, to analyze how deregulation affects innovation from firms of different sizes, I link total firm asset holdings from the CRSP/Compustat database to patent applicants.

UCSD Health Insurance Claims Extract. Insurance claims data from UCSD Health provide information on how healthcare prices respond to deregulation. To my knowledge, no available data, including ECRI *PriceGuide*, reliably measures the direct prices that providers pay for medical devices before 2011. As another option, device prices could also be reflected in insurance claims data, provided that device costs comprise a substantial share of procedure costs. However, insurance claims databases before 2011 do not measure exact paid amounts at the procedure level, the granularity necessary for attributing costs to device usage. Thus, I acquire claims data from UC San Diego Health that detail prices at the Current Procedural Terminology (CPT) level. I then identify claims with procedures that use medical device types that were down-classified since 2006.[26] To form control groups, I collect a set of procedures that use matched control device types and randomly select 100 procedures. Together, these data contain nearly 500,000 unique patient claims from 2005–2020. I then take the average amount paid for a given procedure in a given year, forming a panel of procedure-year prices.[27]

[25]To construct a comprehensive list of keywords related to medical device safety, I use Word2Vec, an algorithm that maps text to a vector space, with proximity indicating semantic similarity. After gathering semantically similar keywords, I search patent claims to identify whether a patent contained any of the keywords of interest and calculate the fraction of patents that mention these keywords in a given device-type-year. If no patents were filed in a given year, I set the fraction of patents mentioning safety to zero (i.e., no scientific advancements in product safety). See table A10 for a list of keywords used.

[26]In total, five Class III to II down-classified medical device types fit this criterion. All Class II to I down-classifications that I analyze are outside the time coverage of the claims database.

[27]Although the average UCSDH procedure amount paid is close to the average procedure amount paid by Medicare, using only UCSDH claims data is a limitation of my study.

1.4 Empirical Strategy

My strategy for estimating the effects of deregulation includes "stacked" difference-in-differences and event-study designs. After describing each design, I underscore how I address potential issues when generating causal estimates in my context.

The first regression specification uses a staggered difference-in-differences design. I use a "stacked" regression, similar to Cengiz et al. (2019), which avoids potential biases from using staggered treatment designs in the presence of heterogeneous treatment effects within-unit over time (Goodman-Bacon 2018, de Chaisemartin and d'Haultfoeuille 2019).[28] This approach assembles event-specific panel data using each treated group $r \in \{1, \ldots, N^1\}$ and all admissible controls. Then, all event-specific panels are stacked while allowing unique time and group fixed effects for each panel. Thus, the estimating equation is given by

$$Y_{t,c,r} = \gamma_{c,r} + \gamma_{t,r} + \beta_1 1\{\text{reclass}\}_{t,c,r} + \varepsilon_{t,c,r}. \tag{1.4.1}$$

In equation 1.4.1, c denotes the medical device type, t denotes time, r denotes the event, and $1\{\text{reclass}\}_{t,c,r}$ is an indicator equal to one when down-classification has occurred in device type c. The outcomes of interest are denoted by $Y_{t,c,r}$. Event-by-time fixed effects ($\gamma_{t,r}$) and event-by-device type fixed effects ($\gamma_{c,r}$) are included. The coefficient of interest, β_1, estimates the differential change in the outcome variable for treated device types relative to control device types after down-classification. I estimate equation 1.4.1 separately for Class III to II events and Class II to I events.

The number of FDA-initiated Class III to II events is relatively low ($N^1 = 13$). Thus, I follow Conley and Taber (2011), who provide a method of constructing reliable confidence intervals for differences-in-differences estimates in the presence of a small number of policy changes. This approach uses information from control group residuals to form confidence

[28] I find that my results do not change meaningfully when I consider another estimator in the heterogeneous treatment effects literature from Borusyak et al. (2021) (see tables A11, A12, and A13).

intervals.

Like all difference-in-differences designs, my specification relies on the assumption that differential trends in the outcomes of interest do not pre-date the down-classification events. To test this assumption, I estimate a stacked event-study design using OLS, given by

$$Y_{t,c,r} = \gamma_{c,r} + \gamma_{t,r} + \sum_{t \neq 0} \beta_t 1\{\text{Treated}\}_{c,r} \times 1\{\text{Years from Reclass}\}_{t,r} + \varepsilon_{t,c,r}. \tag{1.4.2}$$

In equation 1.4.2, the omitted interaction between the treated group indicators (i.e., $1\{\text{Treated}\}_{c,r}$) and the time dummy variables (i.e., $1\{\text{Years from Reclass}\}_{t,r}$) aligns with the year the event occurred. Thus, each parameter β_t represents the difference-in-differences estimate of the change in the outcome in a given period relative to that reference period. Standard errors for each β_t are calculated using Conley and Taber (2011).

Down-classification rulings are typically announced a year before enactment. Since innovators could respond to a down-classification announcement, $1\{\text{reclass}\}_{t,c}$ is equal to one for all device-type-years after an announcement occurs in device type c. However, FDA administrative data will not reflect changes until the year of enactment since firms cannot market devices under new regulations before enactment. Thus, for FDA-derived outcome data, the indicator $1\{\text{reclass}\}_{t,c}$ is equal to one for all device-type-years after a down-classification is enacted in device type c. For the event-study, the event-time $t = 0$ follows accordingly.

Identifying control device types that track the counterfactual development of the outcome variables is a central challenge in my empirical context. Controls could be unsuitable for several reasons. Control device types, for example, could be affected by unique scientific developments, have lower scientific potential, or face different market forces. Alternatively, some device types could be affected by spillovers from treated device types. Lastly, the FDA selects device types for down-classification based on inherent risk. Thus, down-classified devices may be less dangerous than those not chosen.

I provide four control groups, each addressing aspects of these concerns, and find that my results are robust across these groups. The first control group broadly comprises all Class III and II devices (for III to II events) and all Class II and I devices (for II to I events) that have not been down-classified. This group provides baseline DID estimates. The second group includes "later-treated" control device types that were down-classified after treated device types and after the latest sample year.[29] This "later-treated" group allows me to compare only device types that the FDA deemed appropriate for the same kind of down-classification. If later-treated device types are different from those treated earlier, the later-treated group may produce biased estimates. To ensure comparability, I form the third control group, a data-driven matched control group computed using nearest neighbor matching on baseline adverse events and innovation rates. Although I do not find evidence for spillovers in my context, I ensure that matched control device types do not treat the same medical ailments as treated device types.[30]

Finally, I provide a set of "intuitive" controls. This fourth set of controls includes medical device types that target similar diseases. I also ensure that device risk is intuitively and empirically comparable. For example, I avoid inappropriate comparisons between external-use devices and implantable or life-sustaining devices (e.g., contact lenses versus pacemakers), as these devices would have drastically different safety profiles. Instead, I compare like with like (e.g., daily- vs. extended-wear soft contact lenses). Profiles of the treatment and intuitive control groups are given in table A15 for Class III to II down-classifications, and in tables A16 and A17 for Class II to I down-classifications. Although the estimates are similar across control groups, the matched control groups constitute my preferred specification.

Additionally, some medical device types may never exhibit adverse events or innovative activity and thus would be incomparable to those that do. Thus, I also provide results from

[29] Specifically, for Class III to II events, I gather controls from all Class III to II events that occurred after 2015, censoring the outcome data after 2015. For Class II to I events, all device types moved from Class II to I in late 2019 constitute the control group. The 21st Century Cures Act drove this Class II to I event and was the first time FDA-initiated down-classifications of Class II devices occurred since 1998 (the year of the event I analyze). Importantly, the FDA used the same explicit down-classification criteria in both events.

[30] See table A14 for spillover estimates.

analyses that consider only treated and control device types with positive counts of a given outcome in the appendix tables A18, A19, and A20. My findings are robust to these restrictions.

As with every non-experimental research design, selection into treatment is a primary concern. Since the FDA selects medical device types to down-classify based on baseline yearly adverse event rates, down-classification may be endogenous to changes in adverse event rates.[31] Thus, I cannot ascertain how deregulation would affect the adverse event rates for a randomly chosen device type. However, I can speak to the optimality of the FDA's decisions on the margin of their rule (i.e., the most dangerous down-classified devices).

1.5 Results

This section presents estimates of equations 1.4.1 and 1.4.2, which capture the effect of deregulation on various outcomes of interest. Subsection 1.5.1 presents the effects on the flow and quality of innovation. Subsection 1.5.2 provides the effects on market structure. Subsection 1.5.3 details how the effects of deregulation on innovation and market structure differ by firm characteristics. Subsection 1.5.4 presents the effects on device safety.

1.5.1 Changes in Innovation

Table 1.2 reports estimates of equation 1.4.1 for my innovation outcomes.[32] Panel A provides estimates for Class III to II events, and panel B provides estimates for Class II to I events. Column (1) reports a 5-year pre-treatment mean of the outcomes for treated groups. Columns (2)–(5) report the estimates of equation 1.4.1 when comparing treated groups to a matched control group, intuitive controls, "later-treated" device types, and all untreated device types, respectively. Conley-Taber standard errors are reported below the estimates.

Table 1.2, panel A indicates that Class III to II events led to statistically significant increases in patenting rates, unique device submissions, mean citations-per-patent, and mean

[31] See appendix A5.1 for more details.

[32] Table A18 presents the results from only including device types with some positive outcome counts.

patent values across control group comparisons (columns 2–5). Depending on the control group, the results reveal that these events generated 189%–470% more patents and new device submissions per year per affected device type (pre-means: 8 patents/yr; 0.5 devices/yr). Patents filed after these events received 180% more citations and exhibited similar increases in market values. Panel B of table 1.2 shows that patents filed after Class II to I events (i.e., complete deregulation) received 330%–1,070% more citations and yielded 10%–50% higher market values, suggesting a divergence between scientific and private value. These results are robust across comparison groups (columns 2–5). Although economically significant, the increase in patenting rates from Class II to I events was not statistically significant under my preferred specification.

I examine the dynamics of the innovation responses by estimating the event-study equation 1.4.2. The top subpanels of figures 1.3 and 1.4 plot the innovation responses (i.e., β_t coefficients) for Class III to II and II to I events, respectively, when using the "matched" control groups.[33] The results of this analysis provide several insights for interpreting my findings. First, trends in all outcomes were similar in treatment and control groups for ten years before deregulation; trends were also similar for other control groups (not shown). This insight strengthens the identifying assumptions that (i) treatment and control groups would have exhibited similar trends in outcomes absent the policy change, (ii) policies were not anticipated, and (iii) policies were not endogenous to increases in innovative activity. Second, figures 1.3 and 1.4 indicate a persistent increase in the flow of innovation, suggesting that these events led to investments in new technologies that would not otherwise have occurred, rather than a forward shift in the timing of those investments.

Lastly, the event-study estimates for Class III to II events suggest that the increase in new technologies (i.e., patents) was slow, whereas the upsurge in access to new *and* existing technologies (i.e., unique devices submitted) was fast. This distinction, thus, is driven by rapid changes in the availability of existing technologies. First, firms may have "on-the-shelf" ideas and products that they have not commercialized due to the expensive approval process. Second,

[33] Figures A4 and A5 show event-study estimates for the innovation quality variables.

firms may promptly repurpose existing technologies for new indications. Third, deregulation accelerates the approval pipeline, leading to a sudden influx of products at different ex-ante stages of approval. Lastly, since, until recently, E.U. regulations were more lenient, firms may have introduced their E.U.-approved devices to U.S. markets after deregulation (Grennan and Town 2020a). By contrast, patenting rates increase gradually after deregulation, consistent with the time-intensive R&D process. U.S. patenting rates, unlike device submissions, are not affected by sudden influxes of existing technologies as these technologies are either already patented or are not patentable. In particular, if a firm files a patent in one country, it must file patents in other countries where it desires protection within one year to receive protection in those countries (Popp 2005). Applying for patents in multiple countries is inexpensive as firms can concurrently file patents in up to 153 countries through the Patent Cooperation Treaty (WIPO 2020).

1.5.2 Changes in Market Structure (Firm Entrants and Prices)

To investigate the effect of deregulation on market structure, I reestimate equation 1.4.1 for five different outcomes: new and incumbent firm entry measured separately by each data source and prices for procedures that use device types of interest. Table 1.3 presents the estimates.[34] The structure of table 1.3 is similar to that of table 1.2, with the exception of an additional comparison group matched on pre-event prices (column 2). Panel A reveals that Class III to II events led to statistically significant increases in incumbent and new firm entry across control groups (columns 3–6) and data sources (patents and FDA devices). Strikingly, these events increased the rate of new firm entry by 840%–1,000% (pre-mean: 0.1 firms/yr) when measured by FDA data and by 150%–420% when measured by patent data.[35] The discrepancy between the magnitudes of these two estimates suggests a strong increase in the availability of existing technologies. Regarding the effects on incumbent firms, these events increased

[34]Table A19 presents results from including only device types with some positive outcome counts.

[35]Supply-side factors may not be the sole driver of these dramatic changes in market structure. As shown in figure A6, there were considerable equilibrium forces at play: After the number of suppliers of treated device types increased, demand increased for procedures that use treated devices three years after deregulation, plausibly driven by lower prices. No significant pre-trends are measured.

incumbent entry by 400% when measured by FDA data and by 130%–240% when measured by patent data.

The procedure price estimates are reported in the first row of table 1.3. The results show that Class III to II events are associated with a statistically significant decrease in the prices of procedures that use treated device types when using two out of three control groups (columns 2 and 3). The estimates translate to a 33–40% drop in prices, plausibly driven by the increase in firm entry and competition (Busso and Galiani 2019).[36] There are several reasons why these price results should be interpreted with some caution. First, my price data is only available after 2004, restricting the number of treated device types I study to five. Second, the estimate generated using the entire sample of procedures as controls (column 6) is quite noisy, indicating that the results are less robust. Lastly, UCSD healthcare claims data only cover one regional hospital system.

Table 1.3, panel B shows the effect of Class II to I events on new and incumbent firm entry as measured by patent data (device data is unavailable for Class I). The results indicate that these events increased new firm patenting by 50%–145%, though the estimate under my preferred specification is only marginally significant. By contrast, incumbent firm entry is statistically and economically insignificant under my preferred specification. The distinction between the new and incumbent results suggests that litigation may obstruct new entry less than regulation, but both environments similarly impact incumbent firms.

To help interpret these findings, I present event-study estimates of equation 1.4.2 for my market structure outcomes. The β_t coefficients are shown in the bottom subfigures of figures 1.3 and 1.4 for Class III to II and Class II to I events, respectively.[37] The figures suggest that

[36]The example of spinal implant deregulation highlights the plausibility of these price estimates. There are several margins along which a drop in the price of spinal implants could affect the overall costs of spinal fusion procedures. First, spinal implants account for roughly 40% of the costs of spinal fusion procedures (Beckerman et al. 2020). Thus, the direct effect of a drop in the prices paid for spinal implants could measurably change the procedure price. Moreover, new technology could be labor-saving, reducing the costs of labor required to perform the procedure. Lastly, a lower price for spinal implants could attract more providers to offer the procedure, potentially driving down prices further.

[37]Figure A7 plots these coefficients for the Class III to II price outcome, and figure A8 plots these coefficients for the Class III to II market structure outcomes measured using patent data.

identifying assumptions (i)–(iii) (listed above) are satisfied and that, when present, the estimated effects are persistent. For similar reasons given above, figures A8 and 1.4 illustrate a gradual increase in the rate of new firms patenting (slow R&D), while FDA device data reveals a sharp increase in device submissions from new firms (includes existing technologies). Figure A7 reveals that procedure prices dropped two years after the events, despite sharp increases in firm entry. This lagged response is consistent with the contractual nature of healthcare markets; prices are "sticky" as hospitals periodically renegotiate contracts with suppliers and insurers (Reinhardt 2006, Grennan and Swanson 2020).

1.5.3 Heterogeneity in Firm Proficiency and Size

The average treatment effects estimated in the last two sections overlook heterogeneity in firm size and regulatory proficiency. In this subsection, I separately estimate equation 1.4.1 across firm size and proficiency quantiles for the outcomes of interest. I link this heterogeneity analysis to the propositions in section 1.2 to gain further insight into the mechanisms that drive the overall results. The identified mechanisms highlight design elements that may make regulation more amenable to small and inexperienced firms.

Firm Proficiency. To examine how regulation affects firms with different regulatory proficiencies, I estimate equation 1.4.1 for the device submission outcome across proficiency quartiles. I center this analysis on FDA data, allowing a cleaner linkage between firms, proficiency, and innovation. Panel A of figure 1.5 presents the results expressed as percent changes relative to pre-event averages. Class III to II events generated statistically significant increases in new device submissions across proficiency quartiles. However, the events were associated with much higher increases among inexperienced firms. Firms in the first proficiency quartile exhibited a 1,000% increase in new device submissions compared to a 50% increase from firms in the top quartile.[38] These results indicate a quickly diminishing response while moving up the

[38] Strategic judgment proofing is not driving these results. In other words, these effects are not driven by larger firms forming small subsidiaries to shield themselves from liability. For example, only 1 out of 20 new spinal implant manufacturers entering the market after deregulation were subsidiaries.

proficiency distribution. This pattern is consistent with the estimated learning curves presented in figure 1.5, panel B as firms in the lowest proficiency quartile benefit from the highest reduction in approval delays. This reduction translates into outsized decreases in commercialization costs for inexperienced firms and, thus, higher increases in commercialization activity (as claimed in proposition 3).

Designing regulation that is simpler and standardized could help less regulation-proficient firms.[39] For example, Stern (2017) shows that when the FDA sets approval expectations by publishing guidance documents, approvals times of new firms drop by roughly 40 percent. To simulate the impact of these types of efforts on innovation, I iteratively shrink the gap in delays between inexperienced and proficient firms by lowering the learning rate γ while measuring R&D response from a hypothetical distribution of firms (see figure A9 and appendix A3.2 for more details). Table A21 presents the results of this simulation. The results suggest that flattening the learning curve could increase the number of unique devices approved up to 63%, with the least proficient firms exhibiting the largest gains.

Firm Size. To assess how regulation impacts firms with different levels of internal capital, I estimate equation 1.4.1 across capital terciles for the patenting rate outcome. I perform this analysis for both down-classification types. Figure 1.6, panels A and B present the results.[40] Both event types are associated with larger increases in patenting rates among firms in the bottom tercile of asset holdings.

Interpreting the heterogeneous effects of regulation through the lens of my conceptual framework indicates that profits increase after deregulation and that small firms face lower financing costs after deregulation, despite incurring potentially higher safety effort costs. These results confirm aspects of the propositions in section 1.2 and suggest that small and inexperienced firms face relatively high regulatory costs to innovate.

[39] In multiple interviews, inventors described to me the FDA approval process as "byzantine" and "too much for us to navigate alone."

[40] I focus on patents for two reasons. First, they can be linked easily to patent applicants and capital holdings. Second, patents allow comparisons across down-classification types.

The results of this subsection should be interpreted with some caution. Other factors may be correlated with firm size and proficiency that also contribute to these R&D responses. However, in addition to the striking similarity between the empirical results and the predictions made in section 1.2, device manufacturers express that regulatory proficiency and financing costs are key factors that influence R&D decisions.[41]

1.5.4 Changes in Device Safety

I examine whether deregulation is associated with decreased device safety by reestimating equation 1.4.1 for two different outcomes: the rate of adverse events and the rate at which inventors emphasize safety. Table 1.4 details the results and is structured like table 1.2.[42] Table 1.4, panel A reveals that Class III to II events are not associated with statistically significant changes in adverse event rates and inventor emphasis across control groups. However, these events are associated with economically significant increases in hospitalization rates under my preferred specification.

Table 1.4, panel B shows that Class II to I events are associated with statistically significant *reductions* in the rates of hospitalizations and deaths across three out of four control groups. In contrast to Panel A, all but two estimates are significant at the 10% level, and all suggest improvements in device safety. The results indicate an associated 93–97% reduction in hospitalizations and a 49–69% reduction in deaths per year per treated device type (pre-mean: 0.3 deaths/yr). Panel B reveals that these events are also associated with a statistically significant 100% *increase* in the share of patents that emphasize an advancement in product safety, corroborating the results generated by the FDA adverse event report outcomes.

How could deregulation *improve* device safety? A compelling answer is that deregulation exposes firms to more litigation, which may increase the net incentives to improve device safety.[43]

[41] Firm size, the most obvious potential confounder, is uncorrelated with firm FDA experience (see table A22). This lack of correlation may result from publicly traded companies having high baseline assets relative to the average MedTech firm.

[42] Table A20 presents the results from including only device types with some positive outcome counts.

[43] Several other potential mechanisms may contribute to improved product safety after Class II to I down-

To shed further light on liability as the mechanism for this change, I use variation in ex-post exposure to legal liability by firm size. Small firms can avoid worst-case damages through bankruptcy, while large firms cannot. If liability risk plays a central role, deregulation should lead to disproportionate increases in device safety among larger firms. Indeed, the top subfigure of figure 1.7 shows that larger firms in the top tercile of asset holdings exhibit a significant 100% increase in the likelihood of demonstrating at least one safety innovation per year per treated device type. By contrast, smaller firms respond much less dramatically. The bottom subfigure of figure 1.7 mirrors this finding and shows a more significant drop in the likelihood of serious adverse events among larger firms.

Figures A10, A11, and A12 illustrate the dynamics of my device safety findings. These figures plot the β_t coefficients estimated from event-study equation 1.4.2. Figure A10 shows that Class III to II events are associated with a gradual increase in hospitalization rates and serious event rates as new devices are invented and marketed within treated device types. Figure A11 shows that Class II to I events are associated with a persistent and gradual decrease in adverse events as inventors increase their emphasis on safer technologies (see also figure A12).

A few caveats accompany my device safety analysis. First, the FDA explicitly down-classifies device types for which prospective regulation adequately mitigates harm. Thus, the insignificant adverse event results associated with Class III to II events should be interpreted as a local average treatment effect. For Class II to I events, however, I use the FDA decision rule described in appendix A5.1 to assess whether the FDA's decisions are optimal on the margin (i.e., at higher "DPM scores"). Accordingly, I separately estimate equation 1.4.1 for each treated device type relative to a matched control (matched based on DPM score) and plot the relationship between the effect size and the score value. Figure A13 shows that marginal device types are

classifications. For example, deregulation may increase competition among firms, which may encourage them to focus more on product safety as a means of differentiation. Additionally, deregulation can lead to increased innovation, which may result in more product safety innovations. However, I do not observe similar safety improvements after Class III to II down-classifications, where innovation and market competition tend to increase more significantly. It is also possible that, after deregulation, firms are no longer constrained by regulatory parameters such as substantial equivalence, allowing them to more freely innovate in the realm of product safety.

associated with *fewer* deaths when compared to control groups, relative to less dangerous treated device types. This pattern may generalize to most current Class II device types, of which 95% exhibit fewer adverse events than the most marginal deregulated device type.

Second, the FDA does not normalize adverse event rates by device utilization due to data limitations. Growth in utilization would increase the likelihood of adverse events. Thus, fluctuations in adverse event rates reflect changes in product safety *and* utilization. Hence, using adverse event rates as a signal of product safety provides a conservative estimate of the net benefit of deregulation as deregulation increases utilization. Figure A6 shows that, although no pre-trends are present, utilization rates of treated medical device types significantly increase three years after Class III to II deregulations, plausibly due to increased supply. Although I do not have similar utilization data for Class II to I events, treated device types also exhibit increased supply after deregulation. All else equal, if the demand curve is not perfectly inelastic, an outward shift in the supply curve would increase utilization.

Lastly, media and regulatory decisions may influence adverse event reports. Manufacturers, for example, could be less likely to report adverse events if they are subject to less regulatory scrutiny or if reports are more likely to make news after deregulation. However, I focus on mandatory reports of deaths or severe injuries from hospitals and device manufacturers, which are less sensitive to these factors than voluntary reports of less severe injuries (FDA 2020c). The FDA enforces the reporting of serious events using financial penalties and criminal resolution (Bragg et al. 2018, Emergo 2022).[44] Lastly, when the FDA announced the largest Class II to I down-classification event in 1995, it created new authorities that enabled closer monitoring of the affected devices to "take appropriate remedial action, if necessary" (FDA 1995), suggesting that adverse event reports would be more challenging to conceal.

[44]Both user facilities (i.e., hospitals) and manufacturers are required to report serious adverse events to the FDA. Thus, if either entity fails to report an event, but the FDA is notified by the other (or other sources like end users), then it is implicated in incompliance. Additionally, the FDA *increased* its monitoring of deregulated device types to take appropriate remedial action if products had become less safe, which would make it it more difficult for firms marketing affected devices to hide adverse events relative to those marketing unaffected devices (FDA 1995).

1.6 Back-of-the-Envelope Calculation: Costs & Benefits

This section presents the costs and benefits of deregulation, which are measured by the three core results derived in section 1.5. First, deregulation increases patenting rates. The value of this increase is determined by the sum of each additional patent's market value, accounting for creative destruction and increases in value from deregulation. Second, deregulation decreases market concentration and healthcare prices. To value lower healthcare prices, I convert price changes to changes in expenditures by assuming constant utilization. Lastly, complete deregulation reduced adverse event rates. The resulting drop in deaths is appraised at the statistical value of all lives saved, while prevented hospitalizations are valued according to Moses et al. (2019). The assumptions and math underlying these calculations are detailed in table 1.5.

Table 1.5 presents the measured costs and benefits of down-classification decisions. To justify the FDA's decision rule for Class III to II down-classifications, the unmeasured costs (e.g., political risks) associated with these events would have to be larger than the measured costs. Class II to I down-classifications do not exhibit any measurable costs as they are associated with *fewer* adverse events and more innovative activity. The benefits of these down-classifications, including fewer adverse events, amount to roughly $24 million per year per treated device type, even at the margin of the most dangerous treated devices ex-ante. Since there are 2,500 Class II devices, the yearly forgone net benefits from stalling deregulation could amount to as much as $60 billion, or nearly 34% of the value of medical devices consumed each year.

I do not include all costs and benefits of deregulation in these calculations. For costs, I do not measure the value of efficacy assurances provided by the FDA, which are lost after down-classification (see Grennan and Town (2020a)). However, one criterion for down-classification is whether device efficacy is easily verifiable and maintained after deregulation, so these costs are likely small. Second, waiting to deregulate to learn more about a device type's inherent risk is valuable if deregulation could lead to increased adverse events (i.e., the option value of waiting). However, Class II regulations are associated with increased adverse event rates relative to Class

I, so waiting to deregulate may not provide value. Lastly, there are potential political costs of misguided deregulation that I do not measure.

The unmeasured benefits of deregulation include reductions in FDA administrative costs, price reductions from Class II to I events, the value of new jobs created with firm entry, the benefits of innovation from private firms, and the scientific value of innovation.

1.7 Discussion and Conclusion

This paper analyzes the effect of regulation on medical device innovation, market structure, and adverse events. My theoretical model clarifies how "learning by doing" and financing costs make regulation especially burdensome for small and inexperienced firms investing in the development of new technologies. In turn, the model shows that deregulation increases the profitability of innovation most for these types of firms and may raise the net incentives to improve product safety by exposing firms to greater liability risk. I then investigate these insights, and my broader questions, empirically in the context of the medical device industry, where complex regulations prevent litigation. For my empirical analysis, I develop a data set that combines eight underlying sources on innovation, market dynamics, firm characteristics, and product safety. I find that deregulation disproportionately benefits small and inexperienced firms and broadly accelerates technological progress and firm entry. This change in market structure reduces related healthcare prices. Lastly, Class II to I down-classifications are associated with a significant decrease in adverse events, providing evidence that legal liability risk creates strong incentives to improve product safety relative to the requirements of medical device regulation. Increases in product safety are highest among devices originating from large firms that have the most assets at risk in liability proceedings, providing additional evidence supporting liability as the driver of this result.

A back-of-the-envelope calculation suggests that deregulation exhibited higher measured benefits than costs. Class II to I events are associated with net benefits amounting to $24 million

per year per treated device type. These benefits are higher for marginal, higher-risk device types, suggesting my results may generalize to other Class II devices.[45] These results align with sentiments from the National Institute of Medicine and physician commentators, which have criticized the effectiveness of Class II regulations and have advocated for alternatives that ensure quality and encourage innovation. My results suggest that deregulating Class II devices, relying instead on the deterrent effects of litigation, is one such alternative: litigation can improve product safety, hasten innovation, and lower administrative costs.

Class III to II events, however, are difficult to evaluate. On the one hand, I find that the benefits of deregulation, namely a 470% increase in the availability of new technologies, are quite large. In the short run, the magnitude of this increase is consistent with deregulation removing the wedge between the available technologies in the E.U. and the U.S. For example, over 80% of cardiac stents marketed in the E.U. are unavailable in the U.S., a potential byproduct of regulation (Grennan and Town 2020a). In the long run, the increase in access to new technologies is persistent. In practice, however, these events present the FDA with asymmetric costs and benefits; an increase in salient device-related deaths could degrade the regulator's reputation and undermine its more cost-effective efforts elsewhere (Carpenter 2004a;b). In contrast, the technological benefits that come from deregulation are more abstract. Thus, the FDA's optimal strategy may be "too conservative" (Isakov et al. 2019) relative to the social optimum to uphold its reputation at the expense of innovation. This asymmetry is evident in FDA documents outlining the criteria for down-classification as the value of forgone innovation is not considered. This study seeks to clarify these forgone benefits. However, more empirical research is needed to assess the costs of regulatory mistakes and the value of regulator reputation.

My study focuses on the large and growing medical device market, but the results may also be relevant to other settings with similar regulations. For instance, FDA regulations for Class III devices are similar to those in the EU, and requirements for these devices resemble

[45] Moreover, 95% of current Class II devices have lower adverse event rates than the most dangerous deregulated device type before deregulation.

those for brand-name drugs in the US and other countries (Van Norman 2016).[46] Additionally, Class II device regulations are similar to those used abroad and resemble those for generic drugs—which are also protected from product design tort claims after FDA approval—and genetically modified (GM) foods (Schwartz and Appel 2020, Schauzu 2000). These similarities suggest that medical technology and food regulations may slow innovation and increase market concentration worldwide. Lastly, my analysis highlights the potential issues that arise when regulators use imperfect proxies or heuristics to evaluate product quality, such as the "substantial equivalence" heuristic used for Class II devices, generic drugs, tobacco products, and GM foods. These heuristics may be particularly pervasive when product quality is hard to verify or when regulators are under-resourced. In such situations, a robust legal system with impartial judges and high damage caps may better incentivize product safety through litigation.

[46] Tabarrok (2000) offers some evidence that FDA pharmaceutical regulations are too stringent.

Figures and Tables

Figure 1.1. Background on Medical Device Regulations. Note: This figure presents background on FDA Medical device regulations and the deregulation policy changes I leverage in my analysis. Device types are placed into one of three classes, each corresponding to a level of perceived risk. Higher perceived risk requires a longer approval process and additional costs to conduct testing and maintain business operations before a product is approved. The time and cost values are averages within the given class and are derived from Makower et al. (2010). While learning about a device type's underlying risk, the FDA can deregulate a device type by moving it from a higher-risk class to a lower-risk class (called "down-classification"). This decision dramatically reduces the approval delays and costs that device manufacturers confront. The FDA rarely reclassifies device types into a higher-risk class. The last column includes examples of Class III, II, and I devices, namely, pacemakers, x-ray machines, and tongue depressors, respectively. *Medical devices with attendant "special controls" requirements (Class II devices) are often protected from product liability (Costello and Pham 2016). However, there is no supreme court precedent that guarantees preemption; thus, courts exercise some discretion in their interpretation of federal preemption with Class II devices.

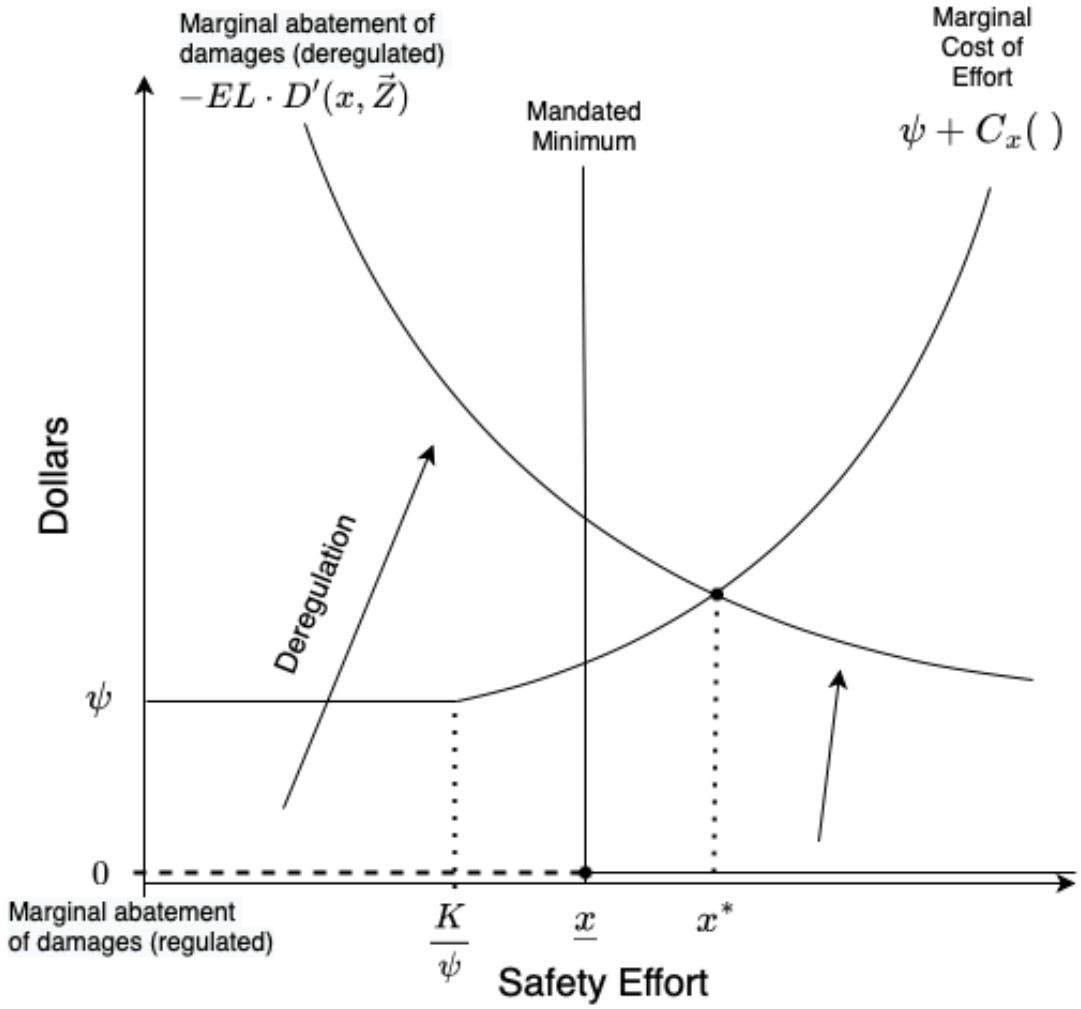


Figure 1.2. Theoretical Change in Safety Effort after Deregulation. Note: This figure presents a possible change in the level of safety effort after deregulation. This scenario is one in which deregulation could lead to an increase in safety effort, given a sufficient increase in damages, as described in section 1.1. The x-axis indicates the level of safety effort exerted. The y-axis denotes the monetary value. The marginal cost of effort curve indicates a marginal cost of psi at initial values of safety effort before financing costs are incurred, at which point marginal costs increase with effort. The marginal abatement of damages curve under regulation is always equal to zero due to federal preemption. The counterfactual dotted section of the marginal abatement curve under regulation represents the marginal abatement of damages from exerting effort below mandated levels while still achieving FDA approval. Deregulation shifts the marginal abatement curve as legal damages are no longer prevented by federal preemption. The value x-star represents the optimal level of safety effort after deregulation (i.e., where the marginal cost of safety effort is equal to the marginal abatement of expected damages). The value x-underbar represents the mandated level of safety effort. The vector Z contains other factors that affect a firm's legal damages in expectation, which might be specific to the given legal system, like damage caps.

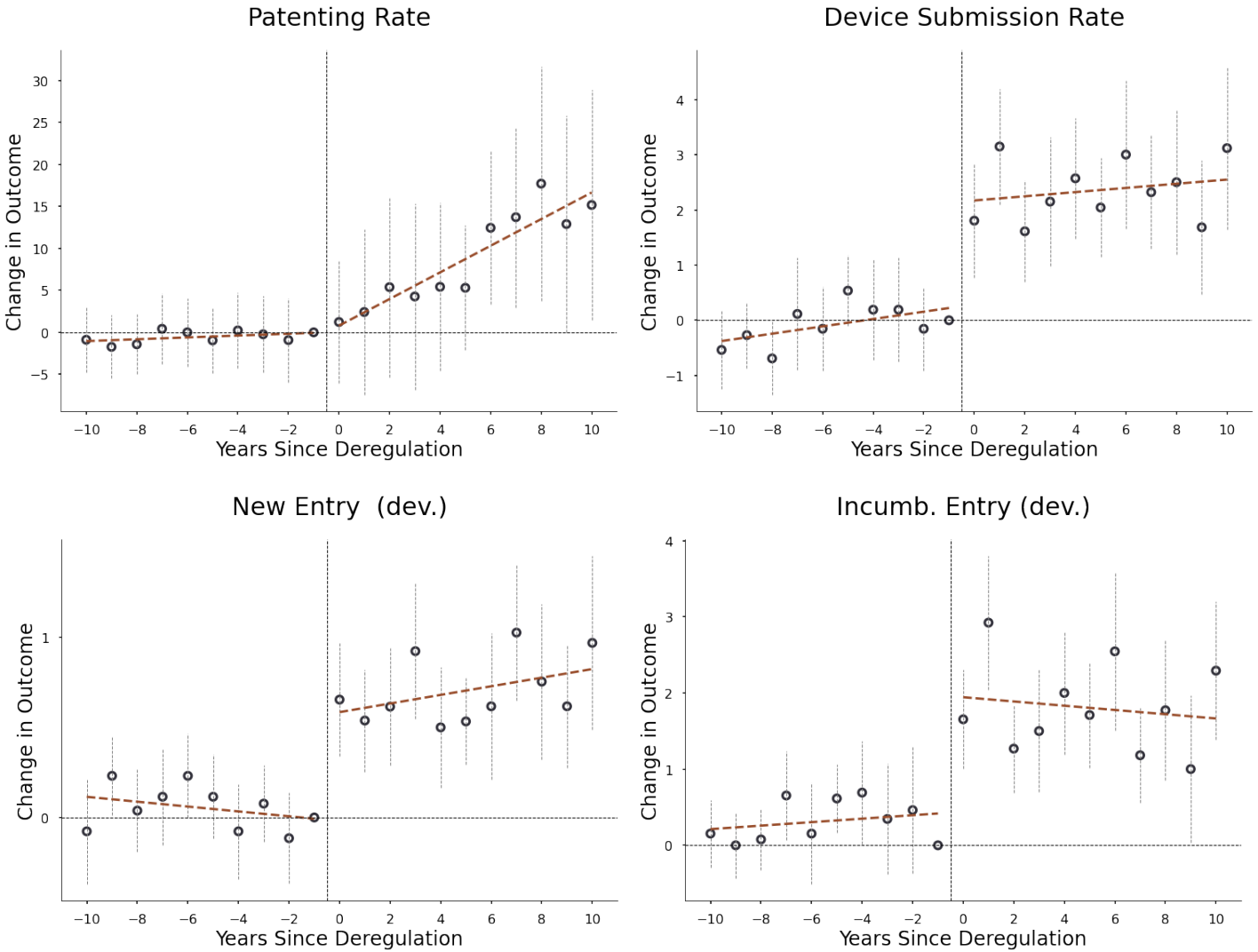


Figure 1.3. Effects of Class III to II Events (High to Moderate Regulation). Note: This figure presents the estimates of the coefficients from the event-study equation 1.4.2 for some innovation and market structure outcomes. Only Class III to II down-classification events are considered. Controls are device types matched on baseline averages of the outcome. Data are analyzed at an annual frequency. The top-left subfigure illustrates the evolution of patents filed per year in treated device types relative to matched control groups. The top-right subfigure describes the evolution of unique devices approved per year by the FDA for treated device types relative to control groups. The bottom-left subfigure illustrates the evolution of the rate of new firm entry (counts per year), calculated using device submission data relative to matched control groups. New firm entry represents firms that have never before submitted FDA documentation. The bottom-right subfigure illustrates the evolution of the rate of incumbent firm entry (counts per year of firms that have previously submitted FDA documents) in treated device type relative to controls. Standard errors are calculated following Conley and Taber (2011).

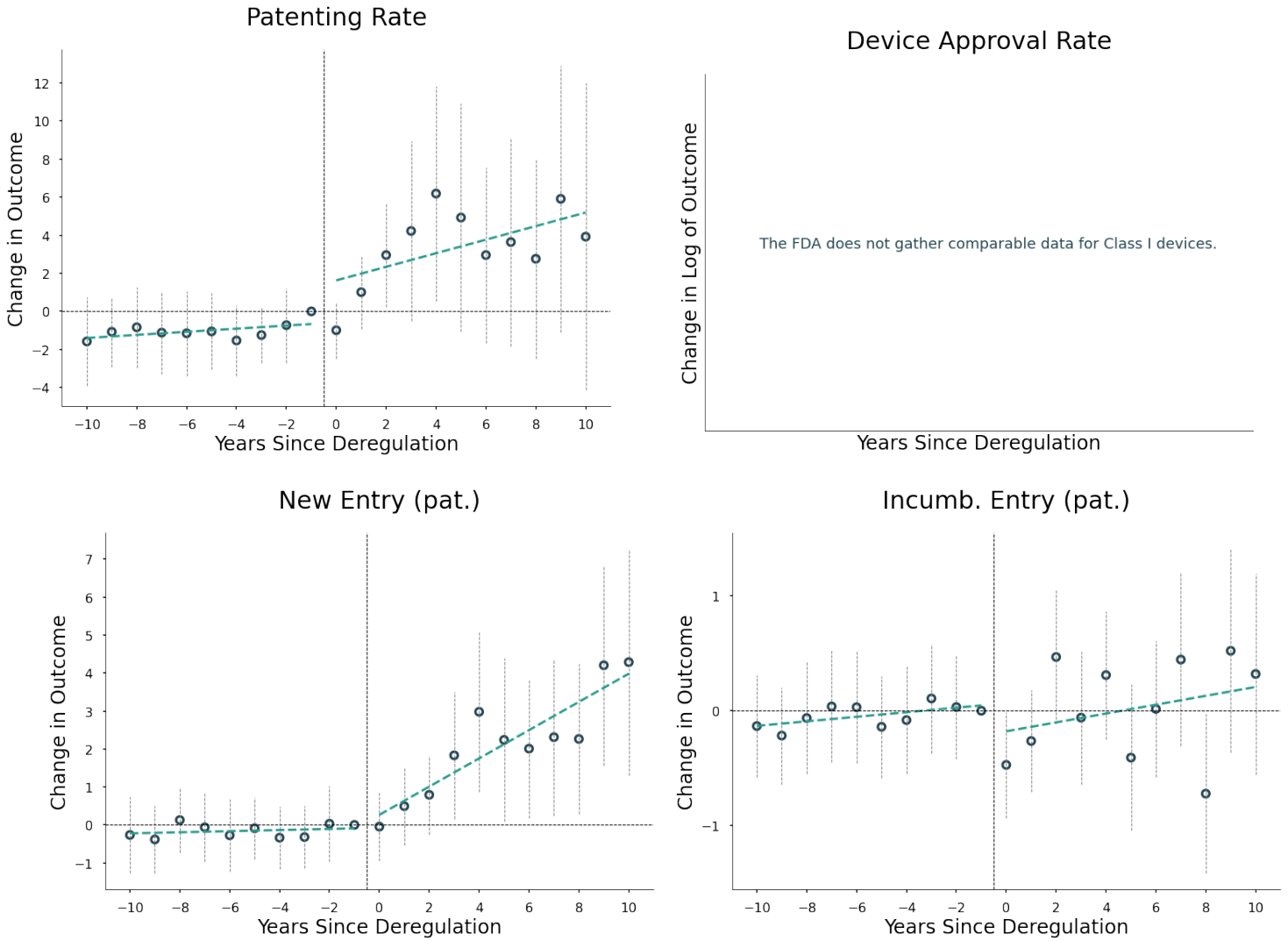
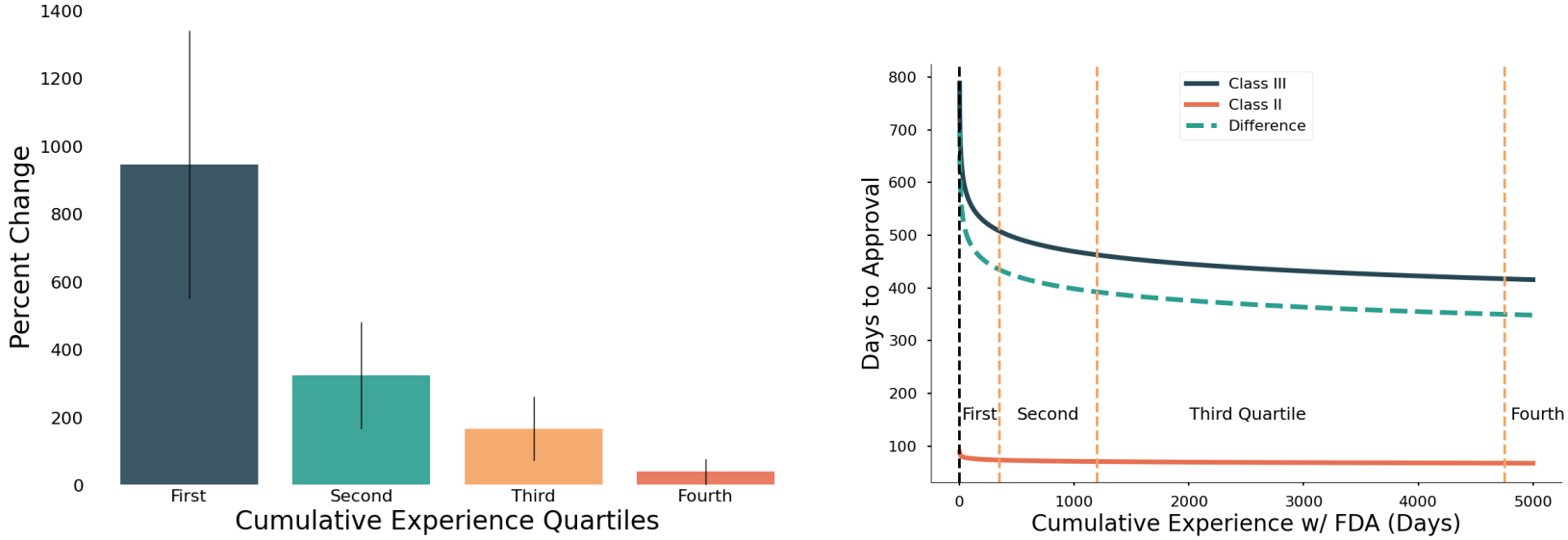


Figure 1.4. Effects of Class II to I Events (Moderate to Low Regulation). Note: This figure presents the estimates of the coefficients from event-study equation 1.4.2 for some innovation and market structure outcomes. Only Class II to I down-classification events are considered. Controls are device types matched on baseline averages of the outcome. Data are analyzed at an annual frequency. The top-left subfigure illustrates the evolution of the patenting rate of treated device types relative to matched control groups. The top-right subfigure is blank, as there is no comparable data for Class I approved devices. The bottom-left subfigure illustrates the evolution of the rate of new firm entry (measured by new firms patenting) relative to matched control groups. The bottom-right subfigure illustrates the evolution of the rate of incumbent firm entry (firms that have received a granted patent), entering treated device types relative to matched controls. I do not include FDA-approved device measures of new and incumbent entry as I do not have reliable data on new Class I devices from FDA sources. 95% confidence intervals are provided.



Panel A: Change in New Devices by Proficiency

Panel B: Estimated Learning Curves

Figure 1.5. Effects on Innovation by Experience and Estimated Learning Curves. Note: This figure presents the experience-specific changes in the rates of newly marketed devices stemming from class III to II down-classification events and the learning curves estimated in equation A3.1. Panel A provides the DID estimates of the rate of newly marketed devices in treated device types, relative to controls, by experience quartiles. DID estimates and standard errors are converted to percent changes. Firm experience is calculated by aggregating each firm's total time spent satisfying FDA regulations up to the time of submitting an approval for the current device. Panel B presents the estimated learning curves for satisfying Class III and Class II regulations. The difference between Class III and Class II approval delays at a given level of FDA experience is also provided. The x-axis indicates the number of days spent on previous approvals. The y-axis describes the number of days taken for a current Class III or Class II device to be approved. I provide divisions of cumulative experience quartiles seen in the data. I exclude observations with no prior experience to avoid undefined outcomes and biases from the extensive margin in the estimation. The 95% confidence intervals overlay the estimates. The simulated confidence intervals are calculated using a Monte Carlo procedure. This procedure produces estimates across repeated random draws from the empirical distribution of firm characteristics.

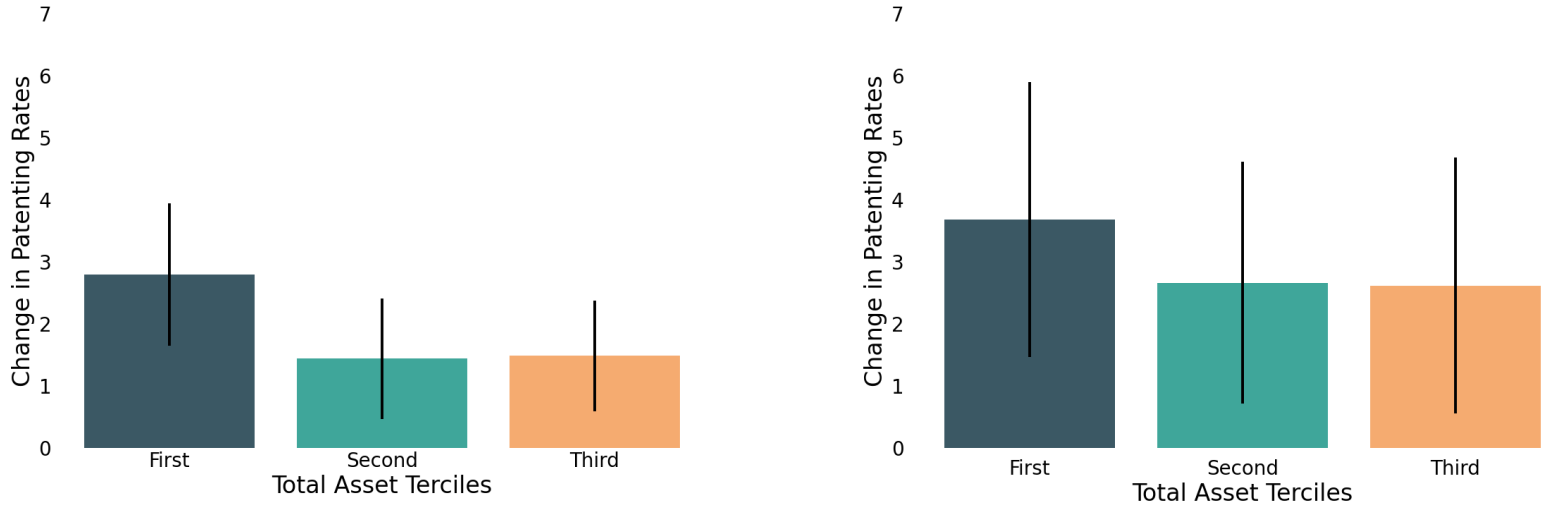


Figure 1.6. Effect of Down-Classification on Patenting Rates by Asset Terciles. Note: This figure presents the DID estimates from equation 1.4.1 for the patenting rate across down-classification types and firm asset terciles. For the empirical estimates, I exclude patent data for private firms since I only observe firm asset data for publicly traded firms. Panel A presents the change in patenting rates in my Class III to II treated medical device types, relative to matched control groups, across asset terciles. The first tercile represents the bottom 33rd percentile of assets, the second represents the 33rd–66th percentile, and the third represents the 66-100th percentile. Panel B presents the change in patenting rates in my Class II to I treated medical device types, relative to matched control groups, across asset terciles. 95% confidence intervals overlay the estimates. Simulated confidence intervals are calculated using a Monte Carlo procedure. This procedure produces estimates across repeated random draws from the empirical distribution of firm characteristics.

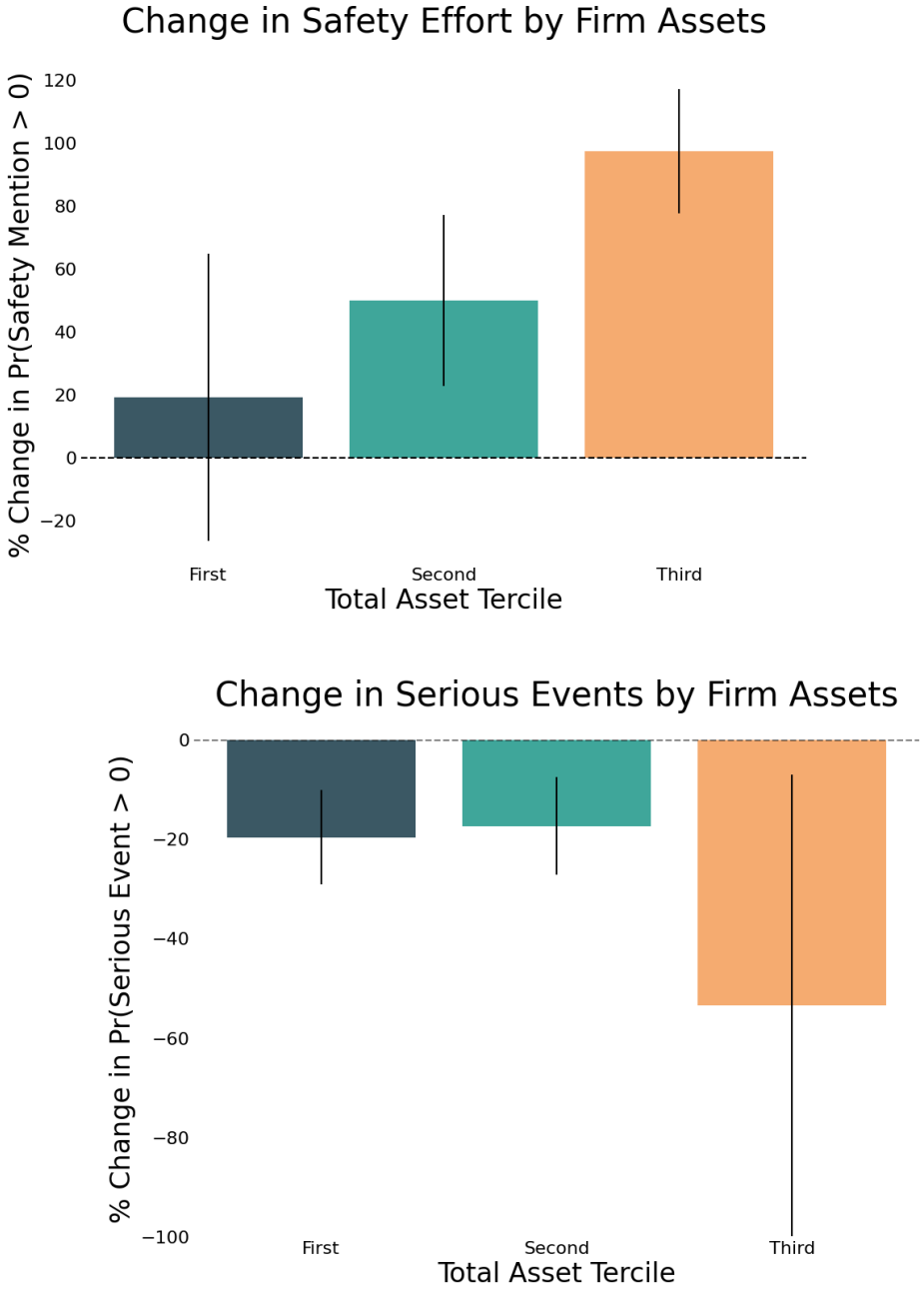


Figure 1.7. Change in Emphasis on Safety by Firm Asset Terciles (II to I). Note: This figure presents separate DID estimates of equation 1.4.1 for the change in the likelihood of device types exhibiting at least one annual occurrence of the given outcome variable by firm asset terciles. I set all outcomes greater than zero to one (LPM) as safety mentions and serious events are rare. The baseline outcome values across asset terciles are roughly equal and do not drive these disparate effects. The top figure presents the change in the likelihood of safety-related innovations, and the bottom figure illustrates this change for serious adverse events (death, hospitalization, or life-threatening event). Terciles are formed using the asset totals from firms that are publicly traded. The x-axis describes the tercile: first, second, or third, and the y-axis conveys the percent change in the likelihood. 95% confidence interval bars are provided.

Table 1.1. Summary Statistics. Note: Tables A7, A8, and A9 provide summary statistics for each class independently. See Kogan et al. (2017) for more information on the patent market valuation data, which was merged into my patent dataset. The CRSP/Compustat database was used to derive the total assets of the firms applying for patent protection and is a proxy for firm size. Market values and applicant assets are only available for patents filed by publicly traded firms, representing roughly 25% of the total sample of patents. Missing observations account for the discrepancies between (i) the number of total FDA device types (5,542) and the number of device types represented in device submissions, adverse event reports, and patents (many device types have no associated patents), (ii) the total number of patents and the number of patents with market valuations and applicant assets, and (iii) the total number of claims and claims containing amounts paid. *"Regulatory proficiency" indicates the total number of days a firm has experienced approval delays across all its submitted devices.

	N	Mean	SD	Range
FDA Admin. Data—Device Submissions (PMA and 510(k) Databases)				
Total	168,880	-	-	-
per Device Type	4,710 (Types)	35.5	110.8	[1, 2,457]
Total Submitting Firms	20,343	-	-	-
Firms per Device Type	4,710 (Types)	15.7	39.5	[1, 1,048]
Firm Regulatory Proficiency	4,660 (Types)	19.5yrs	65.4yrs	[0, 686.2yrs]*
FDA Admin. Data—Adverse Event Reports (MAUDE)				
Total	9,238,733	-	-	-
per Device Type	4,111 (Types)	2,353.3	18,939.9	[1, 0.6M]
Serious Events per Dev. type	2,400 (Types)	571.7	5186.8	[1, 0.15M]
Assets of Offending Firm	7,139,727	$3.76B	$5.77B	[$0, $0.79T]
USPTO Device Patents				
Total	1,248,292	-	-	-
per Device Type	2,113 (Types)	590.8	2077.4	[1, 23,056]
Citations	1,248,292	14.6	88.8	[1, 5,817]
Market Valuation	377,465	$13.1M	$30.7M	[$45, $1.9B]
Applicant Assets	377,465	$26.7B	$54.8B	[$0.07M, $1.1T]
UCSD Healthcare Claims Extract				
Total	495,519	-	-	-
per Procedure Code	528 (Codes)	880.4	2397.5	[1, 18,915]
Unique Patients	55,621	-	-	-
Price	453,079	$135.7	$389.0	[$0, $0.01M]
Price per Proc. Code	528 (Codes)	$354.8	$576.1	[$0, $5,401]

Table 1.2. Effect of Down-Classifications on Innovation. Note: The table presents estimates of equation 1.4.1, which is a difference-in-differences (DID) style OLS regression model. Outcomes are derived from USPTO patent databases, FDA administrative data, and Kogan et al. (2017). Column (1) presents the 5-year baseline average of treated device types for the outcomes listed on the left-hand side. Columns (2)–(5) present DID estimates for the listed outcomes using different control groups: namely, a matched control group, intuitively similar device types (treat similar diseases), "later-treated" device types (treated after sample window), and the full sample, respectively. Device submissions are derived from FDA data and are not available for Class I devices. For column (4), Class III to II, control device types are treated after 2015; thus, all observations after 2015 are dropped. Confidence intervals are calculated using Conley–Taber test statistics. +, *, **, and *** correspond with statistical significance at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

		DID Estimates			
	Pre-mean	Matched	Intuitive	Later	Full
Down-Classification	(1)	(2)	(3)	(4)	(5)
A. Class III to II:					
Patenting Rate	7.95	14.99**	25.61**	26.65*	18.14
	(9.27)	(5.57)	(8.98)	(10.36)	(20.58)
Device Submission Rate	0.47	2.69***	2.36**	2.26**	2.22***
	(1.03)	(0.59)	(0.77)	(0.73)	(0.33)
Citations-Per-Patent Rate	9.06	16.59*	21.86*	19.43**	26.24***
	(20.65)	(7.48)	(9.81)	(6.41)	(5.62)
Average Patent Value	4.36	8.24***	11.29***	11.58***	10.50***
	(6.12)	(1.81)	(2.91)	(2.96)	(1.59)
Sample Size		1540	1056	920	60456
B. Class II to I:					
Patenting Rate	16.32	7.34	7.06	13.32**	29.17***
	(37.11)	(4.86)	(6.77)	(5.01)	(7.18)
Citations-Per-Patent Rate	0.64	6.85**	2.12*	3.98***	6.00***
	(0.48)	(2.30)	(1.08)	(0.84)	(1.43)
Average Patent Value	6.49	3.37***	0.90+	2.04***	6.13***
	(14.19)	(0.67)	(0.47)	(0.46)	(0.56)
Sample Size		15180	20592	27764	32472

Table 1.3. Effect of Down-Classifications on Market Structure. Note: The table presents estimates of equation 1.4.1, which is a difference-in-differences (DID) style OLS regression model. Column (1) presents the 5-year baseline average of treated device types for the outcomes listed on the left-hand side. Columns (2)–(6) present DID estimates for a given outcome using different control groups. These groups are (2) matched on baseline prices, (3) matched on baseline innovation and adverse event levels, (4) an intuitively comparable group, (5) a later-treated group, and (6) the full sample of controls, respectively. Column (5) of Panel A uses control device types treated after 2015, so all observations after 2015 are dropped. Procedure prices were only available after 2004, restricting sample size. There are no price estimates in columns (4) and (5) due to data limitations. Confidence intervals are calculated using Conley–Taber test statistics. +, *, **, and *** correspond with statistical significance at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

		DID Estimates				
Down-Classification	Pre-mean (1)	Price (2)	Matched (3)	Intuitive (4)	Later (5)	Full (6)
A. Class III to II:						
Procedure Price	95.31	-58.25**	-43.54**	-	-	-27.50
	(123.95)	(21.16)	(15.66)	-	-	(144.11)
Sample Size		160	176	-	-	36240
Incumb. Entry (dev.)	0.40	-	1.58***	1.48**	1.46**	1.44***
	(0.91)	-	(0.36)	(0.54)	(0.52)	(0.22)
New Entry (dev.)	0.07	-	0.67***	0.70**	0.59**	0.63***
	(0.31)	-	(0.19)	(0.22)	(0.19)	(0.13)
Incumb. Entry (pat.)	1.47	-	1.91**	2.78**	3.56**	2.98*
	(1.78)	-	(0.59)	(1.01)	(1.34)	(1.48)
New Entry (pat.)	3.78	-	5.63***	11.19**	11.94**	8.88
	(4.76)	-	(1.61)	(3.75)	(4.31)	(6.32)
Sample Size		-	1364	1056	920	60456
B. Class II to I:						
Incumb. Entry (pat.)	2.26	-	0.04	0.32	0.61*	1.36**
	(4.33)	-	(0.45)	(0.36)	(0.29)	(0.42)
New Entry (pat.)	7.27	-	3.85+	2.60	4.87**	10.55***
	(16.87)	-	(1.99)	(2.10)	(1.57)	(2.07)
Sample Size		-	13552	20592	27764	32472

Table 1.4. Effect of Down-Classifications on Adverse Events. Note: The table presents estimates of equation 1.4.1, which is a difference-in-differences (DID) style OLS regression model. Column (1) presents the 5-year baseline average of treated device types for the outcomes listed on the left-hand side. Adverse event outcomes are derived from the FDA MAUDE database. Columns (2)–(5) present DID estimates for the listed outcomes using different control groups: namely, a matched control group, intuitively similar device types (treat similar diseases), "later-treated" device types (treated after sample window), and the full sample, respectively. For column (4), Class III to II, control device types are treated after 2015; thus, all observations after 2015 are dropped. Confidence intervals are calculated using Conley–Taber test statistics. +, *, **, and *** correspond with statistical significance at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

		DID Estimates			
Down-Classification	Pre-mean (1)	Matched (2)	Intuitive (3)	Later (4)	Full (5)
A. Class III to II:					
Emphasis on Safety	0.16	0.073+	-	-	-
	(0.21)	(0.039)	-	-	-
Life-Threatening Event Rate	0.07	0.65	0.89	-0.92	-2.40
	(0.31)	(0.55)	(0.83)	(0.64)	(1.83)
Hospitalization Rate	0.25	2.38+	3.07	1.39	-3.48
	(0.84)	(1.27)	(1.94)	(1.16)	(3.72)
Mortality Rate	0.08	-1.21	1.08	-0.07	0.26
	(0.46)	(2.21)	(0.68)	(0.59)	(2.53)
Sample Size		616	672	552	38472
B. Class II to I:					
Emphasis on Safety	0.065	0.05***	-	-	-
	(0.218)	(0.012)	-	-	-
Life-Threatening Event Rate	0.07	-2.18	-0.36+	-3.24*	-3.18*
	(0.43)	(2.02)	(0.19)	(1.63)	(1.56)
Hospitalization Rate	0.17	-2.05***	-3.04+	-4.87*	-5.44*
	(0.94)	(0.60)	(1.56)	(2.35)	(2.54)
Mortality Rate	0.26	-0.43**	-0.27	-0.46+	-0.57*
	(2.13)	(0.14)	(0.20)	(0.26)	(0.27)
Sample Size		10332	13104	17668	20664

Table 1.5. Costs and Benefits of Down-Classification. Note: This table provides the back-of-the-envelope calculations of the costs and benefits of Class III to II and Class II to I down-classification events. Assumptions are detailed at the header of the table. Patent estimates are calculated using only publicly traded companies for which I can obtain patent values as calculated in Kogan et al. (2017). I provide 95% confidence intervals for the costs and benefits. Costs and benefits are annualized and averaged at the device type level (as defined by the FDA). The column "Value" is the value per unit of the estimate. In my data, procedures using treated medical device types generate, on average, $26,849 a year of health expenditures. Scaling this total to a national level ($26,849/ 0.0008, where .0008 is the share that UCSDH executes) gives roughly $33 million a year spent per treated procedure, on average. This total is similar to the average yearly cost of medical procedures seen when Medicare data is scaled to national expenditures, at $34.7 million a year per procedure. Since I find that costs, as measured by paid amounts, decrease by 44–62% a year, I use these percentage decreases in prices to calculate annual national expenditure changes per treated medical device type. These calculations are presented in the "Prices" row for Class III to II down-classifications. "Patented Inn." represents innovation that is patented by public firms, and "Hospital." represents hospitalizations.

Assumptions

-Cost of mortality is EPA's VSL of $10 million.
-Average inpatient hospital stay costs $22,000. No other costs.
-Creative destruction of 4/5 from value of patents.
-Do not consider private firm patent values.
-Do not consider scientific value of innovation.
-No value of efficacy information from regulations.
-No value from firm entry (e.g., not considering value of new jobs).
-UCSDH performs .08% of total U.S. procedures (calculated from data).

		Outcome	Estimate	95% C.I.	Value	Total	95% C.I.
Class III to II	**Costs**	Mortality	1.08	[-0.3,2.4]	$10m	$10.8m	[-$3m, $24m]
		Hospital.	2.38	[-0.1,4.9]	$.02m	$.05m	[$0m, $0.1m]
						$10.9m	[-$3m, $24m]
	Benefits	Patented Inn.	5	[3.2,8.1]	$13m/5	$13m	[$8.2m, $21.1m]
		Prices	-$14.7m	[-$2.6,-$26.8]	-1	$14.7m	[$2.6m, $26.8m]
						$24.7m	[$11m, $48m]
Class II to I	**Costs**	Mortality	-0.43	[-0.7, -0.16]	$10m	-$4.3m	[-$7m, -$1.6m]
		Hospital.	-2.1	[-3.3, -0.9]	$0.02m	-$0.04m	[-$0.06m, $0]
						-$4.3m	[-$7m, -$1.6m]
	Benefits	Patented Inn.	9	[3.1, 14.9]	$10m/5	$18m	[$6m, $30m]
						$18m	[$6m, $30m]

1.8 Chapter Acknowledgments

Chapter 1, in full, is currently being prepared for publication of the material. The dissertation author was the sole investigator and author of this paper.

Chapter 2

Demand Shocks, Procurement Policies, and the Nature of Medical Innovation: Evidence from Wartime Prosthetic Device Patents

From 1960 to 2019, U.S. health spending rose from 5 to nearly 18 percent of GDP. Research has documented that the advance of medical innovation underlies a substantial share of this cost growth (Smith et al. 2009, Cutler 2004), which raises a variety of questions. First, what factors drive the volume of medical innovation? Second, what leads inventors to focus on reducing costs (e.g., by streamlining production processes) versus improving quality? More generally, what factors shape the specific problems with which medical innovators choose to engage?

Wars and pandemics, among other events, can create acute needs for medical innovation. The COVID-19 pandemic, for example, generated demand for new vaccines, new diagnostic tests, testing infrastructure, and personal protective equipment. The value of new vaccines is widely recognized. Improvements in medical equipment, reductions in production costs, and expansions in productive capacity can also have substantial value when demand rises sharply. This motivates us to study how demand shocks and procurement environments shape the volume of medical innovation, its emphasis on the production process, and its emphasis on dimensions of product quality.

We analyze the effects of demand shocks and procurement environments on the quantity of medical innovation and the product and production process attributes it emphasizes. Our empirical analysis considers two important periods in the history of prosthetic device innovation: the U.S. Civil War and World War I. We begin by presenting key details of these historical contexts, including differences in demand, differences in procurement incentives, and differences in the stated goals of the public procurers. We show that both the Civil War and World War I led to substantial increases in prosthetic device patenting. A point of contrast is that the Civil War led to a much greater focus on cost-conscious innovation while World War I did not. To the best of our knowledge, this analysis provides the first evidence that cost-conscious procurement environments can indeed steer medical innovation in a cost-conscious direction.

Empirically assessing how incentives shape the emphases of inventors requires overcoming two primary challenges. First, existing data sources that categorize patents or clinical trials do not provide information on an invention's detailed economic attributes. Extracting this information requires going deeper into an invention's details. Second, linking procurement environments to the specific attributes on which inventors focus requires analyzing settings across which those environments exhibit variation.

To gain insight into how inventors advanced the frontier of prosthetic device technology, we use machine learning tools to construct a novel data set. We begin by closely reading 1,200 patents from the periods surrounding the U.S. Civil War and World War I. Our selection comprises prosthetic device patents and patents from other medical and mechanical technology classes. Based on these close readings, we code variables describing the economic traits emphasized in each patent. These variables include three traits that we interpret as production-process attributes, three traits that capture distinctive dimensions of product quality, and two additional traits that are less clearly defined as quality or production process traits. We then use machine learning tools to extend our data set to include a much larger set of patents.

The U.S. Civil War and World War I generated dramatic increases in demand for artificial limbs, as amputations were remarkably common. The associated public procurement

environments created incentives that differed across the two wars. Our empirical analysis of these episodes includes a combination of time series and difference-in-differences methods. In the time series analysis, we directly examine changes in prosthetic device patents. In the difference-in-differences analyses, we use patents from other medical and mechanical technology classes to construct control groups.

Our first result quantifies the effects of the Civil War and World War I on the quantity of prosthetic device innovation. For several years during these historical episodes, prosthetic device patenting rose by nearly 100 log points relative to patenting in our control groups. Despite analyzing only two events, the relative increases in prosthetic device patenting are strongly statistically distinguishable from zero. Our evidence from patents filed with the U.S. Patent and Trademark Office (USPTO) is supplemented by patents from the short-lived Confederate patent office, as well as from the British and Spanish patent authorities.[1]

For the Civil War period, we have sufficient information to infer an elasticity of innovation with respect to potential revenues. We estimate an elasticity on the order of one for both patenting and firm entry; this is higher than typical estimates of long-run elasticities of medical innovation with respect to long-run changes in market size (Dubois et al. 2015). Innovation may respond more rapidly to crisis-driven shocks than to standard changes in market size, as Agarwal and Gaule (2021) have observed in the context of the COVID-19 pandemic.

Second, we find that the demand shock associated with the Civil War generated substantial effort to reduce the cost of producing prosthetic devices. During the Civil War, the average prevalence of production process traits doubled in prosthetic device patents but was essentially flat within other technology classes. There was a far more modest shift towards production process traits during World War I. The Civil War era shift towards cost-oriented innovation is consistent with an important role for procurement incentives. As discussed in section 2.1, the U.S. government's Civil War era procurement program involved modest, fixed-price payments

[1] In the British patent data, we see a large increase in prosthetic device patenting during World War I and no increase during the U.S. Civil War. Spain participated in neither conflict and the Spanish data exhibit no increase in prosthetic device patenting.

to artificial limb manufacturers, which can create strong incentives for innovation to reduce production costs.[2] As further suggestive evidence for the role of procurement incentives, we show that patents for artificial arms, for which profit margins were lower than for artificial legs, exhibit a more substantial shift in emphasis towards cost reduction during the Civil War.

Third, the prosthetic device patents of the Civil War and World War I diverged with respect to dimensions of quality. Civil War-era prosthetic device patents exhibit a substantial increase in emphasis on comfort. By contrast, World War I-era prosthetic device patents de-emphasize comfort and place greater emphasis on occupation-oriented "appliances." The latter shift connects quite directly to the historical narrative, which highlights an emphasis of governments and medical professionals on the re-employment of veterans with amputated limbs. Civil War and World War I-era differences in emphasis on comfort are plausibly linked to a World War I-era shift in choice away from veterans and toward medical professionals. As detailed below, the historical narrative provides validation for the channels through which the Civil War and World War I-era procurement environments may have altered these dimensions of inventor effort.

Our analysis adds to a broad line of research on the effects of potential profits on innovation. This includes labor economics applications (Acemoglu 1998, Hémous and Olsen 2022) as well as a substantial environmental economics literature summarized by Popp (2010; 2019). In the context of health care, research on the effects of potential profits on innovation has focused primarily on pharmaceutical innovation (Finkelstein 2004a, Acemoglu and Linn 2004b, Budish et al. 2015).[3] Exceptions include analyses of medical equipment and device patenting

[2] With fixed prices set moderately below baseline costs, for example, sales are not profitable until manufacturers find ways to reduce production costs. More generally, even when the fixed price exceeds cost, a lower baseline profit per unit increases the returns to innovating to reduce cost relative to the returns to innovating to increase market share by increasing quality.

[3] Additional papers include Blume-Kohout and Sood (2013a), who find that research on drugs with high Medicare market shares rose following the introduction of Medicare Part D, Yin (2008), who finds positive effects of the Orphan Drug Act, Dubois et al. (2015), who find that potential profits affect the number of new molecular entities that come to market, and Agarwal and Gaule (2021) who study medical innovation in the context of the COVID-19 pandemic.

by Clemens (2013) and by Galasso and Luo (2017; 2022).[4] We contribute to this literature by providing novel evidence on the effects of large demand shocks on prosthetic device innovation. We additionally provide evidence that innovation may respond more aggressively to crisis-driven shocks than one would infer on the basis of long-run elasticity estimates.

We also contribute to the literature on medical innovation by applying text analysis methods to gain insight into innovators' emphases on cost versus dimensions of product quality. Analyses of patent texts have become increasingly common in the innovation literature.[5] We apply text analysis methods to develop the novel data required to make progress in understanding whether procurement environments can shape the particular dimensions of the technical frontier on which inventors focus. Methodologically, we develop several practical insights into best practice methods for this class of machine learning applications. The substance of our findings provides evidence that cost-conscious procurement environments can indeed steer medical innovation in a cost-conscious direction.

The paper proceeds as follows. Section 2.1 provides historical background and section 2.2 summarizes the hypotheses that are motivated by our historical settings. Section 2.3 discusses our novel data set and section 2.4 our empirical strategy. Section 3.4 presents our results and section 2.6 concludes.

2.1 Civil War and World War I Demand for Artificial Limbs

The U.S. Civil War and World War I were both associated with dramatic increases in demand for prosthetic devices. In this section, we begin by describing the size of these demand shocks. We then provide background on the relevant systems for rehabilitating veterans and procuring artificial limbs.

[4]Clemens (2013) studies medical equipment patenting surrounding the introduction of Medicare. Galasso and Luo (2017) study the effects of tort reform on medical equipment and device innovation, while Galasso and Luo (2022) study the effects of liability risks faced by the suppliers of medical implants.

[5]See, for example, Khoury and Bekkerman (2016), Bergeaud et al. (2017), Iaria et al. (2018), Watzinger and Schnitzer (2019), Arts et al. (2018), Cockburn et al. (2018).

2.1.1 The Magnitude of Wartime Demand Shocks

The U.S. Civil War was contested between the armies of the Union and the Confederacy from April 1861 to May 1865. An estimated 35,000 veterans with amputated limbs survived the war on the Union side alone (Linker 2011; p. 98). Because the government had not formed a permanent bureaucracy for addressing veteran health care needs prior to the war, both the Union and Confederacy implemented ad hoc artificial limb procurement systems as the scope of need became clear. Wartime production levels (Barnes and Stanton 1866, Hasegawa 2012) far exceeded pre-war production as documented in the 1860 Census of Manufacturing. In developing our evidence of the effects of Civil War-era demand on innovation, we draw primarily on patents filed with the USPTO, but also consider patents filed with the short-lived Confederate patent office.

World War I produced an estimated 300,000 veterans with amputated limbs worldwide. Relative to the Civil War, demand associated with 4,000 U.S. veterans was relatively modest. Because production capacity was low among the European powers and high in the United States, the U.S.-based artificial limb industry played an important role in satisfying global demand. Great Britain, for example, which was home to an estimated 41,000 surviving veterans with amputated limbs (Guyatt 2001; p. 98), invited the largest American prosthetic companies "to set up workshops at the main amputee center" (Linker 2011; p. 99). In developing our evidence of the effects of World War I-era demand on innovation, we study patents from both the United States and Great Britain.

2.1.2 Background on Civil War and WWI-Era Procurement

During the Civil War, the manufacturers of artificial limbs faced a competitive environment in which they were reimbursed on a "fixed-price" basis. To become eligible for purchase through the Union's limb allowance program, artificial limb models had to be certified by a board

of physicians.[6] If the board deemed a prototype to be "serviceable," its manufacturer entered the list of manufacturers from which soldiers could select the provider of their artificial limb. Fixed-price reimbursements were set at modest levels relative to manufacturers' stated costs from the pre-war period, and balance billing was prohibited (Hasegawa 2012; p. 37-38).[7]

By World War I, the U.S. had substantively formalized the treatment of veterans with amputated limbs. This occurred within a broader effort to formalize veterans' health care. In addition to being formalized, care for veterans with amputated limbs was mostly centralized at large facilities, including the recently built Walter Reed Hospital.[8]

Progressive Era policymakers worried that veterans with amputated limbs would, like many of their Civil War predecessors, fail to return to gainful employment. A perception of limbless Civil War veterans "pocketing" their allowances and opting out of the labor force impacted World War I-era views regarding care and rehabilitation (Linker 2011). As Linker (2011; p. 13) writes, "The veterans of America's First World War were expected to become citizen-workers once their military service was over; they were to make useful lives, not to languish at the expense of the US Treasury."

Between the Civil War and World War I, discretion in the choice of artificial limb shifted from veteran to government. During World War I, veterans underwent extensive rehabilitation prior to their return to civilian life, including obligatory use of standard-issue prosthetic limbs. Linker (2011; p. 101) writes that "the OSG [Office of the Surgeon General] forcefully mandated artificial limb wear, creating legislation that made it virtually impossible for US amputee soldiers

[6] As Hasegawa (2012) documents, General William Hammond convened a panel of physicians to, in Hammond's words, "determine what kind of Artificial Limbs should be adopted for the use of mutilated soldiers."

[7] During the latter half of the war, the price for artificial legs was set at $75 (roughly $1,500 in 2018 dollars) and the price for artificial arms was set at $50. A small number of products were authorized for sale at higher rates (Hasegawa 2012; p. 40). In such cases, the veteran was responsible for the difference between the approved price and the government's allowance of $75 per leg or $50 per arm. These products were meant to be sold at the approved prices on a fixed rate basis with no balance billing. Hasegawa (2012) documents that a leading manufacturer told the government his costs were $150 per artificial leg.

[8] Treatment of veterans with amputated limbs also took place at Letterman hospital in San Francisco. As Linker (2011; p. 80) writes, "Surgeon General Gorgas designated two general hospitals to become permanent installations for rehabilitative care: Letterman General Hospital in San Francisco and Walter Reed General Hospital in Washington. Later in the war, the list of military rehabilitation hospitals would grow to 14, but Letterman and Walter Reed remained the flagship facilities during and after the war."

to be discharged from military service without months of rehabilitation and daily routine artificial limb wear." In contrast with the Civil War, demand for artificial limbs was thus shaped to a significant degree by the veterans' medical bureaucracy and to a lesser degree by wounded veterans.

The incentives facing artificial limb manufacturers were shaped by the preferences of World War I-era medical bureaucracies in both the U.S. and Europe. While we cannot know the precise criteria each bureaucracy used in their procurement of artificial limbs, the historical record provides clues regarding approaches to rehabilitation. Medical professionals of the World War I-era de-emphasized comfort in favor of a strict rehabilitation program. Linker (2011; p. 109-114) writes, for example:

> Once surgical healing had been attained... the 'toughening' of the stump by 'pounding it on a firm surface' should be 'vigorously pursued'... Following stump pounding exercises, 'patients usually complained of discomfort'... Another report stated that when amputees were forced to wear artificial limbs soon after surgery, they often 'expressed gratitude when the artificial limb [was] removed.'

In addition to driving a relatively severe program of physical rehabilitation, the desire for social reintegration spurred an emphasis on re-employment. The British government had similar views on the importance of rehabilitation and re-employment.[9] The historical record thus suggests that World War I-era procurers placed substantial emphasis on artificial limbs' capacity to restore an individual's employability.

2.2 Implications of Wartime Demand Shocks for Innovation

We draw on the historical narrative regarding Civil War and World War I-era demand shocks and procurement environments to develop hypotheses regarding the potential effects of these events on prosthetic device innovation. The hypotheses motivated by the historical record are as follows:

[9] See, for example, the discussions of British World War I-era rehabilitation and artificial limb manufacturing in Novotny (2017) and Guyatt (2001).

First, the large demand shocks associated with both the Civil War and World War I increased incentives for developing novel prosthetic devices. The hypothesis that these demand shocks would increase flows of innovation is perhaps the most standard hypothesis in the literature on demand-driven innovation.

Second, the Civil War-era procurement environment featured a low, fixed-price reimbursement regime. We hypothesize that this regime may have generated an increase in inventor emphasis on cost-conscious innovation. This hypothesis is linked in part to the fact that production costs must be driven below the reimbursement level before sales become profitable.

Third, we hypothesize that the emphasis of World War I-era procurers on the reemployment prospects of wounded veterans may have increased inventor emphasis on the capacity for artificial limbs to enhance their wearer's social reintegration and employability. Social reintegration could be facilitated by limbs that more faithfully mimicked the appearance of a natural limb. Employability could be facilitated by a line of artificial limb technology we call "appliances." In this context, the word "appliances" refers to interchangeable artificial limb attachments which serve functions that connect directly to occupational tasks.

Fourth, we hypothesize that the Civil War-era procurement environment may have increased inventors' emphasis on characteristics demanded by veterans, who could choose across products, while the more centralized World War I-era procurement environment prioritized the preferences of the veterans' medical bureaucracy. This final hypothesis has less precise empirical content than hypotheses one through three. It may be relevant to such traits as an artificial limb's comfort and appearance.

2.3 Patent Data and Text Analysis Methods

We begin this section with a discussion of the historical patent data we use to estimate the effects of wartime demand shocks on overall patent flows. We then discuss the new data we generated through text analysis (or natural language processing) using a combination of close

readings and machine learning techniques.

2.3.1 Historical Patent Data

The first question we attempt to answer is if wartime increases in demand for prosthetic devices increased the rate of prosthetic device patenting. This analysis requires information on 19th and early 20th century patents by technology class. Until relatively recently, the patent data sets analyzed by economists did not facilitate this type of historical analysis. The groundbreaking NBER patent database (Hall et al. 2001), for example, begins with patents granted in 1963. Economists have recently developed databases extending to the earliest surviving records of the U.S. Patent and Trademark Office (USPTO). To identify historical patents based on their technology classes, we use the database assembled by Berkes (2018).[10] We supplement these data with additional data on Confederate patents, British patents, and Spanish patents.[11]

One shortcoming of the Civil War era patent data is that, before 1873, patents reported the date the patent was issued, but not the date it was filed (Berkes 2018). Consequently, we organize patents according to their date of issuance throughout our analysis. Patents from 1873 onward allow us to gauge the typical lag between patent filing and issuance during the period we analyze. From 1873 through the end of our World War I sample, the average lag between filing and issuance was 1.2 years for the full set of technologies we analyze and just over 0.9 years for prosthetic devices.[12] We test whether indexing by patent issuance dates changes our findings relative to indexing by filing dates using data from the World War I era. We find that the time series for both our treatment and control classes are shifted forward by roughly one year when indexed by patent filing year, as shown in panels A and B of Figure B26. This has little influence

[10] In a comparison of several recent efforts to compile data sets on the universe of U.S. patents, Andrews (2019) concludes that the database laid out in Berkes (2018) is "currently the gold standard." Additional analyses of 19th and early 20th century patents, including those by Berkes and Nencka (2019) and Berkes et al. (2019) have been made possible by these data.

[11] Sáiz (2000) and Sáiz et al. (2008) generously provided Spanish patent data.

[12] In the technology classes we analyze, the average lag between filing and issuance has exceeded three years during the 21st century. Lags between filing and issuance have thus been much longer in recent years than during our sample.

on our reading of the evidence.

Figure 2.1 provides an initial look at time series on prosthetic device patents and other broad categories of patents during the historical episodes we analyze. The dashed vertical lines in each panel encompass the years we subsequently associate with war-induced booms in prosthetic device patenting. It is quite clear from the panels of Figure 2.1 that both the Civil War and World War I were associated with substantial increases in the rate of prosthetic device patenting among combatant nations (i.e., the United States during the Civil War and World War I, the Confederacy during the Civil War, and the United Kingdom during World War I), but not among non-combatant nations (i.e., the United Kingdom during the U.S. Civil War and Spain during both the U.S. Civil War and World War I). However, quantifying the causal effect of wartime demand shocks requires constructing counterfactuals, which we discuss in section 2.4.

There are limitations when using patent counts to measure innovation. Primarily, patent counts do not necessarily measure changes in meaningful innovation. Thus, during the period surrounding World War I, we follow standard practice in the literature by using citations as a proxy for patent quality. As shown in Panel B of Figure B24, the average number of citations per patent was fairly stable during World War I, suggesting that the prosthetic device patent boom was associated with patents of similar impact as the pre-war patents. Citation measures of quality for Civil War patents are less reliable. As described by Berkes (2018), 19th-century patents have less complete and noisier citation data. Panel A shows that, during the Civil War period, the sparsity of citation data likely renders this exercise uninformative. To validate the quality of Civil War era patents, we look to information reported in Tables 2.1 and 2.2, which we describe below in detail.

Several features of the Civil War period allow us to establish that changes in patenting connect to real industry responses. The most striking point is that we directly observe the entry of new manufacturers. Further, as reported in Table 2.1, we are able to establish links from patents to manufacturers, from manufacturers to sales through May 1866, and from both sales

and manufacturers to expert assessments of quality.[13] Twelve out of the thirteen most notable manufacturers of artificial legs and eight out of the nine most notable manufacturers of artificial arms from the Civil War period can be linked to at least one patent. Through May 1866, these patent-holding manufacturers accounted for nearly all of the artificial legs and nearly 90 percent of the artificial arms furnished to Union Army veterans. As shown in Table 2.2, contemporaneous sources reveal a dramatic increase in the number of artificial limb manufacturers, artificial limbs produced, and the total value of artificial limb output during the U.S. Civil War. Finally, medical histories document that these episodes were, in fact, episodes of substantial advance in artificial limb technologies.[14]

2.3.2 Coding Patent Attributes

Beyond measuring patent flows, our analysis aims to understand the economic attributes that are emphasized in each patent. We pursue this to understand how inventors distributed their efforts across improving aspects of production processes and/or particular dimensions of each product's quality. Because the data required for this analysis did not previously exist, we developed a novel data set.

Our data set contains information that quantifies the economic attributes emphasized in historical patent documents. To generate this information, we first created a program to scrape historical patent documents from Google Patents. Using the text of each patent document, we then coded a set of product and/or production process attributes on which the patent places emphasis. We describe three of these attributes, namely cost, simplicity, and adjustability, as cost-oriented

[13] A limitation of this analysis is that we can only estimate market shares for the 6,075 artificial limbs documented in Barnes and Stanton (1866). Because this memorandum was submitted on May 11, 1866, it cannot document market shares for artificial limbs delivered after that time.

[14] Post- and late-war rankings of artificial limbs by quality further support a link between quality and market share (Barnes 1865, Houston et al. 1866). The top three rated artificial legs accounted for just under 60 percent of sales through May 1866, while the top four rated artificial arms accounted for just over 60 percent of sales through May 1866. The highly-rated limbs with low market shares were those developed relatively late during the war, namely the artificial arms of John Condell and the National Arm and Leg Company. The low market shares we observe for these limbs in sales through May of 1866 are thus largely mechanical, as they were not on the market when most of the limb purchases for which we have documentation occurred. Low-rated limbs with non-trivial market share tended to be either unpatented or to involve pre-war patents, suggesting an incumbency advantage.

production process traits. That is, these traits involve aspects of a product's production. We use the term "adjustability," for example, to describe patents that emphasize uniform production of outputs that can subsequently be fitted (or "adjusted") to the needs of a specific consumer. Three traits, namely comfort, appearance, and occupation-oriented appliances, are quality-oriented attributes. We also code two additional traits, namely materials and durability, that we have not explicitly labeled as either product or production-process traits.

Table 2.3 presents a concise verbal definition of each economic attribute. The table also summarizes three important aspects of each attribute related to the quality of the information we capture with each variable. The first aspect, summarized in column 3, is the strength of the linkage between each trait and the hypotheses we have generated based on the historical record (i.e., the hypotheses laid out in section 2.2). The second aspect, summarized in column 4, is our assessment of the extent to which our text analysis procedure generated a variable that successfully captures the economic content we sought to capture. The third aspect, summarized in column 5, is our assessment of the challenges associated with identifying comparison technology classes to construct control groups for our analysis of a given trait.

How successfully can the variables we generate capture the intended economic content of patents? A key point regarding this important methodological question is that the difficulty of identifying economic concepts in text can vary substantially from concept to concept. In the remainder of this section, we illustrate the underlying issues with a small number of examples. Appendices B1 and B2 provide substantially more detail.

Some economic concepts are straightforwardly conveyed in text. We found this to be true, for example, of the traits cost and simplicity. One patent, for example, describes the mechanism underlying an artificial knee joint as having "great simplicity, and therefore cheapness." A second states "The object of my invention is to imitate this eccentric motion of the knee-joint in the simplest manner." For both simplicity and cost, there is little difference between the performance of our close readings, our fully refined machine learning model, and a straightforward keyword search.

Other concepts are more inherently difficult to track in text than cost or simplicity. Tracking the use of new materials, for example, proved difficult because establishing a set of keywords requires knowing what materials are common and what materials are newly introduced in manufacturing products in a given technological class. These difficulties are sufficiently severe that we place little emphasis on our findings for the "materials" trait.

Other traits can capture clear and distinctive technological developments despite being very specific to a particular technological class. The trait we term "appliances" exemplifies this third scenario. As illustrated through a set of examples, occupation-oriented "appliances" were a critical, clearly defined dimension of prosthetic device innovation during World War I. This dimension of prosthetic devices, however, does not have a strong analogy in other technology classes. This fact casts doubt on the potential utility of constructing a control group for analyses of such a trait, as conveyed by our designation of appliances as "weak" in column 5 of Table 2.3. For a trait like "appliances," evidence from simple time series differences may be more informative than analyses that incorporate counterfactuals based on other technology classes.

2.3.3 Text Analysis

This section provides an overview of the text analysis tools we developed and implemented. Appendix B2 describes these tools in greater detail and underscores several best practices to consider when generating variables with machine learning algorithms.

Our text analysis methods can be concisely described as the output of a keyword search that has been informed by domain-specific knowledge and enhanced by machine learning tools. We developed domain-specific knowledge by closely reading just over 1,200 patent documents. While reading these patents, we completed two tasks. First, we form the data set to train our machine learning model by indicating whether each patent has specific attributes. Second, we construct the initial sets of keywords that we associate with each of the attributes.

The set of closely-read patents (i.e., the "training set") covers the domains relevant to our analysis. That is, our training set includes patents from both the prosthetic device class and

candidate control classes, as well as from both the Civil War and World War I-eras. To achieve this coverage, we randomly selected our sample of closely-read patents after stratifying across technology classes and war episodes. As summarized in Table B31, the manually coded data set contains 195 prosthetic device patents and 399 other medical or mechanical patents from the Civil War period, as well as 302 prosthetic device patents and 305 other medical or mechanical patents from the World War I period.[15]

Our text analysis task faces a common problem of dimensionality. With just over 1,200 patents in our training set, algorithms will perform poorly if we attempt to use every word from every patent document as an input. We thus implement an approach to limit the algorithm's attention to the most relevant words, or "features," in each patent document's text.[16] The features we selected are a set of keywords, synonyms, and a small neighborhood of textual context surrounding the keywords and synonyms (see appendix B2 for more details). We developed our initial lists of keywords based on our 1,200 closely read patents. We next augment these keywords with synonyms that appear in similar linguistic contexts, which we selected using the"Word2Vec" algorithm (Mikolov et al. 2013). Finally, to aid our algorithm in identifying context-specific word meanings, we gather a "spread" of contextual words surrounding the appearance of each keyword. Our augmented set of keywords and their accompanying contextual "spread" are the features from each patent that we use as inputs into our machine learning model. After training and validating our model, we use the model to extend our encodings to roughly 750,000 patent texts that span our treatment and control groups.

[15]The attribute "appliances" is an exception. The relevance of occupation-oriented appliances was drawn to our attention by a referee in August 2021, which was several years after we completed the close readings underlying the coding of other traits. Our coding of appliances is thus based on a keyword search that is informed by close readings of a smaller number of patents.

[16]This approach, which is called "feature selection," has been shown to improve the efficiency of predictive models (Guyon and Elisseeff 2003). The familiar Lasso procedure, for example, limits the number of features in the model by applying a penalty factor within its objective function.

2.3.4 Novel Data Set on Patent Attributes

Our final data set, produced by our machine learning approach, describes the economic attributes of 745,558 patents, with the earliest coming from 1840 and the latest from 1940. There are 814 prosthetic device patents, 19,666 other medical patents, and 725,078 mechanical patents. Our regression analyses focus on samples of our 745,558 patents for which the patent year is in relatively close proximity to each conflict. These samples extend from 1855 to 1867 and from 1910 to 1922.

Across this large set of patents, appendix Table B33 shows that the economic traits we coded are only modestly correlated with one another. The primary exceptions are cost and simplicity. Among prosthetic device patents, cost and simplicity share a correlation of 0.378 with an associated r-squared of 0.142. Similarly, across all patents in our data set these traits share a correlation of .303 with an associated r-squared of 0.092. Correlations across all other trait pairs are between -0.12 and 0.13, highlighting that the traits capture independent dimensions of innovation.

2.4 Empirical Strategy

We now present our specifications for analyzing changes in patenting rates and in the economic characteristics emphasized in patent documents. After presenting each estimation framework, we highlight the key challenges we face when attempting to generate causal estimates of the effects of wartime demand shocks.

2.4.1 Analyzing Patent Counts

We begin by estimating the effects of the Civil War and World War I on patent counts using the regression equations below. The first is specified as an Ordinary Least Squares model for predicting the log of patents per year:

$$ln(N_{t,c}) = \alpha_{c,w(t)} + \alpha_t + \beta_1 1\{\text{War}\}_t \times 1\{\text{Prosthetic}\}_c + \varepsilon_{c,t}. \quad (2.4.1)$$

The second is specified as a Poisson model of patent counts:

$$E[N_{t,c}|X_t] = exp(\gamma_{c,w(t)} + \gamma_t + \beta_1 1\{\text{War}\}_t \times 1\{\text{Prosthetic}\}_c + \varepsilon_{c,t}). \quad (2.4.2)$$

In both equation (2.4.1) and equation (2.4.2), c denotes patent classes, t denotes time (multi-year time periods for these specifications), and $w(t)$ denotes war episodes (Civil War and World War I). $N_{t,c}$ denotes the number of patents in class c at time t. The specifications include time fixed effects (α_t or γ_t) and episode-by-patent class fixed effects ($\alpha_{c,w(t)}$ or $\gamma_{c,w(t)}$). The coefficient of interest is β_1, which is an estimate of the differential change in the patenting rate for prosthetic devices relative to the control classes during war episodes relative to pre-war periods. The periods over which the wars influenced prosthetic device patenting are defined to extend from 1862 to 1866 for the Civil War and from 1916 to 1922 for World War I.

The key challenge in developing causal estimates is to construct control groups that approximate the counterfactual development of patenting rates for prosthetic devices. Technology classes might generate inappropriate counterfactuals for a variety of reasons. They might, for example, be affected by very different sets of scientific developments (e.g., nuclear technology vs. prosthesis). Alternatively, a plausibly comparable technology class will be a poor control class if it is directly affected by wars (e.g., firearms) or if it is shaped by spillovers from prosthetic device innovation.

Our selection of a complementary set of control groups follows the logic of Finkelstein (2004a), whose analysis of vaccine clinical trials is analogous to our setting in some key respects. The patents we use to construct control groups come from broad categories of medical and mechanical innovations. In all analyses, we exclude technology classes for which there was one or fewer patents per year within the time periods into which we divide the data. Our largest control group incorporates all medical and mechanical technology classes that meet this criterion.

We also consider sub-groups chosen to either increase comparability or reduce the likelihood that the control group contains patent classes that could be directly affected by the wars. Like Finkelstein (2004a), we also consider data-driven control groups. For our analysis of patent flows, the data-driven approach selects the control group to match baseline flows of prosthetic device patents in levels.

2.4.2 Analyzing Patent Traits

Our analysis of the traits emphasized by wartime prosthetic device patents confronts challenges that differ from the challenges facing our analysis of patent counts. The variables of interest in this analysis describe the share of patents within a given technology class and time period that emphasize the characteristic of interest:

$$\text{Category Trait Share}_{period} = \frac{\text{\# Category Patents with a Trait}_{period}}{\text{\# Category Patents}_{period}}.$$

For our analysis of patent traits, it is less clear what might constitute a reasonable control group. It may simply be less relevant, for example, to worry that the traits emphasized by prosthetic device patents will shift markedly for reasons unrelated to the wartime demand shocks on which our analysis focuses. As an initial estimator, this leads us to consider simple time series changes among prosthetic device patents:

$$\beta^{TS} = [\text{Prosth. Trait Share}_{wartime} - \text{Prosth. Trait Share}_{prewar}] \tag{2.4.3}$$

This is captured by β^{TS} from equation (2.4.3).

We also consider difference-in-differences estimates, which net out changes in the emphasis on a given trait among the patents within a control group. For analyses of this sort,

selecting control groups is non-trivial because some traits of interest are only relevant to a small set of the technology classes within our broadest control group. As shown in Table B32, for example, this is true of traits including "appearance" and "comfort." This leads us to select control groups using several complementary approaches, which include the construction of synthetic control groups as well as a simple matching procedure.[17] We discuss additional aspects of our application of the synthetic control procedure in Appendix B3. The resulting estimator takes the form below:

$$\begin{aligned}\beta^{DD} = &[\text{Prosth. Trait Share}_{wartime} - \text{Prosth. Trait Share}_{prewar}] \\ &- [\text{Other Trait Share}_{wartime} - \text{Other Trait Share}_{prewar}],\end{aligned} \tag{2.4.4}$$

We interpret our findings as being robust if we obtain similar results whether we rely on the time series variation, as in equation (2.4.3), or any of several plausible difference-in-differences strategies, as in equation (2.4.4).

2.5 Results

This section presents estimates of equations (2.4.1), (2.4.2), (2.4.3), and (2.4.4). Subsection 2.5.1 presents estimates of the effects of the Civil War and World War I demand shocks on flows of prosthetic device patents. Subsection 2.5.2 discusses the magnitudes of our estimates. Subsections 2.5.3 and 2.5.4 present estimates of changes in the attributes emphasized in prosthetic device patents during the wartime patent booms relative to the pre-war periods.

[17] When implementing the synthetic control approach for our Civil War sample, patent flows for many technology classes were limited, including prosthetic devices. In each of 1858 and 1861, for example, there was a single prosthetic device patent. The maximum across the pre-Civil War years was seven, which occurred in 1859. The share of patents emphasizing a given trait is thus highly volatile across the Civil War baseline when expressed at an annual frequency. Matching year-to-year trends would amount to matching noise. For our baseline method, we thus match levels and trends in four-year moving averages. As a natural robustness check, we have confirmed that our results are little changed by matching levels and trends on either three-year moving averages or five-year moving averages.

2.5.1 Overall Patent Flows

Table 2.4 presents estimates of equation (2.4.1). The estimates presented across the columns differ exclusively with respect to the patent classes used as controls. The estimate in column 1 reveals that wartime changes in prosthetic device patenting were roughly 95 log points larger than changes in patenting in all other medical or mechanical patent classes. Columns 2 through 7 reveal that this estimate is only moderately sensitive to using subsets of the broader set of controls. The subsets include other categories matched based on baseline patenting rates (column 2), other medical categories only (column 3), the "miscellaneous" mechanical classes (column 4), metalworking mechanical classes (column 5), materials processing mechanical classes (column 6), and all classes except those that would be plausibly affected by wartime demand shocks (column 7).[18] The estimates range from 85 log points to 102 log points. Panels B and C reveal substantial increases in prosthetic device patenting during each war episode, with economically larger increases occurring during the Civil War than during World War I.

Appendix B4 provides additional evidence relevant for interpreting these findings. First, Table B30 presents estimates of the Poisson model described by equation (2.4.2). Second, Figure B21 presents an "event study" analysis, which provides evidence against the concern that wartime increases in prosthetic device patenting were driven by pre-existing trends. Third, Figure B22 illustrates why, despite having only two class-by-time period treatment events, the wartime increases in prosthetic device patenting are nonetheless strongly statistically distinguishable from zero when we conduct inference using "randomization tests" (Imbens and Rosenbaum 2005). Each observation underlying Figure B22's histograms represents the change in patenting

[18] Our restriction of the control group to other medical technology classes (column 3), is similar to the approach taken by Moser et al. (2014) in their analysis of chemicals patenting. We obtain similar, though modestly smaller, results when further narrowing our control group to the sub-category "Miscellaneous-Drugs and Medicine," which also contains Prosthesis innovation. This sub-category is quite small during these periods, however, as it comprises only two other classes, namely "Optics: Eye Examining, Vision Testing and Correcting" and "Dentistry." A further issue facing this approach to selecting control classes is that optics and dentistry are medical categories for which it is plausible that the Civil War and World War I may have had a direct effect. This may contribute to why we obtain moderately smaller point estimates when using these control classes rather than a broader control group. For details, we refer readers to the descriptions of the technology classes that are available on the website for the NBER patent database: http://www.nber.org/patents/.

in a patent class in our broadest control group. The dashed vertical lines are placed at the value of the change for prosthetic devices. In the Civil War histogram (Panel A), the change in prosthetic device patenting is the rightmost point in the distribution; this underlies the uniformly low p-values in Panel B of Table 2.4. The change during World War I is quite close to the right end of the distribution (Panel B). Figure B23 presents the results of the randomization inference procedures we implement, which are described in greater detail in the appendix.

2.5.2 Interpreting Magnitudes

The estimates in Tables 2.4 and B30 capture the short-run responsiveness of patent flows to large shocks to market size. The magnitudes of both the shock and industry response are more readily translated into elasticities in the context of the Civil War than in the context of World War I.[19] Between data from Barnes and Stanton (1866), Hasegawa (2012), and the 1860 Census of Manufacturers, we can infer that the Civil War elevated annual revenues across the artificial limb industry by an average of roughly 100 log points over four years.[20] The estimates in Panel B of Table 2.4 thus suggest that, during the Civil War, the elasticity of short-to-medium run patenting with respect to the short-to-medium run shock to potential revenues was slightly greater than 1. We can similarly infer an elasticity of firm entry with respect to the Civil War era demand shock. As reported in Table 2.2, there were five artificial limb manufacturers in the 1860 Census of Manufacturing, and at least 17 manufacturers in 1865, implying an increase of at least 120 log points. This implies an elasticity of firm entry of greater than 1. These elasticity estimates

[19] It is less feasible to infer elasticities for the World War I period due to a combination of conceptual hurdles and data limitations. The key conceptual hurdle is that the conflict's global nature makes it difficult to infer the precise markets to which the firms who were patenting with the USPTO were responding. The key data limitation is that we lack sources on the number of manufacturers either during or preceding the war. In the 1910 Census of Manufacturing, for example, artificial limb manufacturers have been merged with a broader category including surgical appliances.

[20] From the 1860 census of manufacturers, we know that the value of the industry's output was roughly $53,000 in 1859. From Barnes and Stanton (1866), we know that over the first four years of the Union Army's artificial limb program, an average of roughly $91,000 in artificial limbs were procured. Viewing this as an increase over baseline demand from causes outside of the war, we estimate a 100 log point increase by comparing $ln(53,000)$ to $ln(53,000+91,000)$. The increase in units sold exceeded the increase in revenues because the Civil War limb allowances were substantially lower than pre-war prices.

are larger than typical estimates of the long-run effects of potential market size on innovation, as discussed by Dubois et al. (2015). Consistent with recent findings from Agarwal and Gaule (2021), who analyze the COVID-19 context, we find relatively sharp short-run responses of innovation to crisis-driven demand shocks.

Interestingly, wartime booms in prosthetic device patenting were not sustained over the long run. This might initially seem puzzling, given that the government's commitment to providing limbs was ongoing. Historical context provides evidence, however, that sustained demand for U.S.-manufactured prosthetic limbs was short-lived during both episodes. Following World War I, demand for U.S.-manufactured devices was short-lived because the European powers made conscious efforts to develop their own prosthetic device industries. By 1920, moreover, veterans with amputated limbs in Germany, Canada, and the United States were documented to prefer adapting to life without a prosthetic (Linker 2011; p. 114,118). The same was true following the Civil War; an overwhelming majority of Union veterans chose cash over replacement artificial limbs when they were given that choice during the post-war years.[21] Substantial demand for replacement limbs thus may not have materialized. In both settings, the preference for cash over replacement limbs is suggestive that, contemporaneous innovation notwithstanding, quality remained low in an absolute sense.

2.5.3 Traits of Wartime Prosthetic Device Patents

We now turn to estimating the effects of wartime procurement on the economic characteristics of prosthetic device patents. Our estimates of equations (2.4.3) and (2.4.4) are presented in Table 2.5, while the underlying time series are presented in Figures 2.2 and 2.3, with additional

[21] Over the decades immediately following the Civil War, the U.S. government provided allowances for the regular replacement of artificial limbs. Notably, veterans were allowed to choose between a replacement limb and cash, which was referred to as a commutation payment (Hasegawa 2012; p. 76). Statistics from annual reports of the army's Surgeon General reveal that veterans overwhelmingly preferred cash; from 1870 to 1891, "arm amputees chose a new device over commutation only 1.4 percent of the time, and leg amputees selected a new leg 21.9 percent of the time" (Hasegawa 2012; p. 76). This suggests, perhaps unsurprisingly, that quality was low in an absolute sense. The shock to artificial limb purchases was thus a pronounced shock spanning a period of four to five years. Our estimates will thus tend to capture the short-to-medium response of industry to a large but temporary shock to demand.

detail in Appendix Figures B27, B28, B29, B30, and B31. Several facts of interest emerge from this analysis.

We find that the Civil War was associated with across-the-board increases in emphasis on our cost-oriented production process traits. The average across these traits (namely "cost," "simplicity," and "adjustability") more than doubled from a base of 0.16, as shown in Figure 2.2. This estimate is statistically distinguishable from zero at the 0.01 level using either the simple time series or synthetic control estimator, as it is a true outlier relative to the distribution of randomization test outcomes. In contrast, the average across cost-oriented production process traits moved quite modestly during World War I. While both periods ushered in substantial increases in emphasis on adjustability, Civil War-era prosthetic device patents also exhibit economically substantial shifts towards emphases on "cost," and "simplicity" as shown in Figure 2.3. Changes in the latter two traits were relatively modest during the World War I episode, as can be seen in Appendix Figure B28. This contrast is plausibly linked to procurement incentives, as the low, fixed-price reimbursements of the Civil War period created strong incentives for innovation to reduce costs. While we do not know the precise details of World War I procurement arrangements for artificial limbs, cost-plus contracts, which blunt incentives for innovation to reduce costs, were "the most common type of contract" during that period (Graske 1941; p. 17).[22]

A comparison between patents for artificial arms and legs provides an additional, suggestive piece of evidence that the emphasis of Civil War era prosthetic device patents on production processes can be linked to the Union's procurement policy. The government's procurement arrangement, namely fixed-price reimbursement of $50 per arm and $75 per leg (roughly $1,000 and $1,500 in 2018 dollars), created a strong incentive for cost-oriented production process innovation because these payments were modest relative to manufacturers' costs. Cost data from the 1860 manufacturing census indicates that payments for artificial arms implied a lower charge-to-cost ratio than for artificial legs (roughly 2/3 vs. 3/4), creating an even greater incentive

[22] Withrow Jr (1942) links the predominance of cost-plus contracts during the World War I-era to the reluctance of firms to submit bids on a fixed-price basis given the risks associated with rapidly rising prices for raw materials.

for cost-reducing innovation. As shown in Figure B25, patents for artificial arms did indeed exhibit a more dramatic increase in their emphasis on production process improvements, and in particular on cost reduction, in comparison with patents for artificial legs.

An alternative possibility is that the emphasis of Civil War era artificial limb patents on the production process might simply have reflected the industry's natural trajectory. That is, if artificial limbs were a "new" technology during the pre-war period, a surge in production-process innovation might naturally be expected. This is not plausible, however, as the pre-war state-of-the-art technology had existed for quite some time. Patents held by Benjamin Franklin Palmer, the pre-war artificial limb industry's leading manufacturer, extended back to 1846. Throughout the 1850s, the rate of production process innovation evolved quite smoothly for artificial limb patents as well as for patents in our control groups. The early-1860s spike in production process innovation for artificial limbs is a distinctive break from this pattern.

We next consider dimensions of quality, for which two findings are both empirically robust and connect directly to historical narratives. First, both our simple time series and synthetic control estimators provide evidence that World War I-era patents exhibit an increase in emphasis on occupation-oriented appliances (see Table 2.5 and Figure 2.3). This finding has a strong connection to the historical records regarding both the intentions of World War I-era artificial limb procurement and the specific technologies to which this period's patents gave rise. Regarding the specific technologies, these "appliances" involved interchangeable, occupation-oriented attachments like the hammer, welding, and woodwork oriented attachments shown in Figures B4, B5, and B6 in appendix B1. Notably, as shown in column 5 of Table 2.5, British World War I-era patents offer a strong piece of supplemental evidence that the demand associated with employment-oriented rehabilitation programs generated increases in emphasis on occupation-oriented appliances. This is relevant in part because the shift towards occupation-oriented appliances in the U.S. patents is, despite representing a substantial increase in percent terms, not an outlier within the relevant placebo distribution and is thus on the margins of statistical significance.

Second, both our simple time series and synthetic control estimators yield strong evidence that Civil War-era prosthetic device patents exhibit a substantial increase in emphasis on comfort (see Table 2.5 and Figure 2.3). By contrast, World War I-era prosthetic device patents de-emphasized comfort (see Table 2.5 and Figure 2.3). These findings are plausibly linked to shifts in demand, which came directly from veterans during the Civil War and from the veterans' medical bureaucracy during World War I. Of course, such a difference in innovation across wars may reflect a variety of factors aside from those that we identify. The historical record, however, as discussed in section 2.1, suggests that the World War I-era medical bureaucracy played a heavy hand. Our findings for this period are very much in line with the bureaucracy's de-emphasis on the veteran's comfort and emphasis on social and labor market reintegration. As with our evidence on occupation-oriented appliances, British patents offer supplemental evidence on the decrease in emphasis on comfort during the World War I period.

2.5.4 Robustness of Analysis of Patent Traits

In section 2.4, we discussed the challenges underlying the construction of control groups in our analysis of the product and production process traits emphasized in patent documents. These challenges motivated our presentation of both a simple time series estimator and a synthetic control estimator in Table 2.5. In this section, we present an additional robustness analysis in which we deploy a range of alternative procedures for constructing control groups. Tables B34, B35, B36, and B37 present difference-in-differences estimates using the following approaches: Table B34 relies exclusively on our full sample of 1,200 manually coded patents; Table B35 uses the full sample of patents as coded using our machine learning model; Table B36 restricts the control group to medical patent classes; finally, Table B37 selects control groups using a simple "caliper" matching procedure.[23]

[23] In yet another robustness check, we have constructed synthetic controls from a sample of medical and mechanical technology classes that excludes all classes that might be directly affected by wars. In addition to classes involving firearms and ammunition, we exclude surgery, classes with plausible linkages to military uniforms (e.g., boot and shoe making, buckles, etc.) camp equipment (e.g., tents), and several others. Excluding these technology classes from the set of potential "donors" to our synthetic control groups has very little effect on our estimates.

The results we have emphasized throughout are findings that are robust to deploying this full set of strategies for constructing control groups, as well as to relying exclusively on the time series change in the emphases of prosthetic device patents as in equation (2.4.3). These include our findings on the Civil War-era increase in emphasis on production process innovation, the Civil War-era increase in emphasis on comfort, the World War I-era decrease in emphasis on comfort, and the World War I-era increase in emphasis on occupation-oriented appliances. In each of these cases, our estimates are robust across the full range of strategies for constructing control groups and imply large percent changes in emphasis on the trait in percent terms.

In contrast with the robust evidence on the findings discussed above, our evidence on appearance and durability illustrate methodological challenges in the analysis of patent texts. The estimates in Tables 2.5, B34, B35, B36, and B37 reveal that our estimates for appearance and durability, and to a lesser extent materials, are sensitive to whether we look to the simple time series change, use the full set of candidate controls, or use a data-driven control group. As we discuss in greater detail in appendices B1 and B2, these traits pose challenges with respect to both the construction of control groups and the implementation of text analysis methods. Consequently, we interpret our evidence on appearance, durability, and materials as weak. Our conclusions thus emphasize the traits for which our evidence is robust and for which we have greatest confidence in the output from our text analysis methods.

2.6 Discussion and Conclusion

Our analysis of Civil War and World War I-era prosthetic device patenting yields several findings of potential interest. First, we find that wartime procurement programs were associated with large increases in the volume of prosthetic device patents. We thus add to an existing body of evidence that finds that innovation can respond quite strongly to changes in demand.

Second, we find that cost-conscious production process innovation increased substantially during the Civil War. This highlights the potential relevance of the Civil War period's procurement

model, which involved fixed-price reimbursement at modest rates. Experts observe that modern medical innovations have tended to bring costly enhancements to quality rather than cost-conscious improvements in productivity (Chandra and Skinner 2012, Skinner 2013). Our findings provide a useful counter-example to this tendency. Demand shocks coupled with cost-conscious payment models can steer innovation in a cost-conscious direction.

Third, we find that the prosthetic device patents of the Civil War and World War I episodes diverged with respect to dimensions of quality. Civil War-era prosthetic device patents exhibited an increase in emphasis on comfort. By contrast, World War I-era prosthetic device patents de-emphasized comfort and emphasized occupation-oriented "appliances." These differences are plausibly linked to a World War I-era shift in choice away from veterans and towards medical professionals. This shift was associated, in turn, with a heightened emphasis on veteran rehabilitation and re-employment. As a caveat, we note these differences between Civil War and World War I-era prosthetic device innovations may stem from several factors that would be difficult to empirically disentangle.

A caveat accompanying our analysis relates to the limitations of text analysis. As discussed in appendix B2, seemingly modest reductions in the accuracy of our text analysis models can substantially attenuate our estimates of the effects of wartime procurement on the direction of prosthetic device innovation. While the accuracy of our models is generally quite high, it varies across the variables we construct. Moderately lower accuracy warrants caution, for example, in interpreting our analysis of the traits we term "materials" and "durability." Further, we highlight a key difference between dimensions of product quality and aspects of the production process. Dimensions of product quality can be highly context-specific, which makes it difficult to select control groups. Consequently, we have more confidence in our analyses of attributes that relate to the production process than in our analyses of attributes that capture dimensions of quality. For researchers who desire to apply similar text analysis tools in other settings, we provide a set of best-practice insights to help guide the development and evaluation of text analysis models.

Readers may also wonder about the rapid pace with which both the patent counts and their emphases evolved during the historical episodes we analyze. An anecdote may help to confirm that the responses we track are real. James Hanger, a renowned prosthetic limb inventor, is documented to have invented and produced a prosthetic limb within six months of being injured during the Civil War's initial skirmishes. Hanger's invention entailed improvements to both function and comfort. Hanger, Inc., the company he subsequently founded, remains in operation today. Beyond this anecdote, the tendency for large shocks to generate rapid innovative responses has been observed elsewhere. Hanlon (2015) finds, for example, that the British textile industry responded quite rapidly to the Civil War's impact on its supply chains. More recently, Agarwal and Gaule (2021) find that the COVID-19 pandemic has had a much greater and more rapid impact on innovation than long-run elasticity estimates would lead one to predict.

We conclude by reflecting on the role of innovation in enabling individuals and societies to respond to large and negative health shocks. Both wars and pandemics can have dramatic effects on the need and demand for medical innovations. Our analysis adds to a body of research on how innovation responds to these societal needs. While the overall consequences of wars and pandemics are devastating, the evidence reveals how their adverse effects can be blunted by the ingenuity of inventors and entrepreneurs.

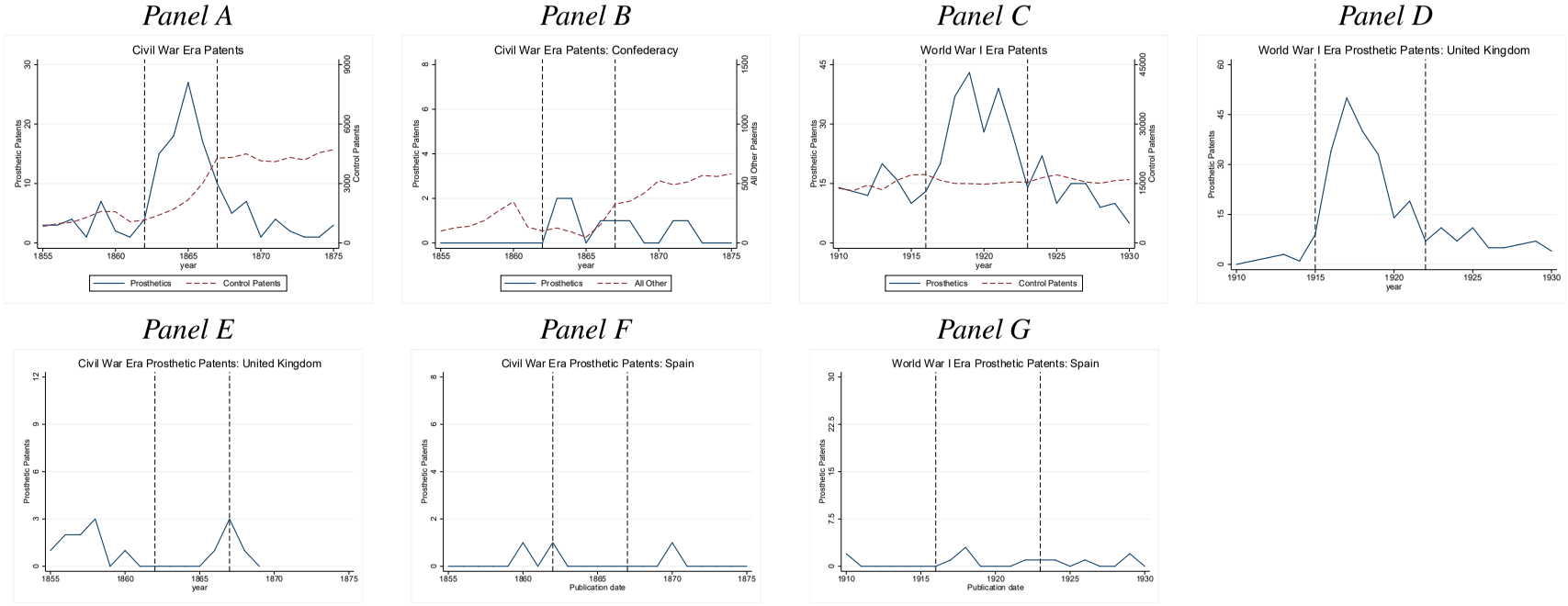


Figure 2.1. Patent Time Series Contrasting Regions Directly Impacted by the US Civil War and World War I with Regions That Were Not. Note: This figure presents annual time series on patents, using USPTO data from Berkes (2018), data from the Confederate patent office as documented by Dobyns (1994) and Knight (2011), as well as data on British and Spanish patents. Dashed vertical lines indicate the periods we associate with wartime prosthetic device patenting in the United States (1862 to 1866 during the Civil War and 1916 to 1922 during World War I) or in Britain (1915 to 1922 during World War I). In USPTO data, the solid blue line corresponds with patents from USPTO class 623 "Prosthesis." The four Confederate prosthetic device patents were identified by the authors based on patent titles. British and Spanish patents were categorized as prosthetic device patents using subject matter indices. In panels using USPTO data, red dashed lines correspond with all other medical and mechanical patent classes, defined using the hierarchical structure of technological categories in the NBER Patent Database (Hall et al. 2001).

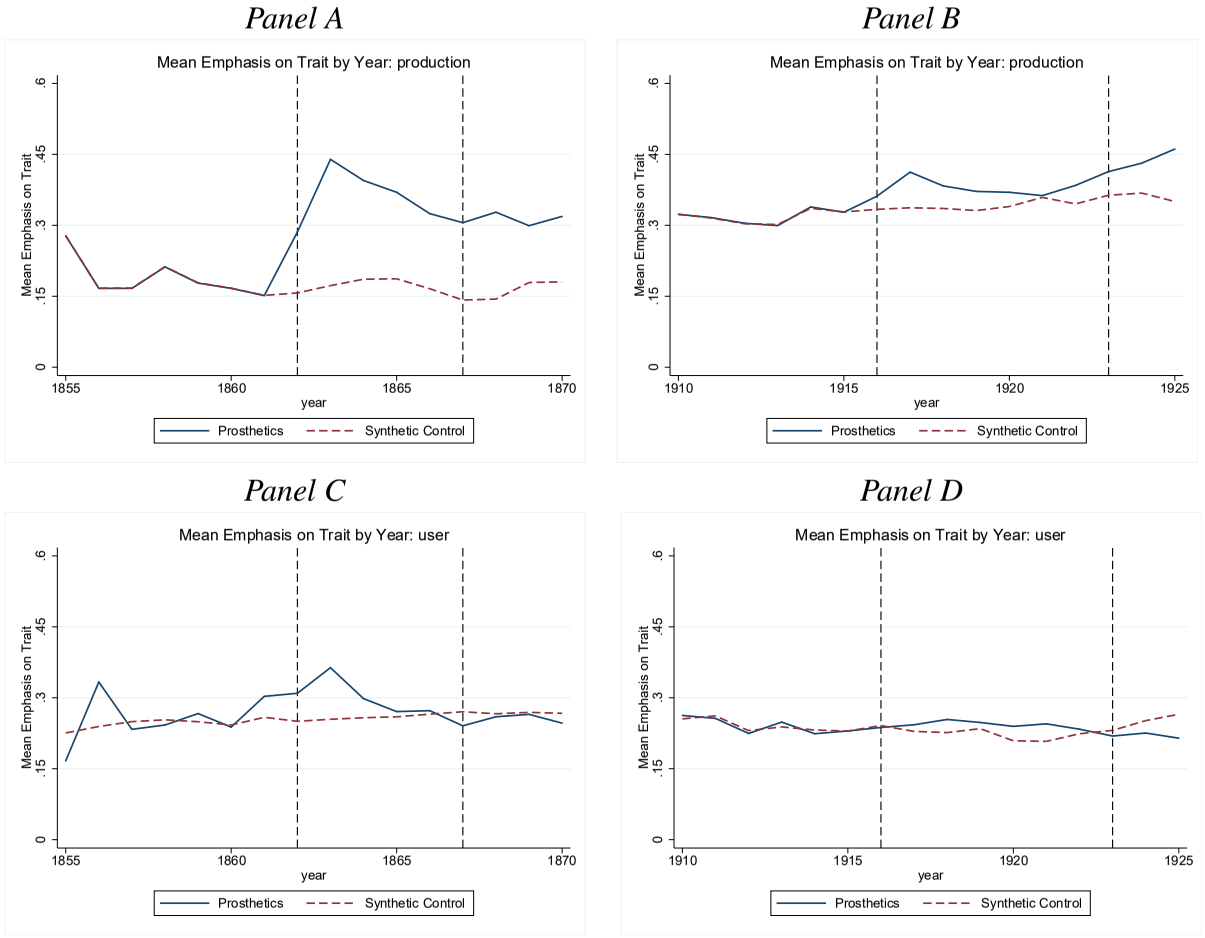


Figure 2.2. Changes in the Averages across Production and User-Oriented Traits. Note: The figure presents data on "treatment" and "synthetic control" series that describe the evolution of patents' emphases on averages across trait aggregates we term "production" (a simple average across "cost," "simplicity," and "adjustability") and "user" (a simple average across "appliances," "appearance," and "comfort") traits. The time series in Panels B and D are calculated as 4-year moving averages. The bar charts in Panels A and C present averages of the "Prosthesis" and "Synthetic Control" series. The series plot the share of patents in a given class ("Prosthesis" or the "Synthetic Control") that emphasize a given trait. In Panels A and C, the "Pre War" baseline extends from 1855 to 1861, and the "Wartime" period extends from 1862 to 1866. In Panels B and D, the "Pre War" baseline extends from 1910 to 1915, and the "Wartime" period extends from 1916 to 1922. We generate the synthetic control group using the "synth" package written by Abadie and Hainmueller (2010). "Donor weights" for panels A and C are chosen to match the treatment group on values extending from 1855 to 1861. "Donor weights" for panels B and D are chosen to match the treatment group on values extending from 1910 to 1915.

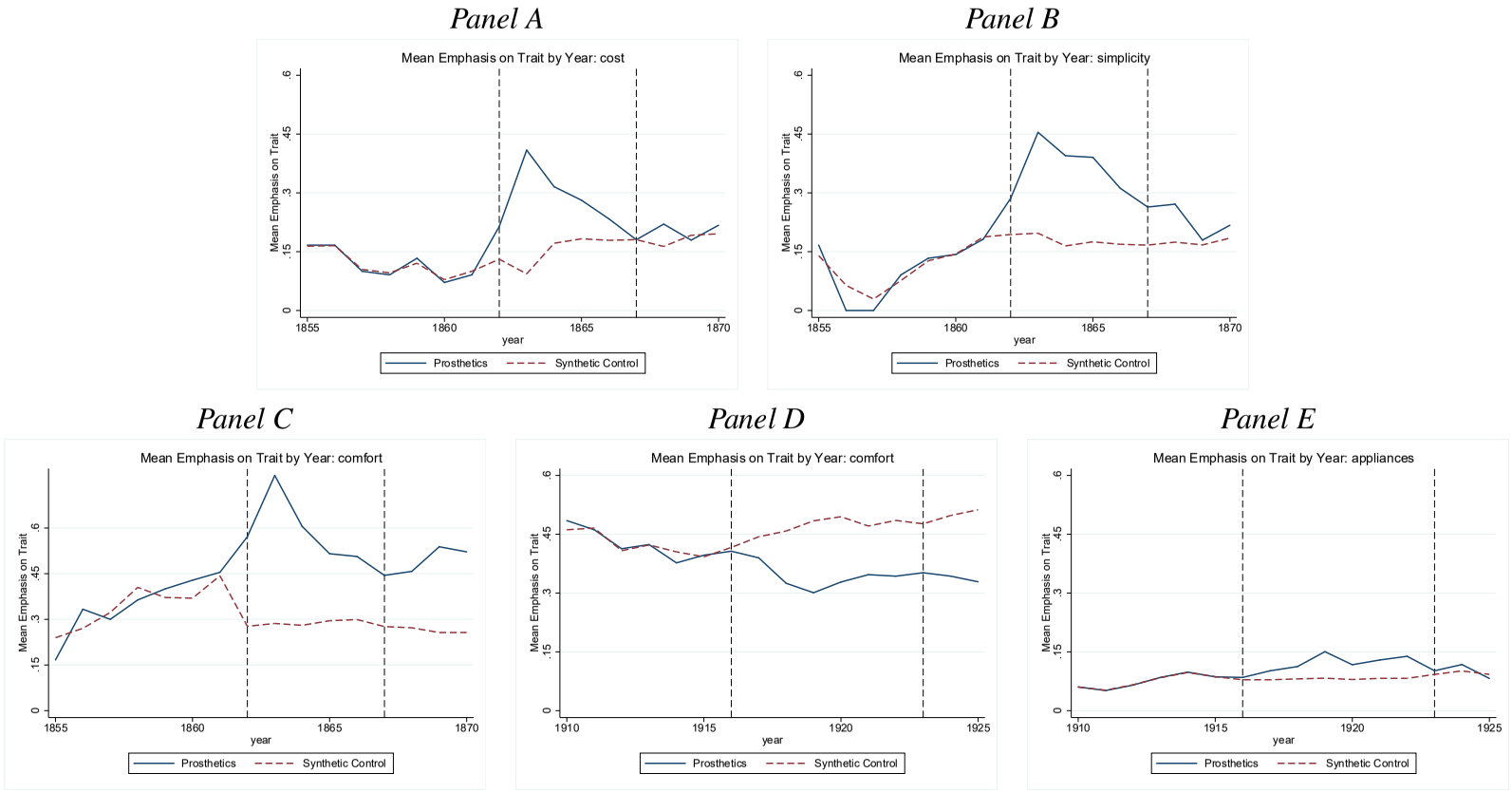


Figure 2.3. Changes in Traits with Strongest Connections to the Historical Record. Note: The figure presents data on "treatment" and "synthetic control" series that describe the evolution of patents' emphases on traits we term "cost," "simplicity," "comfort," and "appliances." The time series in Panels B and D are calculated as 4-year moving averages. The bar charts in Panels A and C present averages of the "Prosthesis" and "Synthetic Control" series. The series plot the share of patents in a given class ("Prosthesis" or the "Synthetic Control") that emphasize a given trait. In Panels A and B, the "Pre War" baseline extends from 1855 to 1861, and the "Wartime" period extends from 1862 to 1866. In Panels C and D, the "Pre War" baseline extends from 1910 to 1915, and the "Wartime" period extends from 1916 to 1922. We generate the synthetic control group using the "synth" package written by Abadie and Hainmueller (2010). "Donor weights" for panels A and B are chosen to match the treatment group on values extending from 1855 to 1861. "Donor weights" for panels C and D are chosen to match the treatment group on values extending from 1910 to 1915.

Table 2.1. Civil War Era Device Manufacturers, Patents, Early Market Shares, and Post-War Quality Rankings. Note: The information in the table comes from a variety of sources. The criteria for a manufacturer's inclusion in the table is that he either a) accounted for at least 0.25 percent of the limbs furnished through May 1866, as documented in Barnes and Stanton (1866), or b) was highly rated by either the Union or Richmond post-war ranking. The Richmond Ranking comes from Houston et al. (1866). The Union Ranking comes from Barnes (1865). An entry of nine-star indicates that a limb was considered and rated unfavorably or, in the case of the Union ranking, that it had been approved for reimbursement but was not included in the reported ranking. Both the Union and Richmond rankings of artificial arms had two distinct categories, resulting in multiple arms rated "1," "2," etc. An entry greater than 4 indicates that a limb was considered and rated favorably, but outside of the top 4. Linkages between manufacturers and patents were generated by the authors using the Google Patent Database and manufacturer names assembled from sources including Hasegawa (2012), Barnes (1865), Houston et al. (1866), Barnes and Stanton (1866). Patent dates come from Berkes (2018).

Manufacturer	Patents	First Patent	Market Share	Richmond Ranking	Union Ranking
Panel A: Artificial Legs					
B. F. Palmer	6122, 9200, 137711	1849	30.04	2	3
Douglas Bly	23656, 24002, 25238, 31438, 38549, 38550, 57666, 87624	1859	23.01	1	1
B. W. Jewett Patent Leg Company	16360, 29494	1857	19.27	9*	>4
E. D. Hudson	Copied Palmer's Design	na	10.92	>4	4
William Selpho/ Sepho and Sons	14836, 26378	1856	4.80	1	2
Salem Leg Company	35686, 35937, 44534,49528, 49529, 51593	1862	4.16	9*	>4
Charles Stafford	15831, 16420	1856	2.68	Not Considered	9*
Richard Clement	47281	1865	2.23	>4	>4
A. A. Marks	40763, 46687, 234596, 366494	1863	1.17	9*	>4
American Arm and Leg Company	40956	1863	0.72	>4	9*
National Arm and Leg Company	39599	1863	0.40	9*	9*
Marvin Lincoln	na	na	0.32	Not Considered	9*
James Hanger	Confederate Patents	1863	Large in South	9*	9*
Panel B: Artificial Arms					
Marvin Lincoln	39487	1863	45.51	2	2
Grenell & Co	44638	1864	13.02	1	4
H. A. Gildea	na	na	10.39	9*	4
D. W. Kolbe	45052, 255796	1864	8.58	9*	1
Selpho and Sons	18021	1857	8.53	9*	3
E. Spellerberg	42515, 51238	1864	6.49	9*	9*
National Arm and Leg Company	46158, 46159, 48002	1865	4.17	1	3
B. F. Palmer	22575, 22576	1859	2.45	9*	9*
John Condell	48659	1865	0.00	2	1

Table 2.2. Facts on Industry Response Surrounding the Civil War. Note: Data for 1865 come from Barnes and Stanton (1866) and Hasegawa (2012). Other years come from Census of Manufacturing tabulations. Patent dates come from Berkes (2018).

	(1859)	(1865)	(1869)
Manufacturing Establishments	5	$\geq$ 17	24
Artificial Limb Output	$\approx$ 350	$\geq$ 3,461	$\approx$ 1,000-2,000
Value of Output	\$53,000	$\geq$ \$223,550	\$160,416
Patents in Surrounding 5 Years	15	87	27

Table 2.3. Patent Attributes with Descriptions. Note: The table describes the definitions we apply in coding each of the economic attributes on which our analysis focuses. The attributes we term cost, simplicity, and adjustability are the attributes we interpret as involving the production process, while appearance, appliances, and comfort are our user-oriented attributes. Columns 3–5 offer our assessments of the relative strengths of the historical narratives, economic interpretations, and control groups for each trait, respectively. By "strong" historical narratives, we mean that there is ample historical evidence that contemporaneous economic factors drove an emphasis on the given trait during one or both wars. By "strong" economic interpretation, we mean that a trait can be cleanly linked to aspects of labor productivity, buyer desires, or mass production. By "strong" control groups, we assess that the keywords describing the given trait have similar meanings and rates of use in control classes as in the prosthetic limb class.

(1) Attribute	(2) Description	(3) Narrat.	(4) Interp.	(5) Controls
Individual Traits				
Cost	Construction is cheap, economical, and less labor intensive	Strong	Strong	Strong
Simplicity	Device construction is simple and less complex/difficult	Strong	Strong	Strong
Adjustability	Manufactured product adaptable to user specifications	Moderate	Weak	Strong
Materials	Made from new materials, substances, and compositions	Weak	Weak	Weak
Durability	Product is able to withstand wear and damage	Weak	Weak	Moderate
Appearance	Natural appearance, life-like, tasteful, and neat	Moderate	Strong	Weak
Comfort	Device noted as comfortable, noiseless, and promoting circulation	Strong	Moderate	Weak
Appliances	Attachable artificial limb components that aid in workplace tasks	Strong	Weak	Weak
Aggregate Traits				
Production	Combination of simplicity, cost, and adjustability traits			
User	Combination of comfort, appearance, and appliances traits			

Table 2.4. Relative Increases in Prosthetic Device Patenting During the Civil War and World War I. Note: The table presents estimates of equation (2.4.1). The control group used for each regression is described in the column heading. The sample for Panel A includes both the Civil War and World War I episodes, while the sample for Panel B consists solely of the Civil War episode and the sample for Panel C consists solely of the World War I episode. For observations associated with the Civil War, the pre-war period extends from 1855 to 1861, while the period over which the war influenced prosthetic device patenting is defined to extend from 1862 to 1866. For observations associated with World War I, the pre-war period extends from 1910 to 1915, while the period over which the war influenced prosthetic device patenting is defined to extend from 1916 to 1922. In Panel A, the standard errors reported in parentheses allow for clusters at the patent class-by-war episode level. In each panel, the p-values reported in rows labeled "Randomization Inf" are based on the position of the point estimate in the distribution of placebo point estimates that are constructed using a procedure along the lines recommended by Imbens and Rosenbaum (2005). Additional details are reported in the main text.

	(1) All Cntrls	(2) Matched	(3) Medical	(4) Misc. Mech.	(5) Metal	(6) Mater. Proc.	(7) Non War
Panel A: Full Sample							
Prosthetics x War	0.951	0.853	0.981	0.883	1.015	1.021	0.945
	(0.267)	(0.298)	(0.294)	(0.194)	(0.269)	(0.338)	(0.255)
N	432	88	34	128	56	92	362
Clusters	216	44	17	64	28	46	181
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Class-by-Episode FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Period Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SEs in Parentheses	Clustered	Clustered	Clustered	Clustered	Clustered	Clustered	Clustered
Randomization Inf.	$P < .01$	$P < .01$	$P < .01$	$P < .01$	$P < .01$	$P < .01$	$P < .01$
Panel B: Civil War							
Prosthetics x War	1.216	0.793	1.259	1.071	1.260	1.348	1.198
Randomization Inf.	$P < .01$	$P < .01$	$P < .01$	$P < .01$	$P < .01$	$P < .01$	$P < .01$
N	188	88	14	56	24	42	156
Panel C: WWI							
Prosthetics x War	0.687	0.571	0.716	0.697	0.774	0.698	0.693
Randomization Inf.	$P < .05$	$P < .2$	$P < .01$	$P < .03$	$P < .01$	$P < .01$	$P < .04$
N	244	32	20	74	34	52	208

Table 2.5. Changes in the Nature of Prosthetic Device Patents. Note: The table presents estimates of the effect of wartime procurement arrangements on the fraction of prosthetic device patents that emphasize a given economic trait. Estimates in columns labeled "Simple Diffs" are of beta-TS from equation (3), while estimates in columns labeled "Synth Estimate" are estimates of beta-DD from equation (2.4.4), where the control group is constructed separately for each trait using the synthetic control procedure described in greater detail in the main text. One-sided p-values are presented in parentheses beneath each point estimate, and two-sided p-values are presented in brackets. In several instances (including Civil War era production process innovation, WWI era de-emphasis on comfort, and WWI era emphasis on employment-enhancing "appliances"), the historical narrative delivers strong one-sided predictions for the evolution of prosthetic device patents. All p-values are generated using randomization inference (Imbens and Rosenbaum 2005), which in this application involves straightforwardly ranking the point estimate for the prosthetic device technology class against the "placebo" point estimates associated with the other technology classes in our sample.

	(1)	(2)	(3)	(4)	(5)	(6)
	US Civil War		US WWI		GB WWI	Notes
	Simple Diffs	Synth Estimate	Simple Diffs	Synth Estimate	Simple Diffs	
Panel A: Aggregated Traits						
Production Average	0.187	0.190	0.074	0.038	0.124	Strong Civil War
	(0.000)	(0.000)	(0.008)	(0.049)		Narrative (+)
	[0.000]	[0.000]	[0.016]	[0.098]		
User Average	0.006	0.036	-0.007	0.019	-0.109	
	(0.330)	(0.054)	(0.139)	(0.115)		
	[0.660]	[0.108]	[0.279]	[0.230]		
Panel B: Individual Traits						
Cost	0.152	0.141	0.079	0.050	0.028	Strong Civil War
	(0.032)	(0.054)	(0.074)	(0.066)		Narrative (+)
	[0.064]	[0.109]	[0.148]	[0.131]		
Simplicity	0.238	0.195	0.043	-0.001	0.226	Strong Civil War
	(0.032)	(0.011)	(0.254)	(0.557)		Narrative (+)
	[0.064]	[0.022]	[0.508]	[0.990]		
Adjustability	0.171	0.076	0.099	0.116	0.118	
	(0.000)	(0.143)	(0.016)	(0.008)		
	[0.000]	[0.286]	[0.033]	[0.017]		
Appliances	0.049	NA	0.065	0.038	0.112	Strong WWI
	(0.106)		(0.049)	(0.066)		Narrative (+)
	[0.213]		[0.098]	[0.131]		
Comfort	0.150	0.303	-0.119	-0.116	-0.230	Strong WWI (-)
	(0.032)	(0.016)	(0.000)	(0.000)		and Civil War (+)
	[0.064]	[0.033]	[0.000]	[0.000]		Narratives
Appearance	-0.182	0.078	0.033	0.068	-0.209	
	(0.043)	(0.037)	(0.107)	(0.008)		
	[0.085]	[0.074]	[0.213]	[0.016]		
Durability	0.016	0.149	0.064	0.025	0.102	
	(0.372)	(0.083)	(0.041)	(0.172)		
	[0.745]	[0.167]	[0.082]	[0.344]		
Materials	0.026	0.035	0.008	-0.005	-0.050	
	(0.138)	(0.104)	(0.328)	(0.496)		
	[0.277]	[0.209]	[0.656]	[0.990]		

2.7 Chapter Acknowledgments

Chapter 2, in full, is currently being prepared for final submission to the *Review of Economics and Statistics*, at which it is conditionally accepted, for publication of the material. This project was co-authored with Jeffrey Clemens. The dissertation author was a primary investigator and author of this paper.

Chapter 3

The Dynamics of Health Care Price Reform

Governments play a significant role in health care markets. The US government's Medicare health insurance program is a key example, accounting for 21% of national health care spending (Cubanski and Neuman 2023). However, the program's high and rising expenses have prompted policymakers to focus on identifying measures to control costs (Arad and McClellan 2022, Navathe et al. 2020). Among the prominent solutions is price reform, which aims to lower health care prices through price ceilings, negotiations, or competitive bidding (Frank and Nichols 2019, Ji 2023). While these reforms can reduce government expenditures in the short run (Ji 2023), there is limited evidence of their long-term consequences, which are important to understand given the government's significant role in these markets. For instance, price regulation could change the profitability of innovating in affected markets, influence the direction of innovation, alter how firms structure supply chains, and affect the quality of products being produced. These potential consequences pose a critical challenge in designing health insurance programs and broader procurement policies: how can we strike a balance between cost savings for patients and taxpayers while mitigating potential adverse impacts on innovation, product quality, and market structure?

In this study, we explore the impact of price regulation on innovation, product quality, and market structure. We focus on the medical device sector, characterized by substantial

research and development (R&D) activity, a wide range of differentiated products, and intricate global supply chains, allowing us to study the effects of price reform on a diverse range of outcomes. Specifically, we focus on durable medical equipment (DME), which are medical devices prescribed for home use, such as insulin pumps, oxygen tanks, and wheelchairs. We investigate the impact of price regulation by leveraging a series of price reforms enacted by the Centers for Medicare and Medicaid Services (CMS). These reforms lowered the Medicare prices paid by 45% for certain DME categories in the largest metropolitan statistical areas (MSAs). By 2019, these reforms cut the total Medicare expenditures for these categories by two-thirds compared to unaffected categories.[1]

To identify the impact of price reform policies, we compared outcomes in DME categories affected by the reform to those that remained unaffected. We employed a stacked difference-in-differences strategy to account for the reforms affecting additional categories over time. As a potential source of endogeneity, CMS chose categories for reform based primarily on pre-reform Medicare expenditures. Despite expenditure differences between affected and unaffected categories, there were no divergent pre-existing trends in the outcomes studied among affected categories relative to unaffected ones. Further, we also leverage within-category variation in the extent to which firms were exposed to price reform by comparing firms with a larger share of their product portfolio within treated categories to those with smaller shares. The firm-level results were consistent with those at the category level, providing further support for our findings.

Our analysis makes use of multiple administrative data sets to capture the multifaceted effects of price reform on various outcomes. These data sources include global patent data, FDA device submissions, FDA adverse events reports, Medicare fee schedules, Medicare claims, and novel data on Medicare-contracted device suppliers and manufacturers. By combining these diverse data sets, we construct three sets of outcomes to examine the long-run market dynamics following the price reform: quantity and direction of innovation, market structure and supply-chain reactions, and product quality.

[1] Authors' analysis of the Medicare claims data.

Our first set of results shows a decrease in innovation and a shift in the direction of innovation following the DME price reform. We define two measures of innovation: the number of new device submissions to the FDA for approval and the number of (US and foreign) patents filed in DME categories. We estimate a statistically significant decline in FDA submissions by 22% in DME categories affected by the reform, with a similar, albeit not statistically significant, decrease of 29% in the number of patents filed. Focusing on firms whose existing product portfolios were affected by the reform, we find that those with above-median exposure to the price reform were 29% less likely to file for new patents in the affected categories, a statistically significant reduction. In contrast, patenting activity in unaffected categories remained unchanged.

The results also show that firms most affected by the price reform increased their emphasis on cost-cutting innovations following reform. Specifically, we define patents as "product" and "process" innovations following Bena and Simintzi (2022). Following the price reform, we find that firms shifted towards "process" innovations, which involve pioneering new methods to improve the production process and reduce production costs, as opposed to "product" innovations like introducing a new product feature. We find a statistically significant 32% increase in the share of patents focusing on process innovations and a statistically significant 15% decrease in the share of patents focusing on product innovation among these most-affected firms relative to those less affected.

Our second set of results reveals a reduced rate of entry by manufacturers and an increase in offshoring the production process following the price reform. Specifically, we find a statistically significant 25% reduction in the number of new entrants into the affected product categories, driven by a 47% reduction in entry by US manufacturers, and a smaller (albeit statistically insignificant) 8% increase in entry by foreign manufacturers. The diverging trends in entry between US and foreign manufacturers may reflect their differential comparative advantages in production, with the latter being more favored in an increasingly cost-conscious environment created by the price reform. Furthermore, manufacturers responded to the price reform by shifting their production overseas. Among manufacturers still operating in affected markets, the

number of firms outsourcing manufacturing to other companies increased by 50%, although the estimate is not statistically significant. However, there was a statistically significant increase of 65% in the number of firms outsourcing production to foreign manufacturers, suggesting that firms adapted to lower prices by leveraging foreign supply chains to cut production costs. These results suggest a shift towards global supply chains and a change in the composition of products sold in the US, with an increasing proportion manufactured by foreign firms.

Our third set of outcomes explores changes in product quality following price reform. We find that these cost-cutting strategies are associated with a decline in product quality, evidenced by increased device repairs and reported adverse events. Our analysis of Medicare claims data suggests a 100% increase in the repair rate for affected DME among Medicare beneficiaries, resulting in an estimated additional 700,000 repairs per year for Medicare. Furthermore, our analysis of FDA adverse event reports[2] suggests a 233% increase in adverse events reported for affected DME categories, despite decreasing utilization (Ji 2023). Notably, adverse event reports increased most significantly for products manufactured by foreign companies and contractors, suggesting that either the utilization of foreign-made DME increased to such an extent that it resulted in a substantial rise in adverse events or that foreign-made DME, especially DME outsourced to foreign manufacturers, was more prone to quality issues.

Our paper contributes to several literatures. First, we add to research on the relationship between market profitability and medical innovation, which has largely focused on the pharmaceutical sector. Although previous studies have shown that expansions in market size lead to large increases in R&D (Acemoglu and Linn 2004a, Blume-Kohout and Sood 2013b, Finkelstein 2004a), there is limited empirical evidence on the effects of price reform.[3] Our research aims to fill this gap, which has been made especially important by recent provisions in the Inflation

[2]CMS increased its surveillance of affected DME categories after price reform, which may have affected the reporting rate of adverse events. Thus, the adverse event outcomes are suggestive. However, results from our claims-based repair rates – which are not directly monitored by CMS – corroborate our adverse event results, strengthening our conclusions drawn from this data.

[3]Existing evidence is limited to theoretical (Filson 2012), simulation-based (Abbott and Vernon 2007, and correlational (Giaccotto et al. 2005, Civan and Maloney 2009) studies. See Philipson and Durie (2021) for a comprehensive review.

Reduction Act that allow CMS to set price ceilings on certain types of drugs by 2026. To our knowledge, our study is the first to measure the effects of health care price reform on innovation using quasi-exogenous variation in price reform policy and the first to examine its effects on product quality, the direction of innovation, and market structure.

Furthermore, our paper contributes to the literature on regulatory tools that affect innovation. Prior research has examined the impact of patent protection (Budish et al. 2015), entry regulation (Rogers 2023, Grennan and Town 2020b), and tort reforms (Galasso and Luo 2017) on innovation. We extend this work by examining the effects of price regulation on innovation and related outcomes.

Lastly, our paper also adds to the literature on procurement policy and innovation (Che et al. 2021, Slavtchev and Wiederhold 2016, Cozzi and Impullitti 2010). We show that procurement price can influence innovation and the flow of trade in a globalized economy, a theoretical insight first pioneered by McAfee and McMillan (1989). Our results indicate that low and uniform prices can cut expenditures but lead to offshoring and potentially lower-quality products. Our study provides unique insights into procurement policy within the large and growing medical device industry. Most closely related, Clemens and Rogers (2020) find that low, fixed-price payments for medical technologies lead to cost-cutting innovations. We find similar results by analyzing quasi-exogenous reforms to procurement policy, holding fixed time-varying factors rather than relying on comparisons across wartime eras. Our findings contribute to this research by suggesting that stringent procurement policies steer innovators toward cost-cutting process innovations and away from product innovation.

This paper is organized as follows: section 3.1 provides background, section 3.2 describes our data, section 3.3 detail our empirical strategy, section 3.4 presents our results, and section 3.5 concludes.

3.1 Setting

3.1.1 Medical Devices and Durable Medical Equipment

Medical devices are instruments or apparatuses intended for the diagnosis, treatment, or prevention of disease. Unlike pharmaceutical drugs, medical devices do not achieve their function through chemical action.[4] Medical devices cover a wide range of products. These include diagnostic devices, such as X-ray machines and electrocardiography (ECG) machines, therapeutic devices like infusion pumps, prosthetics such as prosthetic limbs and dentures, implants like pacemakers and stents, and assistive devices like mobility scooters and communication aids.

Medical devices are regulated by the US Food and Drug Administration (FDA) through both pre-market approval processes and post-market surveillance. In general, medical devices are classified into one of three categories based on their level of risk: Class I (low-risk), Class II (moderate-risk), and Class III (high-risk). Most Class III devices are subject to pre-market approval (PMA), which requires the manufacturer to provide data demonstrating the device's safety and effectiveness. Most Class II devices are subject to either PMA or pre-market notification, known as 510(k), which requires the manufacturer to demonstrate that a new device is substantially equivalent to a previously approved device. Class I devices are generally exempt from either PMA or 510(k), but must be registered with the FDA.[5]

Durable medical equipment (DME) is a category of medical devices designed for home use. These items aid in the recovery process after inpatient hospitalization or facilitate the management of ongoing illnesses. Examples of DME include wheelchairs, glucose monitors, oxygen concentrators, and nebulizers. DME can be Class I, II, or III.

[4] https://www.fda.gov/media/131268/download

[5] https://www.fda.gov/medical-devices/device-advice-comprehensive-regulatory-assistance/overview-device-regulation

3.1.2 Medicare Reform of Payments for Durable Medical Equipment

Medicare, the federal health insurance program for patients aged 65 and above, covers DME under Part B of the program. Historically, Medicare reimbursed for DME based on fee schedules set by CMS, which were largely tied to list prices from the late-1980s and inflation-adjusted over time. In 2006, due to concerns about high prices in the DME sector, Medicare announced a reform that would replace these fee schedules with prices determined through supplier auctions. The reform was initially implemented in nine metropolitan statistical areas (MSAs) in January 2011 and later expanded to an additional 91 MSAs in July 2013, covering approximately half of total DME spending in these areas.[6] Ji (2023) finds that the reform led to an average reduction of 45% in prices, with price reductions observed in all affected product categories. In 2016, the remaining MSAs also began lowering their prices based on prices generated by auctions in the first 100 MSAs. As of 2019, 13 product categories were covered by the price reform, while the remaining 43 product categories continued to follow the existing fee schedules. Figure 3.1 plots the total Medicare DME payments separately for product categories that were and were not subject to the price reform. Both sets of products had similar spending trends up to 2009, but spending in the affected categories declined sharply after that. By the end of 2019, annual Medicare spending in the affected categories decreased from approximately $4.5 billion to $2 billion, while spending in the unaffected categories continued to increase from $4.5 billion to $6 billion.

3.2 Data and Summary Statistics

We describe the data sets we use, the construction of our baseline sample, and report our baseline summary statistics.

[6]The original implementation date was January 2009, but CMS postponed the implementation by two years and instead imposed a one-time 9.5% price reduction for all treatment products in 2009.

3.2.1 Data

FDA Device Submissions (PMA and 510(k) Databases)

We use two FDA databases on device approvals: the pre-market approval (PMA) database and the 510(k) database. The PMA database contains information about medical devices that have undergone a rigorous review process, which typically involves clinical trials and other extensive testing, to demonstrate their safety and effectiveness. Most Class III devices must go through the PMA process before they can be sold in the US market. The 510(k) database contains information about devices that have been deemed substantially equivalent to devices already on the market, and therefore require a less lengthy review process. Most Class II devices are required to complete the 510(k) process. Together, the PMA and 510(k) databases capture the majority of late-stage innovative activity in these device categories. In both data sets, we observe the universe of FDA device submissions including the submitting company name, device brand name, product codes and descriptions, and submission and approval dates. For both databases, we include all submissions between 1996 and 2018.

Dimensions Patent Grants Extract

Dimensions is a comprehensive database that provides detailed information on patents issued by the United States Patent and Trademark Office (USPTO) and global patent offices across 100 countries. This database includes essential patent information such as the patent title, abstract, description, claims, filing date, and approval date, where applicable. To ensure that we observe close to the universe of patents in our sample years, we restrict our analysis to data between 1996 and 2016, as the patent submissions can take up to three years or more before they are publicly posted.

FDA Registration Database

The FDA requires all products sold in the US to be registered in this database. The variables include the name of the registering establishment, proprietary name of the product, product code, device classification (I, II or III), establishment type (e.g. manufacturer, contract

manufacturer, exporter) and the location of the establishment (US state or foreign country). In principle, these data cover the universe of medical devices (including DME) available for sale in the US. In practice, the database has two important limitations. First, the data are reported at the registration event level; for establishments that registered multiple brand names across different product codes, there was not straightforward way to establish one-to-one correspondence between the registered brand name and the associated product code. Second, the data do not include inactive registrations; for example, a firm that registered a given product in 2005 that stopped selling the product (thus stopped registering in subsequent years) would not be captured in later years' data. We address these data issues in two ways. First, we use yearly snapshots from WayBackMachine for 2009, 2010, 2011, 2013, and 2020 to enhance our sample by capturing currently inactive registrations. Second, we focus on firm-level events, which we can identify in the data, rather than device-related events, which we cannot. We record whether the contractors are US-based or foreign. Nonetheless, due to these data limitations, we restrict the use of the registration data only to our analysis of firm contracting behavior.

Medicare Data

We use the 100% Traditional Medicare enrollment and claims data from 2009 to 2019, which encompassed health care claims for all beneficiaries under Traditional Medicare. For each DME claim, we observe the date of the claim, the HCPCS code, the Medicare price, and the quantity purchased. We supplement these data with publicly available Medicare DME fee schedules. We obtained supplier-reported, quarterly data on the manufacturer, model, and make of DME products sold to Medicare beneficiaries between 2011 and 2019 through a Freedom of Information Act (FOIA) request.

FDA Adverse Event Reports (MAUDE).

The FDA's Manufacturer and User Facility Device Experience (MAUDE) database enables us to measure the safety of medical devices based on adverse event reports from 1992 to 2019. These reports include events such as deaths, hospitalizations, and life-threatening incidents,

as well as minor events like product breaks, across FDA device types. Following Ensign and Cohen (2017), we address data and coding issues in the MAUDE database.

3.2.2 Sample and Variable Definitions

DME Category-Level Baseline Sample

We define the set of treatment and control DME product categories based on whether a product category was ever subject to Medicare price reform during our sample period. We assigned all FDA device submissions and registrations to either the treatment or control group by matching the FDA product codes to Medicare DME categories (the level of the treatment). The match was completed by comparing HCPCS code descriptions within each DME category, DME category descriptions, and FDA product code descriptions. We construct a data set of unique device brand names within a device type using text analysis and aggregate them at the DME category-year level. Using the same method, we also construct a data set of unique manufacturer names listed in the FDA device submission database within device types and aggregate them at the DME category-year level to measure firm entry. Our baseline sample includes device submissions across 18 control DME categories and 8 treated DME categories, with over 3,738 unique device submissions spanning 1996 to 2016. Note that the sample does not include all DME categories as not all DME requires FDA approval (e.g. Class I devices generally only require registration), therefore, this measure captures changes in innovation among medium and high-risk devices.

To complement our FDA submissions measure, we also use patents as an indicator of DME innovation. Unlike with pharmaceutical drugs, there is no official database that links devices with their patents, so we create our own using a three-step procedure. First, we compile a list of keywords and patent classification codes (CPC) that correspond to each DME category description. Second, we use Digital Science's Dimensions platform (Hook et al. 2018) to collect all global patents granted that match those keywords and CPC codes in their text. Third, we count the annual number of patents filed within each DME category based on the the filing date.

The result is a panel of yearly patent counts for 51 DME categories from 1996 to 2016, using a collection of 236,656 global patents related to DME categories, of which over 100,000 are granted in the US. Patents are a useful complement to FDA device data for several reasons. First, they capture innovation for Class I devices, which do not need FDA approval and are missing from our FDA submission data. Second, they enable a straightforward analysis of the quality of new innovation, using rich textual and citation information. Lastly, an analysis of two different measures of innovation provides corroborative evidence.

We supplement the patent data with measures of process and product claims made within a patent document created by Bena and Simintzi (2022) (B&S). To identify process-oriented innovation, B&S exploit a US patent-specific policy that requires inventors to indicate process-related claims by beginning the claim with the words "A method for" or "A process for." B&S then tag the claims that begin with these words as "process" claims and those that do not as "non-process." We consider a patent as primarily process-oriented if it has an above-median share of process claims.

We measure changes in contractor relationships by identifying the number contractors registered within a product code in a given year. We then use the crosswalk described above to aggregate the data at the DME level. Our panel measures the establishment of over 1,500 US-based and 4,100 foreign contractor relationships across 36 control and 13 treated DME categories.

We measure changes in product quality using Medicare claims and FDA adverse event data. Using the 100% Medicare claims data for DME, we define a "repair" event as unique claim lines with repair modifiers ("RT", "LT", or "RB"). We define repair rates in each year at the product category level by dividing the number of repair events with the number of claims for each product category. Using the FDA MAUDE data, we also count the number of adverse event reports submitted to the FDA each year within each device type. We then use the crosswalk above to aggregate adverse event report counts at the DME level.

Manufacturer-Level Baseline Sample

We focus on firms that manufacture equipment in DME categories affected by price reform, defined as manufacturers that have either been listed by a Medicare supplier as selling in one of the affected categories or have registered themselves in an FDA product code corresponding to a DME category before price reform. All firms also must have filed a patent in one of the affected categories prior to the price reform. We match their names to global patent assignees using Dimensions. We keep only patents with high similarity scores to firm names, resulting in 486 firms that have patented at least once. We then calculate the share of patents for each firm in a treated DME category, based on our DME-level patent database constructed above. Figure C1 shows the distribution of the share of patents subject to price reform across firms. We define a firm as treated if its share is above the median, and assign the year of treatment as the first year a DME category with the firm's patent was subject to price reform. We count the number of patents filed by each firm from 1996 to 2016 to form our firm-level innovation measure.

3.2.3 Summary Statistics

Tables 3.1 and 3.2 provide an overview of the product categories and DME manufacturers in our study. Table 3.1 presents summary statistics of product innovation and utilization for all DME categories, categories impacted by the price reform, and categories not impacted by the reform. We report summary statistics across groups in 2005, a year before the first announcement of the price reform, and across all sample years from 1996 to 2016. Notably, we observe that the number of affected categories is roughly one-third of the number of unaffected categories. However, despite this, Medicare expenditures, users, and FDA device submissions were similar or higher in affected categories than those of unaffected ones, consistent with Medicare's intent to choose the largest markets for price reform.

Table 3.2 presents summary statistics for patent portfolios of DME manufacturers, both for the year 2005 and the entire sample period. The table shows that the average portfolio exposure to price reform was 24% for firms filing at least one patent in the affected DME

categories. Additionally, the distribution of manufacturer patent filings is highly skewed to the left, with the top quartile of firms filing only one patent per year. As a result of this skewness, the analysis for our firm-level study mainly focuses on the extensive margin of patenting outcomes, namely, whether a firm innovates at all in a given year.

3.3 Empirical Strategy

3.3.1 DME Category Analysis

We estimate the effect of the price reform by comparing outcomes for product categories that were subject to the reform with those that were not subject to the reform during our study period.

Figure C2 plots the raw trends for our two measures of innovation. Figure C2 (a) shows the number of PMA and 510(k) over time separately for product categories subject to the price reform and other product categories. While the former has higher levels of PMA and 510(k) filing prior to 2010, the gap largely shrinks thereafter. Figure C2 (b) shows analogous trends for the number of patents filed. Annual patent counts steadily increased for the two groups at comparable rates up to the early 2010s, when the rate of patent filing plateaus for product categories affected by the reform.

To empirically quantify the impact of the price reform on our measures of innovation, we use an event study specification with a stacked regression design. This approach assembles event-specific panel data for each of the DME categories subject to the reform and all control DME categories (i.e. categories not subject to the reform). All event-specific panels are then stacked while allowing unique time and product category fixed effects for each panel. We estimate the following event study specification:

$$Y_{i,t,k} = \gamma_{i,k} + \gamma_{t,k} + \sum_{r(i,t)\neq -1} \beta_r 1\{\text{Reform}\}_{i,k} \times I_{r(i,t)} + \varepsilon_{i,t,k}. \tag{3.3.1}$$

where i denotes DME categories, t denotes calendar years, and k denotes price reform

events. $\gamma_{i,k}$ and $\gamma_{t,k}$ denote event-by-DME category fixed effects and event-by-calendar year fixed effects, respectively. $1\{\text{Reform}\}_i$ is an indicator for whether DME category i in event panel k is subject to the price reform. $I_{r(i,t)}$ are indicators for years relative to the announcement of the reform, which are normalized to zero for DME categories not subject to the reform. We define $r(i,t) = 0$ as the year Medicare price reforms were announced, since investors and manufacturers can already respond to the change in expected revenue following the announcement, before the formal implementation of the reform. The coefficients of interest, β_r's, quantify the impact of the price reform on the outcome of interest $Y_{i,t,k}$. Since we have a small number of treated DME categories ($N^1 = 13$), to achieve reliable inference, we follow Conley and Taber (2011) and use control group residuals to compute standard errors. The relatively larger size of our control groups ($N^0 = 38$) allows us to reliably estimate standard errors in the presence of relatively few treated groups.

To summarize the impact over the post-period, we also estimate a pre-post version of the same specification where we replace the relative year indicators with an indicator for the period after the price reform has taken place: $1\{\text{Post}\}_t$. Also, since we are estimating the effect over the entire post-period, we only estimate one reform coefficient β_1. The estimating equation is given by

$$Y_{i,t,k} = \gamma_{i,k} + \gamma_{t,k} + \beta_1 1\{\text{Reform}\}_{i,k} \times 1\{\text{Post}\}_{t,k} + \varepsilon_{i,t,k}. \tag{3.3.2}$$

3.3.2 Manufacturer Portfolio Analysis

To explore heterogeneity in impact across manufacturers, we estimate the specifications analogous to equations (3.3.1) and (3.3.2) at the firm level. That is, we assemble event-specific panel data for each DME manufacturer subject to the reform and all admissible controls. Specifically, we estimate the following specification

$$Y_{j,t,k} = \gamma_{j,k} + \gamma_{t,k} + \sum_{r(j,t)\neq -1} \beta_r 1\{\text{Above Median Exposure}\}_{j,k} \times I_{r(j,t)} + \varepsilon_{j,t,k}. \quad (3.3.3)$$

where j denotes manufacturers, t denotes calendar years, and k denotes price reform events. $1\{\text{Above Median Exposure}\}_{j,k}$ is an indicator that the manufacturer's exposure to the price reform is above median among all device manufacturers. We define exposure as the share of a firm's patent portfolio during the pre-period that is affected by the reform. Other variables are defined analogously. We define $r(i,t) = 0$ as the first year Medicare price reforms were announced for a DME category that falls into the firm's portfolio. Our coefficient of interest, β_r's, estimate the differential change in the outcome between firms with above- and below- median exposure.

We also report a pre-post version of the same specification, as shown in equation 3.3.4:

$$Y_{j,t,k} = \gamma_{j,k} + \gamma_{t,k} + \beta_1 1\{\text{Above Median Exposure}\}_{j,k} \times 1\{\text{Post}\}_{t,k} + \varepsilon_{j,t,k}. \quad (3.3.4)$$

As with every non-experimental research design, selection into treatment is a primary concern. Medicare selects DME categories for price reform based on baseline yearly pre-reform expenditures, which may result in differences between treated and untreated categories. However, we do not find significant divergent pre-existing trends in the outcomes of interest. Additionally, we find consistent reductions in innovation following price reform, which, to the extent that higher expenditure categories have a more rapid innovation trajectory, may lead to upward bias and more conservative estimates of the treatment effect. Moreover, our firm portfolio analysis provides additional evidence for the impacts of price reform on innovation, market dynamics, and product safety using variation at the firm level.

3.4 Results

In this section, we present our results of estimating equations (3.3.1), (3.3.2), (3.3.3), and (3.3.4), which capture the static and dynamic effects of price reform at the DME and firm level. Subsection 3.4.1 details the effects of price reform on innovation, subsection 3.4.2 presents the effects on supply chain structure and the direction of innovation, and subsection 3.4.3 provides effects on product safety.

3.4.1 Changes in Innovation

Figure 3.2 displays our event-study estimates of changes in FDA submissions and patent filings. Panel (a) exhibits a sharp and immediate decline in FDA submissions one year after the price reform announcement. These effects persist over time and can amount to a statistically significant 47% decrease in FDA submissions in some years, relative to pre-reform means. Panel (b) illustrates a slower and steadier decrease in global patenting rates in affected DME categories relative to those in unaffected categories. As the time progresses, the effects of price reform become more significant, with estimates five years after the reform announcement growing in magnitude and significance. The long-run estimates suggest that price reforms lead to 100 fewer patents per year, representing a 50% reduction relative to the pre-reform mean, although the point estimates are not statistically significant. We find no significant pre-existing trends in treated groups relative to control groups.

The differential short-run impact of the price reform on patenting rates and FDA submissions can be explained by the cost structure of developing and commercializing early versus late-stage technologies. Patents primarily reflect progress in earlier-stage R&D, where firms incur marginal costs incrementally developing their products. In contrast, FDA submissions capture products that have been developed but must undergo a costly approval process before commercialization, with an average cost of $24–$75 million, depending on the approval method (Makower et al. 2010). As a result, firms may continue to develop existing projects in their

pipeline and patent them due to the relatively low marginal cost and the potential for profits outside the US, while finding it unprofitable to clear products with the FDA in the US. Therefore, panel (b) shows firms continuing to develop existing projects in the short run, which may not have been profitable to start after the price reform, before these types of projects are eventually exhausted. Panel (a), on the other hand, reflects firms abandoning existing and potential products within the US due to the high fixed costs of the approval process, which may outweigh their potential profits.

Figure 3.3 replicates Figure 3.2 (b) separately for US and foreign firms. Panel (a) illustrates that the long-run estimates for foreign patent filings are only marginally significant at times, representing a 42% decrease in patenting. In contrast, Panel (b) shows that the long-run estimates for patents filed in the US are all significant or marginally significant, indicating a decline of 50 filings a year, or a 66% decrease relative to the pre-reform mean. This disparity is consistent with two observations. First, US health policy has substantial implications for global R&D activity. Second, while both foreign and domestic markets are affected, domestic R&D is more significantly impacted by domestic policy.

Table 3.3 displays our static difference-in-differences estimates. In Panel (a), we observe a statistically significant 22% drop in FDA submissions across all post-reform periods and firms, while patent filings decreased by 29%, though not statistically significant. It is worth noting that these estimates do not account for the likely lag in response time between the implementation of the policy and its effect on patenting rates due to the lengthy nature of the R&D process, as described above. Therefore, it is important to examine the impact on patent filing over a longer horizon.[7]

Although patenting and FDA submissions are both decreasing in aggregate, we find a divergence between the number of FDA submissions from US firms and foreign firms. Table 3.3 Panel (a) suggests that the number of FDA submissions from US firms alone decreases by 49%,

[7]To address this limitation, we plan to estimate short and long-run effects separately in future iterations of this paper.

while the number of FDA submissions from foreign firms increases by 54%, both changes are statistically significant. We also find suggestive evidence that patent filing decreased more for US firms than for foreign firms, although the estimates are not statistically significant (see table C38).

It is natural to ask why we observe a decrease in patenting rates among foreign and domestic firms despite the increase in foreign firm FDA submissions. One possible explanation is the distinction in what patents and FDA submissions represent. Patents capture changes in the development of new products or processes, while FDA submissions measure changes in access to either new or existing products. Foreign firms may not introduce new technologies; rather, they may submit existing products for FDA approval to exploit their comparative cost advantages and sell them to US suppliers looking to cut costs. Such incentives would be consistent with an increase in FDA submissions from foreign firms but a decrease in foreign patenting rates.

Panel (b) of Table 3.3 reports the results of our firm-level analysis. We find that firms most affected by the price reform experience a significant 10 percentage point drop in the likelihood of filling a patent in an affected DME category, representing a 29% reduction relative to the less affected firms. In contrast, these firms do not significantly change their patenting behavior in unaffected DME categories, and may even slightly increase patenting in these areas (see Figure C3 for related event studies). We highlight that our firm-level analysis yields similar results to our DME-level analysis, both pointing to a 29% reduction in our measures of innovation.

In addition to an overall reduction in innovation, the price reform also altered the direction of innovation. As shown in Table 3.4, companies that were most exposed to price reform significantly increased their emphasis on process innovations by 32%, measured by a rise in the share of patents with an above-median number of process claims. Such innovations focus on new production methods known to reduce production costs (Bena and Simintzi 2022). In contrast, these firms reduced the share of patents filed focused on product innovations. Our findings suggest that price reform prompted firms to cut production costs by concentrating on process innovations. We also find suggestive evidence that, after reform, patents filed by the most

exposed firms receive 60% more citations than those filed by their less-exposed counterparts (although this estimate is insignificant). This result suggests that, despite the decrease in patent filings from exposed firms, they continued to undertake some high-value efforts to differentiate themselves from their competitors.

3.4.2 Changes in Market Structure

Table 3.5 presents our estimates of the changes in manufacturer entry and outsourcing resulting from equation (3.3.2). Panel (a) shows that there is a significant decrease in the rate of new entrants per year into affected DME categories by 25% relative to unaffected ones (see Figure C4 for the corresponding event study). This decrease is solely driven by a significant 47% drop in domestic entrants, while foreign firm entry rates did not significantly change and may have even increased slightly by 8%. These results support our earlier results indicating that foreign firms may capitalize on their cost advantages to compete in the US market by offering lower-cost products.

As domestic firm entry rates decline and foreign firm entry rates slightly increase over time, foreign-made products likely become more prevalent in the US DME market. However, foreign firm and product entry are not the only factors influencing the dominance of foreign-made products in the US market. Domestic firms may also adapt to price reform by outsourcing production to low-cost foreign contractors to maintain their position in the market.

Table 3.5 shows that the number of contracted manufacturers increases by 54% in affected DME categories relative to those unaffected, although not statistically significantly. However, we observe a statistically significant increase of 65% in foreign contracted manufacturers relative to the pre-reform mean (see Figure C5 for the corresponding event studies). Our results indicate that companies operating in affected DME categories are increasing their global presence by expanding their supply chains across different countries. Even domestic firms, which may face cost disadvantages alone, utilize low-cost foreign production by contracting with foreign firms.

In summary, we observe an increase in the number of foreign firms entering the US

market to compete, possibly due to their comparative advantages. Simultaneously, domestic firms outsource their production to foreign firms, potentially for the same reason. These results are consistent with our finding that firms most affected by the price reform increasingly prioritize process innovation over product innovation, which may further help manufacturers reduce production costs and offset the effects of price reform. However, an important question is whether these efforts to outsource production and prioritize process innovation may come at the expense of product quality, which could have significant implications for consumer welfare.

3.4.3 Changes in Product Quality

To examine the changes in product quality, we analyze the repair rates of DME. Since it takes time for new products to penetrate the consumer market, and the penetration is likely accelerated by the implementation (rather than just the announcement) of price cuts, we adopt later reference points in our preferred specification. Specifically, in addition to using the announcement dates as we do elsewhere in the paper, we also report results using the implementation date of the first price cut for each product category to define relative years, which is our preferred specification. Changes to repair rates likely take time to manifest as suppliers form new contracts with manufacturers to purchase newer products, and consumers take time to adopt and use them to the point where they might need repairs. Nonetheless, our results are robust to either relative year definition. To ensure valid estimation of the changes in the repair rate outcome, we employed a matching procedure to assign two control DME categories to each affected one. This approach not only allowed a sufficient number of control groups for reliable inference using Conley–Taber (i.e., two matched controls instead of one) but also enabled us to select control groups that were sufficiently similar to the treated groups. Without matching, the average pre-event repair rates among unaffected DME categories were 30 times higher than those in affected DME categories, posing challenges for trend comparisons. We note that when we naively scaled the outcomes of the entire sample of unaffected categories to the outcomes of the affected categories, we obtained similar results to those generated using our matching procedure.

We supplement our analysis of repair rates with adverse event rates. Our analysis of adverse event rates is only suggestive, given contemporaneous changes in CMS surveillance of DME categories, which may have influenced reporting rates.

Outcomes such as outsourcing, foreign firm entry, and a shift in focus to more cost-cutting innovation could plausibly lead to poorer product quality. To examine changes in product quality, we estimate changes in repair rates of DME equipment in affected categories relative to unaffected ones using an event-study design. Figure 3.4 presents our results, showing a sharp increase in repair rates after price reform, which persists and continues to rise over a longer horizon. By the last sample year, repair rates have significantly increased by as much as two percentage points, representing a 250% increase from pre-reform rates. Importantly, we do not observe significant pre-existing trends in treated groups relative to controls.

Table 3.6 provides further insights into the impact of price reform on repair rates. Our stacked difference-in-differences analysis reveals a significant 100–125% increase in repair rates over the post-reform sample, suggesting a decline in product quality. We present estimates for two reference periods: at the time of announcement and with a lead time to account for the time it takes for changes to materialize after the announcement. The results are statistically significant for both reference periods, though the lead-time reference is our preferred specification. A 125% increase in repair rates, representing a one-percentage-point increase, would result in an additional 700,000 DME repairs annually.

We also investigate the impact of price reform on adverse event reports related to affected DME categories and find suggestive evidence of a 157–233% increase, although the results are not statistically significant. However, in an effort to explore the mechanisms behind these decreases in product safety, we find substantial heterogeneity by firm type. Specifically, Table C39 shows that products from foreign manufacturers experienced a significant 470% increase in adverse events, with the largest increases observed in products made by foreign contractors, representing a 3,136% increase in adverse events relative to the pre-reform mean. These findings suggest that changes in product quality may be associated with increased contracting with foreign firms

after price reform. However, we are unable to draw definitive conclusions because two plausible explanations may account for this pattern. One possibility is that products made by foreign contractors are of lower quality, resulting in higher adverse event rates. Alternatively, as firms are increasingly outsourcing production to foreign manufacturers, the total number of products made by foreign contractors consumed in US markets increases, which may mechanically increase the number of adverse events due to increased utilization of those products, absent any differences in quality.

There are several caveats to consider when interpreting the adverse event analysis results. First, CMS increased surveillance of products within affected DME categories to assess the impact of price reform on product quality, which might have encouraged more reporting of adverse events, irrespective of changes in underlying safety. However, it is important to note that the reports we analyze originate from manufacturers and not CMS or users, although it is possible that manufacturers may have increased reporting in response to CMS's closer vigilance. Moreover, it is possible that CMS increasingly alerted manufacturers of adverse events, increasing reporting rates. Second, FDA adverse event report rates are not normalized by utilization, so changes in the number of adverse event reports could reflect changes in utilization rather than changes in safety. However, we find that utilization of affected DME decreases after price reform relative to unaffected DME, which suggests that such factors are unlikely to drive our results, as we observe an increase in adverse event reports. Finally, we acknowledge that our adverse event report analysis is suggestive, given these limitations.

3.5 Discussion and Conclusion

This paper examines the impact of health care price reform on innovation, market structure, and product quality using rich data from administrative sources and machine learning methods. We employ a stacked difference-in-differences estimator to compare outcomes of affected DME categories to those unaffected, and supplement this strategy with a comparison

of firms more exposed to price reform to those less exposed. Our analysis yields a set of three results. First, price reform reduces innovation in affected categories, especially among the most exposed firms. It also results in an increase of foreign-made products in the US market, possibly due to the ability of foreign firms to deliver products at a lower cost. Text analysis of the patents suggest a shift toward "process" innovation and away from "product" innovation, consistent with greater emphasis on production costs in the market. Second, US firm entry into affected DME categories decreases following the price reform, while foreign firm entry appears unaffected or slightly increases. US firms outsource production to foreign manufacturers in response to the cost-conscious US procurement environment. Lastly, we find evidence of reduced product quality, as manifested by increased repair rates and adverse event reports in affected categories relative to those unaffected, particularly among foreign firms and contractors.

The results of this study shed light on the long-term implications of government procurement policy. Governments face a crucial trade-off when setting procurement prices: while setting low prices can yield immediate savings, it risks altering the market structure, precipitating production offshoring, disrupting the trajectory of technological advancements, and potentially eroding product quality. These dynamic, long-term effects must be considered when designing policies as they may diminish or even outweigh the immediate cost-savings. Thus, understanding the interplay between price reforms and their far-reaching impacts is crucial for designing optimal procurement policy and is increasingly important in light of recent interest in additional price reforms, such as those proposed in the Inflation Reduction Act, and ongoing efforts to reduce health care costs.

Our paper raises several important questions that remain unanswered. For example, are there labor market responses to health care price reforms: does employment change among DME suppliers and manufacturing plants? Furthermore, are there price or quantity responses from private insurers following the Medicare price reform that may have either amplified or dampened its overall impact? Lastly, could lower-quality products be associated with poorer long-run health outcomes and increased expenditures that may offset some of the initial savings?

Further research is needed to address these questions and gain a deeper understanding of the implications of health care price reforms.

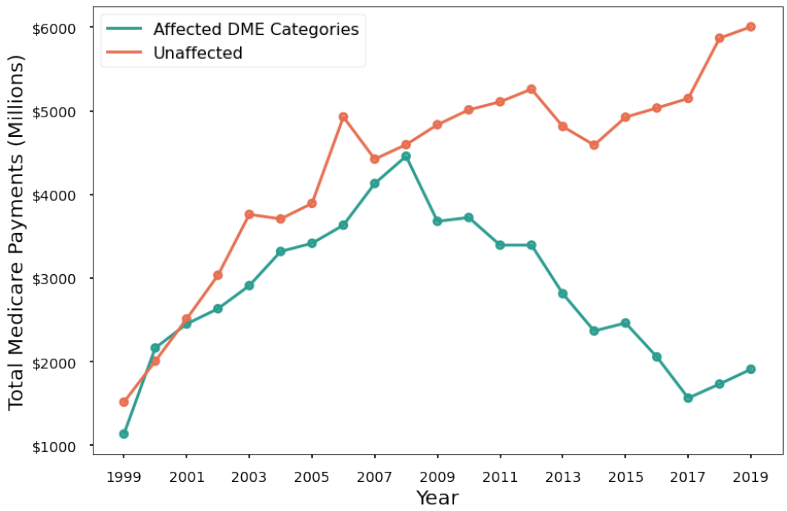
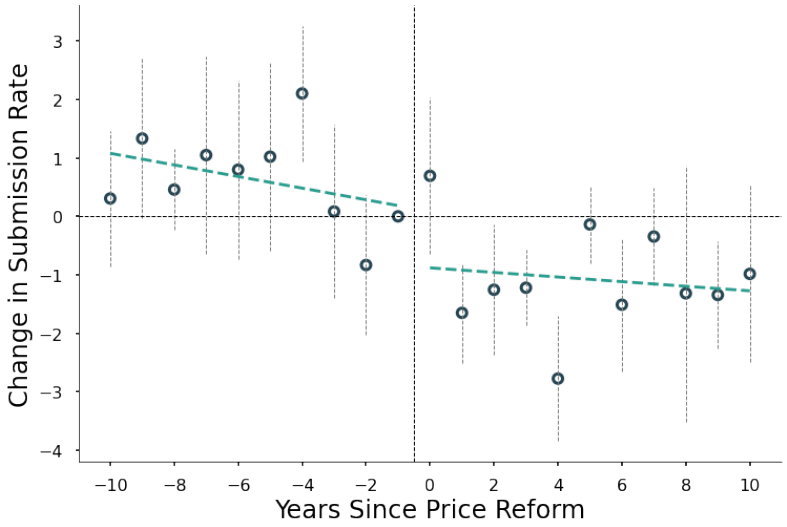
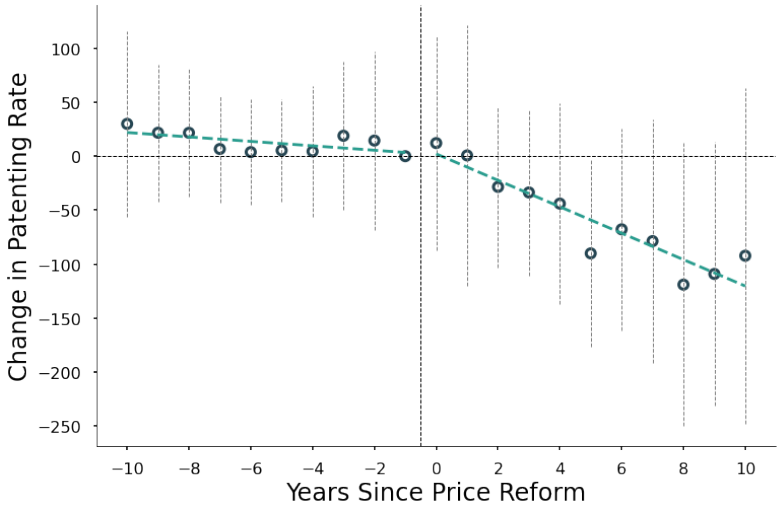


Figure 3.1. Raw Trends of Total Medicare DME Payments. Note: The figure plots total Medicare DME payments separately for DME in categories subject to the price reform and those that are not. The sample includes all Traditional Medicare DME purchases each year. The y-axis values are given in millions of US dollars.

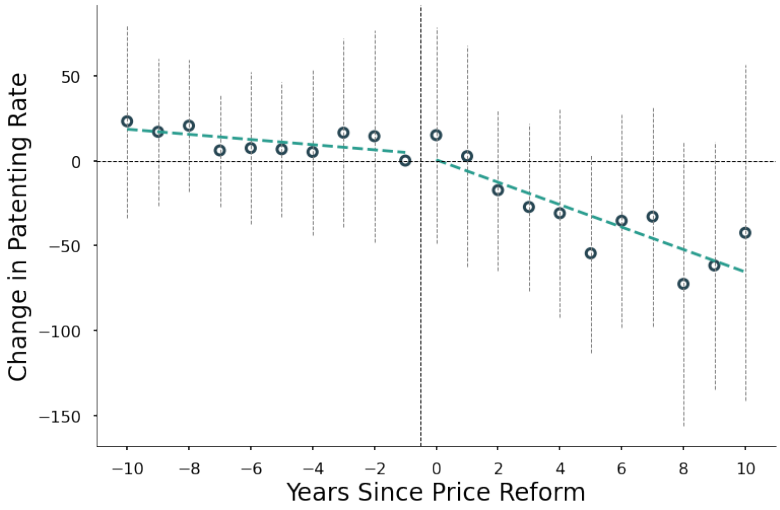


(a) Pre-Market Approval and 510(k) Count

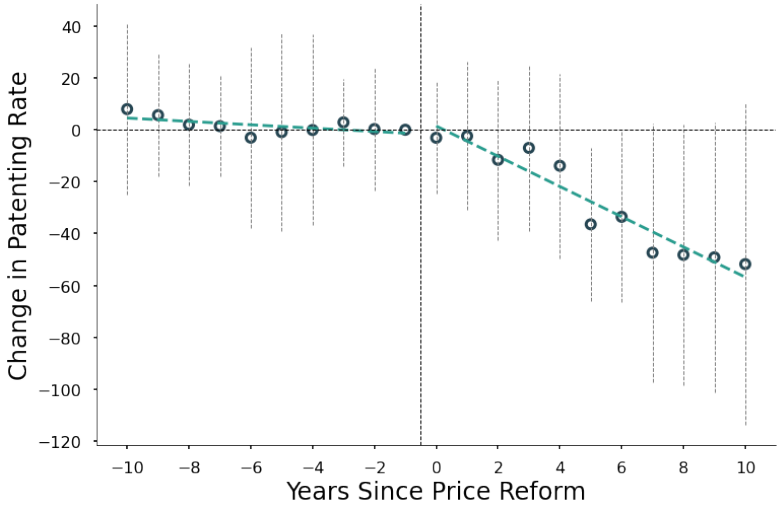


(b) Patent Count

Figure 3.2. Event Study: Innovation. Note: The figure presents the coefficients obtained from estimating equation (3.3.1) for our FDA submissions and patent count outcomes. It illustrates the temporal evolution of outcomes in DME categories affected by the event, relative to those unaffected, with a reference period at $t = -1$. Panel (a) presents our event-study estimates for changes in the number of PMAs and 510(k)s submitted to the FDA and panel (b) provides the estimates for the changes in the number of patents filed annually. 95% confidence intervals are provided.



(a) Foreign Patent Count



(b) US Patent Count

Figure 3.3. Event Study: Patent Count by Origin. Note: The figure presents the coefficients obtained from estimating equation (3.3.1) for the patenting rate outcomes, separately for patents filed in the US and abroad. It illustrates the temporal evolution of outcomes in DME categories affected by the event, relative to those unaffected, with a reference period at $t = -1$. Panel (a) presents the event-study estimates for changes in the patenting rate (per year) of patents filed abroad, while panel (b) presents estimates for the changes in the patenting rate of patents filed in the US. 95% confidence intervals are provided.

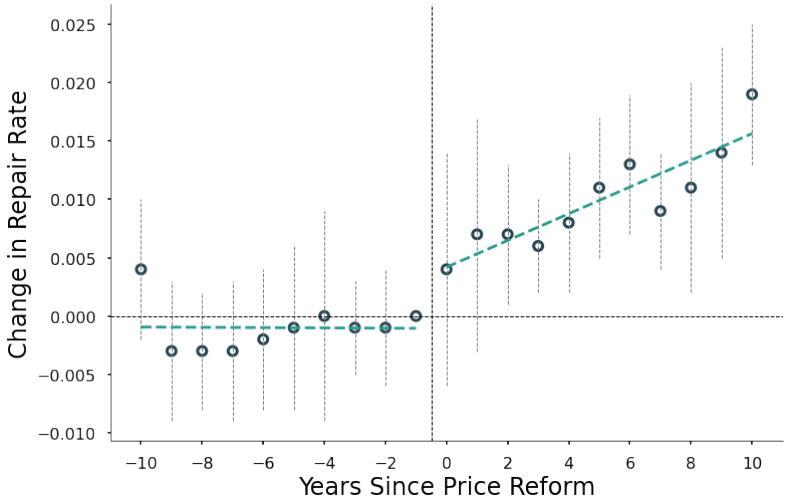


Figure 3.4. Event Study: Change in Product Quality. Note: The figure presents the coefficients obtained from estimating equation (3.3.1) for the repair rate outcome. It illustrates the temporal evolution of outcomes in DME categories affected by the event, relative to those unaffected, with a reference period at $t = -1$. 95% confidence intervals are provided.

Table 3.1. Summary Statistics of Product Categories. Note: The table reports summary statistics from the year before the announcement of the price reform (2005) and all sample years (1996-2016), separately for all product categories, product categories that were subject to price reform, and product categories that were not subject to the reform during the sample period. Panel (a) presents summary statistics for our measures of innovation, and panel (b) presents those for our utilization measures.

	All DME Categories		Categories Affected by Price Reform		Categories Unaffected by Price Reform	
	2005	*1996-2016*	*2005*	*1996-2016*	*2005*	*1996-2016*
***Panel (a)** Product Innovation*						
Number of Patents	10,371	236,656	1,714	43,567	8,657	193,089
Number of PMA/510(k)'s	105	3,738	67	2,306	38	1,432
	2005		*2005*		*2005*	
***Panel (b)** Product Utilization*						
Number of Medicare Users (Millions)	104		59		45	
Medicare Total Expenditures (Millions)	$7,309.6		$3,416.3		$3,893.3	
Number of Categories	56		13		43	

Table 3.2. Summary Statistics of DME Manufacturers. Note: The table reports summary statistics on the number of patents filed of different types across the firms included in our firm-level analysis, separately for 2005 and the full sample. P25 signifies firms in the 25th percentile, while P50 and P75 follow accordingly.

	2005					*1996-2016*				
	Mean	S.D.	P25	P50	P75	Mean	S.D.	P25	P50	P75
Number of Health-Related Patents	16.83	46.57	0	2	13	334.07	746.61	19	68	270
Number of DME Patents	2.83	8.48	0	0	2	57.48	125.02	4	14	50
Number of Affected DME Patents	0.93	2.83	0	0	1	22.65	62.77	2	5	15
Number of Unaffected DME Patents	1.89	7.18	0	0	1	34.84	88.91	0	4.5	29.75
Share of Portfolio Affected by Price Reform	0.24	0.31	0.03	0.10	0.34	-	-	-	-	-
Number of Manufacturers	486									

Table 3.3. Impact of Price Reform on Innovation. Note: The table presents results from estimating equations (3.3.2) and (3.3.4) for our innovation outcomes. Column (1) reports the pre-event (before price reform) mean across treated groups. Column (2) presents the estimates, with standard errors reported in parentheses below the estimates. Column (3) shows the percent change in the outcome relative to the pre-event mean. Panel (a) describes estimates from our DME category-level analysis, with overall totals and differentiation between origins (i.e., US and foreign), and panel (b) presents estimates from our firm-level analysis, measuring changes in the likelihood that a firm files at least one patent in a given year. Statistical significance is denoted by +, *, **, and *** correspond to significance levels of 0.10, 0.05, 0.01, and 0.001 levels, respectively.

		Change with Price Reform	
	Pre-Event Mean (1)	Estimate (2)	% Change (3)
Panel (a) *DME Category Level*			
Number of PMA/510(k)'s per Year	6.38	-1.40***	-22%
		(0.35)	
From US Firms	5.0	–2.431***	-49%
		(0.27)	
From Foreign Firms	1.38	0.75*	54%
		(0.31)	
Number of Patents per Year	196.92	-55.73	-29%
		(27.39)	
Filed in the US	76.30	-20.64	-27%
		(27.38)	
Filed Elsewhere	120.62	-36.66	-30%
		(43.63)	
Panel (b) *Firm Level*			
Pr of Filing Affected DME Patent	0.35	-0.10***	-29%
		(0.03)	
Pr of Filing Unaffected DME Patent	0.15	0.01	0.07%
		(0.03)	

Table 3.4. Impact of Price Reform on Direction and Quality of Innovation. Note: The table presents results from estimating equations (3.3.2) and (3.3.4) for our direction and quality of innovation outcomes. Column (1) reports the pre-event (before price reform) mean across treated groups. Column (2) presents the estimates, with standard errors reported in parentheses below the estimates. Column (3) shows the percent change in the outcome relative to the pre-event mean. Table reports estimates of the change in the direction of innovation or citations at the firm level. The quantity $t = 0$ refers to the reference period (i.e., the period we consider the policy enacted). Statistical significance is denoted by +, *, **, and *** correspond to significance levels of 0.10, 0.05, 0.01, and 0.001 levels, respectively.

	Change with Price Reform		
	Pre-Event Mean	Estimate	% Change
	(1)	(2)	(3)
Share of Patents on Process Innovation	0.31	0.10*	32%
		(0.05)	
Share of Patents on Product Innovation	0.69	-0.10*	-15%
		(0.05)	
Citations per Patent	12.95	8.00	60%
		(6.77)	

Table 3.5. Impact of Price Reform on Manufacturer Entry and Outsourcing. Note: The table presents results from estimating equation (3.3.2) for our entry and outsourcing outcomes. Column (1) reports the pre-event (before price reform) mean across treated groups. Column (2) presents the estimates, with standard errors reported in parentheses below the estimates. Column (3) shows the percent change in the outcome relative to the pre-event mean. Panel (a) describes estimates of the number of entrants at the DME-category level, with overall totals and differentiation between origins, and panel (b) presents estimates of the number of new contractors at the DME-category level, with overall totals and differentiation between origins. Statistical significance is denoted by +, *, **, and *** correspond to significance levels of 0.10, 0.05, 0.01, and 0.001 levels, respectively.

	Change with Price Reform		
	Pre-Event Mean (1)	Estimate (2)	% Change (3)
Panel (a) *Number of Entrants*			
All Entrants	2.38	-0.59** (0.20)	-25%
US Entrants	1.88	-0.88*** (0.14)	-47%
Foreign Entrants	0.50	0.04 (0.09)	8%
Panel (b) *Number of New Contractors*			
All Contractors	4.00	2.15 (1.63)	54%
US Contractors	0.92	0.153 (0.59)	17%
Foreign Contractors	3.08	2.00* (1.01)	65%

Table 3.6. Impact of Price Reform on Product Quality. Note: The table presents results from estimating equations (3.3.2) and (3.3.4) for our product quality outcomes. Column (1) reports the pre-event (before price reform) mean across treated groups. Column (2) presents the estimates, with standard errors reported in parentheses below the estimates. Column (3) shows the percent change in the outcome relative to the pre-event mean. Table reports estimates of changes in product quality measures at the DME-category level. The quantity $t = 0$ refers to the reference period (i.e., the period we consider the policy enacted). Statistical significance is denoted by +, *, **, and *** correspond to significance levels of 0.10, 0.05, 0.01, and 0.001 levels, respectively.

		Change with Price Reform	
	Pre-Event Mean (1)	Estimate (2)	% Change (3)
Repair Rate			
$t = 0$ at announcement	0.008	0.008** (0.003)	100%
$t = 0$ after lead time	0.008	0.01*** (0.003)	125%
Adverse Event Reports			
$t = 0$ at announcement	927.7	2,163.9 (1,340.6)	233%
$t = 0$ after lead time	1,464.5	2,300.3 (1,480.5)	157%

3.6 Chapter Acknowledgements

Chapter 3, in full, is currently being prepared for publication of the material. This project was co-authored with Yunan Ji. The dissertation author was a primary investigator and author of this paper.

Bibliography

Abbott, Thomas A and John A Vernon, "The Cost of Us Pharmaceutical Price Regulation: A Financial Simulation Model of R&D Decisions," *Managerial and Decision Economics*, 2007, *28* (4-5), 293–306.

Acemoglu, D., "Why do new technologies complement skills? Directed technical change and wage inequality," *The Quarterly Journal of Economics*, 1998, *113* (4), 1055–1089.

_ **and J. Linn**, "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry," *The Quarterly Journal of Economics*, 2004, *119* (3), 1049–1090.

Acemoglu, Daron and Joshua Linn, "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry," *Quarterly Journal of Economics*, 2004, *119* (3), 1049–1090.

Acharya, Viral, Ramin Baghai, and Krishnamurthy Subramanian, "Labor Laws and Innovation," *Journal of Law and Economics*, 2013, *56* (4), 997–1037.

_ **,** _ **, and** _ , "Wrongful Discharge Laws and Innovation," *The Review of Financial Studies*, 2014, *27* (1), 301–346.

Agarwal, Ruchir and Patrick Gaule, "What Drives Innovation? Lessons from COVID-19 R&D," *IZA Working Paper 14079*, 2021.

Aghion, Philippe, Antonin Bergeaud, and John Van Reenen, "The Impact of Regulation on Innovation," *NBER Working Paper 28381*, 2019.

_ **, Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt**, "Competition and Innovation: An Inverted-U Relationship," *Quarterly Journal of Economics*, 2005, *120* (2), 701–728.

_ **, Richard Blundell, Rachel Griffith, Peter Howitt, and Susanne Prantl**, "The Effects of Entry on Incumbent Innovation and Productivity," *The Review of Economics and Statistics*, 2009, *91* (1), 20–32.

Andrews, Michael, "Comparing historical patent datasets," *Available at SSRN 3415318*, 2019.

Arad, Nitzan and Mark B McClellan, "Drug Pricing Reform In The Inflation Reduction Act: What Are The Implications? Part 1," *Health Affairs Forefront*, 2022.

Arrow, Kenneth J, "The Economic Implications of Learning by Doing," in "Readings in the Theory of Growth," Springer, 1971, pp. 131–149.

Arts, Sam, Bruno Cassiman, and Juan Carlos Gomez, "Text matching to measure patent similarity," *Strategic Management Journal*, 2018, *39* (1), 62–84.

Athey, Susan, "The Impact of Machine Learning on Economics," in Ajay K. Agrawal, Joshua Gans, and Avi Goldfarb, eds., *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press, January 2018.

Azoulay, Pierre, Joshua S. Graff Zivin, and Bhaven N Sampat, "The Diffusion of Scientific Knowledge Across Time and Space: Evidence from Professional Transitions for the Superstars of Medicine," *NBER Working Paper 16683*, 2011.

Barnes, J.K., *Artificial Limbs* Circular Order, Office of the Surgeon General, 1865.

_ **and Edwin Stanton**, *Artificial Limbs Furnished to Soldiers* Ex. Doc. 108, Department of War, 1866.

Beckerman, Daniel, Melissa Esparza, Sun Ik Lee, Sigurd H Berven, S Samuel Bederman, Serena S Hu, Shane Burch, Vedat Deviren, Bobby Tay, Praveen V Mummaneni et al., "Cost Analysis of Single-Level Lumbar Fusions," *Global Spine Journal*, 2020, *10* (1), 39–46.

Bena, Jan and Elena Simintzi, "Machines Could Not Compete With Chinese Labor: Evidence From Us Firms' Innovation," *Available at SSRN 2613248*, 2022.

Bergeaud, A., Y. Potiron, and J. Raimbault, "Classifying patents based on their semantic content," *PLoS ONE*, April 2017, *12*.

Bergstra, James and Yoshua Bengio, "Random Search for Hyper-parameter Optimization," *J. Mach. Learn. Res.*, feb 2012, *13*, 281–305.

Berkes, Enrico, "Comprehensive Universe of U.S. Patents (CUSP): Data and Facts," *Unpublished Working Paper*, 2018.

_ **and Peter Nencka**, "Novel Ideas: The Effects of Carnegie Libraries on Innovation," *Unpublished Working Paper*, 2019.

_ **, Ruben Gaetani, and Martí Mestieri**, "Cities and Technology Cycles," *Unpublished Working Paper*, 2019.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, "How Much Should We Trust Differences-in-Differences Estimates?," *The Quarterly Journal of Economics*, 2004, *119* (1).

Blume-Kohout, Margaret E and Neeraj Sood, "Market size and innovation: Effects of Medicare Part D on pharmaceutical research and development," *Journal of Public Economics*, 2013, *97*, 327–336.

_ and _ , "Market Size and Innovation: Effects of Medicare Part D on Pharmaceutical Research and Development," *Journal of public economics*, 2013, *97*, 327–336.

Boomhower, Judson, "Drilling Like There's No Tomorrow: Bankruptcy, Insurance, and Environmental Risk," *American Economic Review*, 2019, *109* (2), 391–426.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, "Revisiting Event Study Designs: Robust and Efficient Estimation," 2021.

Bragg, Jennifer L., Maya P. Florence, and Amanda H. Chan, "Failure to Report Adverse Events Results in Criminal Misbranding Settlement and Individual Liability," 2018.

Breiman, Leo, "Random Forests," *Machine Learning*, Oct 2001, *45* (1), 5–32.

Brodersen, Kay Henning, Cheng Soon Ong, Klaas Enno Stephan, and Joachim M Buhmann, "The balanced accuracy and its posterior distribution," in "2010 20th International Conference on Pattern Recognition" IEEE 2010, pp. 3121–3124.

Budish, Eric, Benjamin N Roin, and Heidi Williams, "Do Firms Underinvest in Long-Term Research? Evidence from Cancer Clinical Trials," *American Economic Review*, 2015, *105* (7), 2044–85.

Buera, Francisco J and Yongseok Shin, "Financial Frictions and the Persistence of History: A Quantitative Exploration," *Journal of Political Economy*, 2013, *121* (2), 221–272.

Buettner, Bettina, "Entry Barriers and Growth," *Economics Letters*, 2006, *93* (1), 150–155.

Busso, Matias and Sebastian Galiani, "The Causal Effect of Competition on Prices and Quality: Evidence from a Field Experiment," *American Economic Journal: Applied Economics*, 2019, *11* (1), 33–56.

Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller, "Bootstrap-based improvements for inference with clustered errors," *The Review of Economics and Statistics*, 2008, *90* (3), 414–427.

Carpenter, Daniel, "The Political Economy of FDA Drug Review: Processing, Politics, and Lessons for Policy," *Health Affairs*, 2004, *23* (1), 52–63.

_ , "Protection Without Capture: Product Approval by a Politically Responsive, Learning Regulator," *American Political Science Review*, 2004, *98* (4), 613–631.

_ **, Justin Grimmer, and Eric Lomazoff**, "Approval Regulation and Endogenous Consumer Confidence: Theory and Analogies to Licensing, Safety, and Financial Regulation," *Regulation & Governance*, 2010, *4* (4), 383–407.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer, "The Effect of Minimum Wages on Low-Wage Jobs," *Quarterly Journal of Economics*, 2019, *134* (3), 1405–1454.

Center for Devices and Radiological Health, "Learn if a Medical Device Has Been Cleared by FDA for Marketing," 2018.

Chandra, Amitabh and Jonathan Skinner, "Technology growth and expenditure growth in health care," *Journal of Economic Literature*, 2012, *50* (3), 645–80.

Che, Yeon-Koo, Elisabetta Iossa, and Patrick Rey, "Prizes versus Contracts as Incentives for Innovation," *The Review of Economic Studies*, 2021, *88* (5), 2149–2178.

Civan, Abdulkadir and Michael T Maloney, "The Effect of Price on Pharmaceutical R&D," *The BE Journal of Economic Analysis & Policy*, 2009, *9* (1).

Clemens, Jeffrey, "The effect of us health insurance expansions on medical innovation," *NBER Working Paper 19761*, 2013.

_ **and Parker Rogers**, "Demand Shocks, Procurement Policies, and the Nature of Medical Innovation: Evidence from Wartime Prosthetic Device Patents," *NBER Working Paper 26679*, 2020.

Coase, Ronald H, "The Problem of Social Cost," in "Classic Papers in Natural Resource Economics," Springer, 1960, pp. 87–137.

Cockburn, Iain M., Rebecca Henderson, and Scott Stern, "The Impact of Artificial Intelligence on Innovation: An Exploratory Analysis," in Ajay K. Agrawal, Joshua Gans, and Avi Goldfarb, eds., *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press, January 2018.

Conley, Timothy G and Christopher R Taber, "Inference with "Difference in Differences" With a Small Number of Policy Changes," *The Review of Economics and Statistics*, 2011, *93* (1), 113–125.

Costello, Kevin and Christopher Pham, "Regulatory Preemption of Medical Devices," 2016.

Cozzi, Guido and Giammario Impullitti, "Government Spending Composition, Technical

Change, and Wage Inequality," *Journal of the European Economic Association*, 2010, *8* (6), 1325–1358.

Cubanski, Juliette and Tricia Neuman, "What to Know About Medicare Spending and Financing," Jan 2023.

Cutler, D.M., *Your Money or Your Life: Strong Medicine for America's Health Care System*, Oxford University Press, USA, 2004.

de Chaisemartin, Clément and Xavier d'Haultfoeuille, "Two-Way Fixed Effects Estimators With Heterogeneous Treatment Effects," *NBER Working Paper 25904*, 2019.

Dechezlepretre, Antoine, David Hemous, Morten Olsen, and Carlo Zanella, "Automating Labor: Evidence from Firm-level Patent Data," *Unpublished Working Paper*, 2019.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *CoRR*, 2018, *abs/1810.04805*.

Diamond, A. Abadie A. and J. Hainmueller, "Synthetic control methods for comparative case studies: Estimating the effect of California's Tobacco Control Program," *Journal of the American Statistical Association*, 2010, *105* (490), 493–505.

Djankov, Simeon, Caralee McLiesh, and Rita Maria Ramalho, "Regulation and Growth," *Economics letters*, 2006, *92* (3), 395–401.

Dobyns, Kenneth W, *The patent office pony: a history of the early patent office*, Sergeant Kirklands Museum &, 1994.

Dranove, David, Craig Garthwaite, Christopher Heard, and Bingxiao Wu, "The Economics of Medical Procedure Innovation," *Journal of Health Economics*, 2022, *81*, 102549.

Dubois, Pierre, Olivier De Mouzon, Fiona Scott-Morton, and Paul Seabright, "Market size and pharmaceutical innovation," *RAND Journal of Economics*, 2015, *46* (4), 844–871.

Dumais, Richard Harshman Scott Deerwester Susan T., "Indexing by Latent Semantic Analysis," *JASIS*, 1990, *41*, 391–407.

Ehrlich, Isaac and Richard A Posner, "An Economic Analysis of Legal Rulemaking," *Journal of Legal Studies*, 1974, *3* (1), 257–286.

Emergo, "Emergo Survey: Regulatory Issues Remain Biggest Challenge for Most Medical Device Companies," May 2019.

_ , "FDA eMDR Adverse Event Reporting for Medical Device Companies," 2022.

Ensign, Lisa Garnsey and K Bretonnel Cohen, "A Primer to the Structure, Content and Linkage of the FDA's Manufacturer and User Facility Device Experience (MAUDE) Files," *eGEMs*, 2017, *5* (1), 12.

FDA, "Medical Devices; Proposed Reclassification and Exemption From Premarket Notification for Certain Classified Devices," *Federal Register*, July 1995, *60*, 38902–38916.

_ , "CFR - Code of Federal Regulations Title 21," 2020.

_ , "Fact Sheet: FDA at a Glance," 2020.

_ , "Medical Device Reporting (MDR): How to Report Medical Device Problems," 2020.

Filson, Darren, "A Markov-Perfect Equilibrium Model of the Impacts of Price Controls on the Performance of the Pharmaceutical Industry," *The RAND Journal of Economics*, 2012, *43* (1), 110–138.

Finkelstein, A., "Static and dynamic effects of health policy: Evidence from the vaccine industry," *The Quarterly Journal of Economics*, 2004, *119* (2), 527–564.

Finkelstein, Amy, "Static and Dynamic Effects of Health Policy: Evidence from the Vaccine Industry," *Quarterly Journal of Economics*, 2004, *119* (2), 527–564.

Frank, Richard G and Len M Nichols, "Medicare Drug-Price Negotiation—Why Now... and How," *New England Journal of Medicine*, 2019, *381* (15), 1404–1406.

Friedman, Jerome, "Greedy function approximation: a gradient boosting machine," *Annals of Statistics*, 2001, *29*, 1189–1232.

Fuhr, Ted, Evgeniya Makarova, Steve Silverman, and Vanya Telpis, "Capturing the Value of Good Quality in Medical Devices," Jan 2018.

Gaglani, Shiv, "Investing in Medical Devices: Interview with Venture Capitalist Dave Eichler of Psilos," Jun 2014.

Galasso, Alberto and Hong Luo, "Tort reform and innovation," *The journal of law and economics*, 2017, *60* (3), 385–412.

_ **and** _ , "When Does Product Liability Risk Chill Innovation? Evidence from Medical Implants," *NBER Working Paper 25068*, 2018.

_ **and** _ , "When does product liability risk chill innovation? Evidence from medical implants,"

American Economic Journal: Economic Policy, 2022, *14* (2), 366–401.

Garcia, Diego, "Sentiment during recessions," *The Journal of Finance*, 2013, *68* (3), 1267–1300.

Gentzkow, Matthew and Jesse M Shapiro, "What drives media slant? Evidence from US daily newspapers," *Econometrica*, 2010, *78* (1), 35–71.

_ , Jesse Shapiro, and Matt Taddy, "Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech," *Econometrica*, 2019, *87* (4), 1307–1340.

Giaccotto, Carmelo, Rexford E Santerre, and John A Vernon, "Drug Prices and Research and Development Investment Behavior in the Pharmaceutical Industry," *The Journal of Law and Economics*, 2005, *48* (1), 195–214.

Glaeser, Edward, Simon Johnson, and Andrei Shleifer, "Coase versus the Coasians," *The Quarterly Journal of Economics*, 2001, *116* (3), 853–899.

Goodman-Bacon, Andrew, "Difference-in-differences With Variation in Treatment Timing," *NBER Working Paper 25018*, 2018.

Graske, Theodore Wesley, *The Law of Government Defense Contracts*, Baker, Voorhis & Company, 1941.

Grennan, Matthew and Ashley Swanson, "Transparency and Negotiated Prices: The Value of Information in Hospital-Supplier Bargaining," *Journal of Political Economy*, 2020, *128* (4), 1234–1268.

_ and Robert J. Town, "Regulating Innovation with Uncertain Quality: Information, Risk, and Access in Medical Devices," *American Economic Review*, 2020, *110* (1), 120–61.

_ and Robert J Town, "Regulating innovation with uncertain quality: information, risk, and access in medical devices," *American Economic Review*, 2020, *110* (1), 120–61.

Guendling, Benjamin L, "Product-Liability Risk Exposure in the U.S. And Europe: Similar but Still Separate and Distinct," *Michigan Bar Journal*, 2016, *95*, 18–21.

Guyatt, Mary, "Better legs: artificial limbs for British veterans of the First World War," *Journal of Design History*, 2001, *14* (4), 307–325.

Guyon, Isabelle and André Elisseeff, "An Introduction to Variable and Feature Selection," *J. Mach. Learn. Res.*, mar 2003, *3*, 1157–1182.

_ , Jason Weston, Stephen Barnhill, and Vladimir Vapnik, "Gene Selection for Cancer

Classification using Support Vector Machines," *Machine Learning*, Jan 2002, *46* (1), 389–422.

Hahn, Robert W and John A Hird, "The Costs and Benefits of Regulation: Review and Synthesis," *Yale Journal on Regulation*, 1991, *8* (1), 233.

Hall, B.H., A.B. Jaffe, and M. Trajtenberg, "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools," *NBER Working Paper 8498*, 2001.

Hanlon, W Walker, "Necessity is the mother of invention: Input supplies and Directed Technical Change," *Econometrica*, 2015, *83* (1), 67–100.

Hasegawa, Guy R, *Mending Broken Soldiers: The Union and Confederate Programs to Supply Artificial Limbs*, SIU Press, 2012.

Hémous, David and Morten Olsen, "The rise of the machines: Automation, horizontal innovation, and income inequality," *American Economic Journal: Macroeconomics*, 2022, *14* (1), 179–223.

Hochreiter, Sepp and Jürgen Schmidhuber, "Long Short-Term Memory," *Neural Computation*, 1997, *9* (8), 1735–1780.

Hook, Daniel W., Simon J. Porter, and Christian Herzog, "Dimensions: Building Context for Search and Evaluation," *Frontiers in Research Metrics and Analytics*, 2018, *3*, 23. https://www.frontiersin.org/articles/10.3389/frma.2018.00023/pdf.

Houston, M.H., J. Bolton, and L.S. Joynes, "Report of the Richmond Medical Journal Commission," *Richmond Medical Journal*, 1866, pp. 564–571.

Hua, Jianping, Zixiang Xiong, James Lowey, Edward Suh, and Edward R. Dougherty, "Optimal number of features as a function of sample size for various classification rules," *Bioinformatics*, 11 2004, *21* (8), 1509–1515.

Iaria, Alessandro, Carlo Schwarz, and Fabian Waldinger, "Frontier Knowledge and Scientific Production: Evidence from the Collapse of International Science*," *The Quarterly Journal of Economics*, 01 2018, *133* (2), 927–991.

Imbens, Guido W and Paul R Rosenbaum, "Robust, accurate confidence intervals with a weak instrument: quarter of birth and education," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2005, *168* (1), 109–126.

Institute of Medicine, *Public Health Effectiveness of the FDA 510(k) Clearance Process: Balancing Patient Safety and Innovation: Workshop Report*, The National Academies Press, 2010.

_ , *Medical Devices and the Public's Health: The FDA 510(k) Clearance Process at 35 Years*, The National Academies Press, 2011.

Isakov, Leah, Andrew W Lo, and Vahid Montazerhodjat, "Is the FDA Too Conservative or Too Aggressive?: A Bayesian Decision Analysis of Clinical Trial Design," *Journal of Econometrics*, 2019, *211* (1), 117–136.

Ji, Yunan, "Can Competitive Bidding Work in Health Care? Evidence from Durable Medical Equipment.," *Working Paper*, 2023.

Johnson, Judith Ann, "FDA Regulation of Medical Devices," 2012.

Jr, James R Withrow, "Control of War Profits in the United States and Canada," *U. Pa. L. Rev.*, 1942, *91*, 194.

Kelsey v. Alcon Laboratories, Inc., "Kelsey v. Alcon Laboratories, Inc.," *No. 180902756 [2019] WL 1884225 (Utah Dist. Ct. Apr. 22, 2019).*

Kessler, Daniel P, *Regulation Versus Litigation: Perspectives from Economics and Law*, University of Chicago Press, 2010.

Khoury, Amir H. and Ron Bekkerman, "AUTOMATIC DISCOVERY OF PRIOR ART: BIG DATA TO THE RESCUE OF THE PATENT SYSTEM," *The John Marshall Review of Intellectual Property Law*, 2016, *16.*

Kim, Yoon, "Convolutional Neural Networks for Sentence Classification," *CoRR*, 2014, *abs/1408.5882.*

Knight, H Jackson, *Confederate Invention: The Story of the Confederate States Patent Office and its Inventors*, LSU Press, 2011.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman, "Technological Innovation, Resource Allocation, and Growth," *Quarterly Journal of Economics*, 2017, *132* (2), 665–712.

Kolstad, Charles D, Thomas S Ulen, and Gary V Johnson, "Ex Post Liability for Harm vs. Ex Ante Safety Regulation: Substitutes or Complements?," *American Economic Review*, 1990, *80* (4), 888–901.

Kowalsky, Meaghan Melissa Marie, "Enabling the Great War: Ex-Servicemen, the Mixed Economy of Welfare and the Social Construction of Disability, 1899-1930." PhD dissertation, University of Leeds 2007.

Linker, Beth, *War's Waste: Rehabilitation in World War I America*, University of Chicago Press,

2011.

Magerman, T., B. Van Looy, B. Baesens, and K. Debackere, "Assessment of Latent Semantic Analysis (LSA) text mining algorithms for large scale mapping of patent and scientific publication documents," *University of Leuven Working Paper*, 2011.

Makower, Josh, Aabed Meer, and Lyn Denend, "FDA Impact on US Medical Technology Innovation; A Survey of Over 200 Medical Technology Companies," 2010.

McAfee, R Preston and John McMillan, "Government Procurement and International Trade," *Journal of International Economics*, 1989, *26* (3-4), 291–308.

Meier, B., "Costs Surge for Medical Devices, but Benefits are Opaque," 2009.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean, "Distributed Representations of Words and Phrases and their Compositionality," in C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, eds., *Advances in Neural Information Processing Systems 26*, Curran Associates, Inc., 2013, pp. 3111–3119.

Moll, Benjamin, "Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation?," *American Economic Review*, 2014, *104* (10), 3186–3221.

Moser, Petra, Alessandra Voena, and Fabian Waldinger, "German Jewish émigrés and US invention," *American Economic Review*, 2014, *104* (10), 3222–55.

Moses, Mark W, Paola Pedroza, Ranju Baral, Sabina Bloom, Jonathan Brown, Abby Chapin, Kelly Compton, Erika Eldrenkamp, Nancy Fullman, John Everett Mumford et al., "Funding and Services Needed to Achieve Universal Health Coverage: Applications of Global, Regional, and National Estimates of Utilisation of Outpatient Visits and Inpatient Admissions From 1990 to 2016, and Unit Costs From 1995 to 2016," *The Lancet Public Health*, 2019, *4* (1), 49–73.

Mulligan, Casey B, "Peltzman Revisited: Quantifying 21st Century Opportunity Costs of FDA Regulation," *NBER Working Paper 29574*, 2021.

Munford, Luther T., "Next Decade To-Do: Enforce Preemption for Class Ii Devices With Special Controls," July 2018.

Navathe, Amol S, Ezekiel J Emanuel, Sherry Glied, Farzad Mostashari, and Bob Kocher, "Medicare Payment Reform's Next Decade: A Strategic Plan for the Center for Medicare and Medicaid Innovation," *Health Aff Blog*, 2020.

Novotny, Jennifer, "To'take their place among the productive members of society': Vocational rehabilitation of WWI wounded at Erskine," *Wellcome open research*, 2017, *2*.

Olson, Mary K, "Firm Characteristics and the Speed of FDA Approval," *Journal of Economics & Management Strategy*, 1997, *6* (1), 377–401.

Peltzman, Sam, "An Evaluation of Consumer Protection Legislation: The 1962 Drug Amendments," *Journal of Political Economy*, 1973, *81* (5), 1049–1091.

Philipson, Tomas and Troy Durie, "The Evidence Base on the Impact of Price Controls on Medical Innovation," Technical Report, Tech. rep., Becker Friedman Institute Working Papers 2021.

Philipson, Tomas J, Eric Sun, and Dana Goldman, "The Effects of Product Liability Exemption in the Presence of the FDA," in "Regulation vs. Litigation: Perspectives From Economics and Law," University of Chicago Press, 2010, pp. 137–163.

Pietzsch, Jan B, Marta G Zanchi, and John H Linehan, "Medical Device Innovators and the 510 (K) Regulatory Pathway: Implications of a Survey-Based Assessment of Industry Experience," *Journal of Medical Devices*, 2012, *6* (2).

Pisani, Robert P, "Federal Preemption Defenses in Product Liability Suits: A Continuing Evaluation," *IDC Quarterly*, 2011, *21* (1).

Popp, David, "Using the Triadic Patent Family Database to Study Environmental Innovation," *Environment Directorate Working Paper*, 2005, *2*.

_ , "Innovation and climate policy," *Annu. Rev. Resour. Econ.*, 2010, *2* (1), 275–298.

_ , "Environmental Policy and Innovation: A Decade of Research," *NBER Working Paper 25631*, 2019.

Powell, Spenser F, "Changing Our Minds: Reforming the FDA Medical Device Reclassification Process," *Food & Drug LJ*, 2018, *73*, 177.

Reinhardt, Uwe E, "The Pricing of US Hospital Services: Chaos Behind a Veil of Secrecy," *Health Affairs*, 2006, *25* (1), 57–69.

Riegel v. Medtronic, Inc., "Riegel v. Medtronic, Inc.," *552 U.S. 312, 323-24 (2008).*

Rogers, Parker, "Regulating the Innovators: Approval Costs and Innovation in Medical Technologies," Technical Report 2023.

Romer, Paul M, "Endogenous Technological Change," *Journal of Political Economy*, 1990, *98* (5, Part 2), S71–S102.

Rosenblatt, Frank, "Frank Rosenblatt: Principles of Neurodynamics: Perceptrons and the

Theory of Brain Mechanisms," *Spartan Books*, 1961.

Sáiz, P, F Llorens, L Blázquez, and F Cayón, "Base de Datos de Solicitudes de Patentes (España, 1878-1939)," 2008.

Sáiz, Patricio, "Base de Datos de Solicitudes de Privilegios. España 1826-1878," 2000.

Schauzu, Marianna, "The Concept of Substantial Equivalence in Safety Assessment of Foods Derived From Genetically Modified Organisms," *AgBiotechNet*, 2000, *2* (044), 1–4.

Schwartz, Victor E and Christopher E Appel, "Where's the Beef?: A Guide to Judges on Preemption of State Tort Litigation Involving Branded Drugs," *U. Cin. L. Rev.*, 2020, *89*, 597.

Shapiro, Adam Hale and Daniel Wilson, "Taking the Fed at its Word: Direct Estimation of Central Bank Objectives using Text Analytics," in "in" Federal Reserve Bank of San Francisco 2019.

_ , Moritz Sudhof, and Daniel Wilson, "Measuring News Sentiment," in "in" Federal Reserve Bank of San Francisco 2018.

Shavell, Steven, "The Judgment Proof Problem," *International Review of Law and Economics*, 1986, *6* (1), 45–58.

_ , *Liability for Harm Versus Regulation of Safety*, Routledge, 2018.

Skinner, Jonathan S, "The Costly Paradox of Health-Care Technology," *MIT Tech Rev http://www. technologyreview. com/news/518876/the-costly-paradox-of-healthcare-technology/. Published September*, 2013, *5*, 2013.

Slavtchev, Viktor and Simon Wiederhold, "Does the Technological Content of Government Demand Matter for Private R&D? Evidence from US States," *American Economic Journal: Macroeconomics*, 2016, *8* (2), 45–84.

Smith, Sheila, Joseph P Newhouse, and Mark S Freeland, "Income, Insurance, and Technology: Why Does Health Spending Outpace Economic Growth?," *Health Affairs*, 2009, *28* (5), 1276–1284.

Stein, Jeremy C, "Agency, Information and Corporate Investment," *Handbook of the Economics of Finance*, 2003, *1*, 111–165.

Stern, Ariel Dora, "Innovation under regulatory uncertainty: evidence from medical technology," *Journal of Public Economics*, 2017, *145*, 181–200.

Stewart, Conor, "Global Medical Devices Market 2021," Dec 2022.

Tabarrok, Alexander T, "Assessing the FDA via the Anomaly of Off-Label Drug Prescribing," *The Independent Review*, 2000, *5* (1), 25–53.

Turney, P. and P. Pantel, "From Frequency to Meaning: Vector Space Models of Semantics," *Journal of Artificial Intelligence Research*, 2010, *37*, 141–188.

US Government Accountability Office, "FDA Should Take Steps to Ensure That High-Risk Device Types Are Approved through the Most Stringent Premarket Review Process.," 2009.

Van Norman, Gail A, "Drugs and Devices: Comparison of European and U.S. Approval Processes," *JACC: Basic to Translational Science*, 2016, *1* (5), 399–412.

Watzinger, Martin and Monika Schnitzer, "Standing on the Shoulders of Science," *CEPR Discussion Paper No. DP13766*, 2019.

WIPO, "WIPO-Administered Treaties," 2020.

Wu, Lingfei, Dashun Wang, and James A Evans, "Large Teams Develop and Small Teams Disrupt Science and Technology," *Nature*, 2019, *566* (7744), 378–382.

Y Combinator, "Introducing Startup FDA," Nov 2016.

Yin, Wesley, "Market incentives and pharmaceutical innovation," *Journal of Health Economics*, 2008, *27* (4), 1060–1077.

Appendix: Chapter 1

A1 Bankruptcy Protection Model Extension

Following insights from the literature on the "judgment proof problem" (see Shavell (1986), Boomhower (2019)), when damages exceed the value of a firm's seizable assets, the difference can be discharged through bankruptcy. To reflect the bankruptcy option, I augment the model above to include expected damages that differ by firm assets K_f. I define the unspent capital available to cover damages as u_f. Unspent capital includes the capital not spent on commercialization costs ($K_f - c_f$) and profits from the current period, given by $u_f = \pi + K_f - c_f$. This term incorporates the simplifying assumption that net profits from the last period are distributed as devidends.[8] The upper bound of legal damages is given by $\bar{\phi}$.

Let ν represent the total realized damages from product defects, with probability distribution function $f(\nu; x_f^*, \vec{Z})$. In the presence of bankruptcy, the expected damages are given by

$$\text{Expected Damages} = \begin{cases} D(x_f^*; \vec{Z}) & \text{if } u \geq \bar{\phi}, \\ \underbrace{\left[\int_0^u \nu f(\nu; x_f^*, \vec{Z}) d\nu + \int_u^{\bar{\phi}} K f(\nu; x_f^*, \vec{Z}) d\nu\right]}_{D^T(x_f^*; \vec{Z})} & \text{else.} \end{cases} \tag{A1.1}$$

In words, if the firm's capital stock is at least as high as worst-case damages, the expected

[8] I could relax this assumption by letting u be equal to the unspent capital and the sum of all prior net profits up to a given point in time. This would mean that firms would tend to grow larger and eventually be unable to file for bankruptcy. However, the theoretical insights remain the same as initially smaller firms will face fewer expected damages for some time.

damages are the same as above, and the investment decision is unchanged. Smaller firms, however, confront a truncated damages distribution, where all possible damages outcomes higher than the firm's unspent capital stock u_f are fixed at u_f. Thus, instead of paying these outsized damages, the firm declares bankruptcy and contributes the value of its total assets to partially cover its damages. Hence, expected damages $D^T(x_f^*;\vec{Z})$ are determined by the probability-weighted sum of damages from 0 to u_f, plus the probability-weighted sum of u_f for all damages higher than u_f. Assume that the marginal benefit of safety effort for small firms is less than large firms at the same levels of safety effort, as there are fewer damages to abate (e.g., $-D'_T(x_f;\vec{Z}) < -D'(x_f;\vec{Z})$ for all x_f)

Bankruptcy protection changes the incentives to improve product safety for small firms. Deregulation introduces firms to legal damages; however, bankruptcy protects small firms from worst-case damages, lowering the marginal benefit of exerting safety effort. Thus, small firms exert less safety effort than large firms. I state this formally as follows:

Proposition 1 *(Deregulation introduces bankruptcy distortion) Assume firm A has less internal capital than (i) firm B (i.e., $K_A < K_B$) and (ii) its worst-case damages outcomes (i.e., $K_A < \bar{\phi}$). Firms A and B are otherwise identical. If deregulation leads to an increase in safety effort (see proposition 1 part ii), firm B will increase its safety efforts most (i.e., $x_B^* - \underline{x} > x_A^* - \underline{x}$). This occurs if and only if $x_B^* > x_A^*$ (which can stack with proposition 4 part ii, if capital is also below safety effort costs).*

A2 Proofs

A2.1 Proof of Proposition 1

Assume that $\psi + C_x(\psi\underline{x} - K) \; < - \; EL \cdot D'(\underline{x})$. Assume, by way of contradiction, that $x_f^* < \underline{x}$. Since x_f^* is the optimal safety effort, this implies that

$$\psi + C_x(\psi x_f^* - K) \; = - \; EL \cdot D'(x_f^*). \tag{A2.1}$$

However, since $x_f^* < \underline{x}$, we know that $C_x(\psi x_f^* - K) \leq C_x(\underline{x} - K)$ as costs are strictly increasing in x (given that $K \leq x$). We also know that $D'(x_f^*) < D'(\underline{x})$ as $D'(\ )$ is strictly increasing in x. Thus, $-EL \cdot D'(x_f^*) > -EL \cdot D'(\underline{x})$. Together, these inequalities imply that

$$\psi + C_x(\underline{x} - K) > -EL \cdot D'(\underline{x}). \tag{A2.2}$$

A contradiction. Thus, $x_f^* > \underline{x}$. See figure 1.2 for a graphical illustration of this proof.

A2.2 Proof of Proposition 2

Assume that deregulation leads to an increase in safety effort $x_A^* > \underline{x}$ and $x_B^* > \underline{x}$. I want to show that $x_A^* - \underline{x} < x_B^* - \underline{x}$. It suffices to show that $x_A^* < x_B^*$. Note that safety effort for deregulated firm B is chosen such that

$$\psi + C_x(\psi x_B^* - K_B) = -EL \cdot D'(x_B^*). \tag{A2.3}$$

And for firm A:

$$\psi + C_x(\psi x_A^* - K_A) = -EL \cdot D_T'(x_A^*). \tag{A2.4}$$

Since $D_T'(x) < D'(x)$ for all x, this means that

$$\psi + C_x(\psi x_A^* - K_A) < -EL \cdot D'(x_B^*). \tag{A2.5}$$

Assume, by way of contradiction, that $x_A^* > x_B^*$. This implies that $\psi + C_x(\psi x_A^* - K_A) > -EL \cdot D'(x_B^*)$, since $\psi + C_x(\psi x_A^* - K_A) > \psi + C_x(\psi x_B^* - K_B)$ as $C_x(\ )$ is strictly increasing in x and decreasing in K ($K_A < K_B$, which further strengthens the inequality if $K_A < \psi x_A^*$, or capital is less than safety effort costs). A contradiction. Thus $x_A^* < x_B^*$.

A2.3 Proof of Proposition 3

Note that, under regulation R, $T_A < T_B$, thus $t_{comm,a} > t_{comm,b}$; thus, for firm A, commercialization costs are strictly larger, financing costs are larger (if non-zero), and the effective life of the invention is shorter. Thus, the returns to commercialization are strictly lower for firm A.

Under the litigation environment L, there are no complexity distortions, thus the returns to commercialization are equal between firms A and B. We can formalize these insights as

$$Returns_{A,R} - Returns_{B,R} < 0 \text{ and } Returns_{A,L} - Returns_{B,L} = 0.$$

The difference in the change in the returns to commercialization from deregulation between firm A and B is given by:

$$DiD = (Returns_{A,L} - Returns_{A,R}) - (Returns_{B,L} - Returns_{B,R}). \tag{A2.6}$$

We WTS that this difference is positive or that the increase in returns is higher for firm A. Rewriting equation A2.6, gives:

$$DiD = (Returns_{A,L} - Returns_{B,L}) - (Returns_{A,R} - Returns_{B,R}). \tag{A2.7}$$

From part equation A2.3 we get

$$DiD = -(Returns_{A,R} - Returns_{B,R}) > 0. \tag{A2.8}$$

Thus, the increases in returns to commercialization are greatest at firm A.

A2.4 Proof of Proposition 4

Note that under the given conditions, small firms face lower expected damages and safety effort costs under deregulation than large firms (see proposition 1). Thus, deregulation would lead to larger returns from commercialization for smaller firms than larger firms, all else

equal. Therefore, showing that the returns from commercialization increase most for small firms through the financing channel is sufficient, given that bankruptcy distortions would broaden the conditions under which deregulation disproportionately benefits small firms. Hence, for simplicity, I consider only the financing channel and the conditions that guarantee outsized small-firm benefits.

Consider firm A's profit function with external funds $e_{R,A}$, given by:

$$REL \cdot \pi_R - \chi t_{comm} - \psi \underline{x} - C(e_{R,A}).$$

Note that firm A's external capital is positive (i.e., $e_{R,A} > 0$) since its internal capital is less than its non-financing commercialization costs (i.e., $K_A < c$); thus, due to nonzero capital frictions, its financing costs are positive (i.e., $C(e_{R,A}) > 0$).

Firm B's internal capital is greater than firm A's; thus, its external capital is less than firm A's, and its financing costs are less than firm A's. Firm A and firm B have identical profit functions aside from financing costs; thus, firm B's expected net profit is greater than that of firm A. Thus, either firm A's commercialization activity is the same as that of firm B ("non-marginal") or firm A's commercialization activity is less than firm B's.

Now for the litigation environment L, the returns to commercialization are given by:

$$Returns = EL \cdot [\pi_N - D(x_f^*; \vec{Z})] - \psi x_f^* - C(\psi = x^* - K_f). \tag{A2.9}$$

For a moment, think of x as not fixed. Since $K_A < K_B$, profits π, and EL are the same between the two firm types, at every value of x, the returns for firm A are strictly less than the returns for firm B, due to increased financing costs. If we assume bankruptcy, firm A also has lower expected damages than firm B and $x_A^* < x_B^*$, which would further increase the Assume, by way of contradiction, that exists an optimal safety effort for firm A x_A^* such that returns to firm A are larger than the returns to firm B at its maximum safety effort x_B^*. Since the returns to firm B are strictly larger than the returns to firm A at each value of x, there exists some x' such that

$Returns_B(x') > Returns_A(x^*)$. However, this implies that $Returns_B(x') > Returns_B(x_B^*)$, even though x_B^* is maximizes returns. A contradiction. Thus, firm A's returns are lower than firm B's. Further, commercialization activity is lower than firm B's. However, it could be the case that returns are negative in the litigation environment for both firms. If so, then commercialization is the same across both firms ("non-marginal").

Thus, we have

$$Returns_{A,L} - Returns_{B,L} < 0 \text{ and } Returns_{A,R} - Returns_{B,R} < 0. \tag{A2.10}$$

I want to also show that the sign of the following difference-in-differences is ambiguous: $(Returns_{A,L} - Returns_{A,R}) - (Returns_{B,L} - Returns_{B,R})$. We have that $(Returns_{A,L} - Returns_{A,R}) - (Returns_{B,L} - Returns_{B,R}) = (Returns_{A,L} - Returns_{B,L}) - (Returns_{A,R} - Returns_{B,R})$. We know this difference could be positive or negative. The first and second differences are both negative, thus the sign of the difference-in-differences depends on the relative changes in profits, damages, and delay costs. However, note that if capital is greater than optimal deregulated safety effort costs (i.e., $K_A \geq \psi x_A^*$), despite being lower than non-financing costs before deregulation, then $(Returns_{A,L} - Returns_{B,L}) = 0$ as there would be no financing costs to differentiate the returns of the two firms; thus, the change in returns would be larger for firm A. Note that if we also consider that damages for smaller firms are lower, due to bankruptcy, then $(Returns_{A,L} - Returns_{B,L}) > 0$. Thus, in both cases, the larger change in returns for firm A would translate into a larger increase in net profits if both firms A and B experience increases in net profits from deregulation.

A3 Learning Curve Estimation and Simulations

A3.1 Estimation Framework for the Learning Curve Parameters

Medical device manufacturers that are inexperienced with regulation may face additional costs when bringing a new medical device to market (Y Combinator 2016, Makower et al. 2010).

As presented in section 1.2, I model the additional costs from approval delays using a learning curve. I model the relationship between the approval delay of project N for firm f, $t_{comm,N,f}$ (measured in days), and cumulative experience, $\sum_{s=1}^{N-1} t_{comm,s,f}$, by the following equation:

$$t_{comm,N,f} = \beta(R_c)\left(\sum_{s=1}^{N-1} t_{comm,s,f}\right)^{-\gamma}, \text{ where } \gamma > 0.$$

Recall that $\beta(R_c)$ represents the baseline approval delay in medical device type c under regulation R (R can be Class III or II in practice), while $\sum_{s=1}^{N} t_{comm,s,f}$ represents the sum of approval delays (in days) faced after having submitted $N-1$ past projects.

More novel devices within a given medical device type may face longer approval delays if the FDA is more careful with these devices to ensure that new scientific characteristics do not lead to unexpected harm. However, the structure of Class III regulations helps distinguish between more or less novel innovation. As mentioned in section A5.3, firms that have already submitted an original PMA in a Class III medical device type may use PMA supplements for follow-on innovation within that device type. PMA supplements experience shorter approval delays and face fewer data requirements. On the other hand, the FDA requires original PMAs when firms have not yet submitted a PMA in a given Class III medical device type or when an incumbent firm invents a new device that is sufficiently novel. Thus, I include only approval delays that firms encountered when submitting original PMA documents in my analysis to condition on device novelty. This ensures that novelty is not driving approval delays.[9] For Class II devices, I ensure consistent novelty across devices by only considering documentation submissions for devices with unique brand names.

I log-linearize equation A3.1, to allow for OLS estimation of the parameter γ, and include medical device type and firm-level fixed effects, resulting in the following specification,

[9] I focus only on firms that have spent at least one day navigating FDA regulation to avoid potential confounders related to first-time innovators, including their tendency to "swing-for-the-fence" when confronted with barriers to entry (see Aghion et al. (2019)). This exclusion does not substantially change my results, with results remaining significant. I also perform the same empirical exercise for Class II device manufacturers as the sample size is much larger. For this exercise, I consider only 510(k) documents submitted for unique devices, finding significant, though smaller, results even after including product-code-by-year and firm fixed effects.

$$ln(t_{comm,N,f}) = ln(\beta(R_c)) - \gamma ln\left(\sum_{s=1}^{N-1} t_{comm,s,f}\right) + \alpha_c + \alpha_f + \varepsilon_{c,f}. \quad \text{(A3.1)}$$

For Class III devices, I include device type and firm fixed effects. For Class II devices, I include firm- and device type-by-year fixed effects, as I have enough observations within those more granular fixed effects to estimate the coefficients. Standard errors are clustered at the device-type-firm level. I exclude observations with no experience to avoid undefined outcomes in the estimation

The estimates of the learning curve parameters are significant for both Class III and II documentation submissions (see table A23).

A3.2 Simulation: Flattening the Learning Curve

As described in section 1.2, firm f's decision to innovate under regulation is determined by its return to commercialization

$$REL_f \cdot \pi_{R,f} - \chi t_{comm,f} - \psi \underline{x} - C(e_{R,f}), \quad \text{(A3.2)}$$

where $t_{comm,f} = \beta\left(\sum_{s=1}^{N-1} t_{comm,s,f}\right)^{-\gamma}$. For tractability, I assume that financing costs take the form $C(e) = \max(0, \chi_j t_{comm,f} + \psi \underline{x} - K_f)$. In addition, since I do not observe firm expenditures on safety R&D, the distribution of damages, safety efforts, or worst-case damages, I assume that damages and safety efforts are vanishingly small relative to profits and delay costs. This assumption is likely not innocuous as these costs are substantial, but it allows me to draw broader insights under my limitations by focusing on changes in delay costs that come from reducing regulatory complexity.

The learning curve parameters γ and $\beta(R_c)$ are presented in table A23 for Class III and Class II devices. I simulate the effect of flattening the learning curve on the rate of unique device inventions from Class III device manufacturers to assess the counterfactual of less complex FDA regulations. I calibrate χ to match the cost of approval delays found in Makower et al. (2010) at

the daily level for both Class III and II devices.

To execute these simulations, I first generate distributions of expected profits, firm sizes, and firm FDA regulatory experience. I proxy for expected discounted profits (i.e., $REL_f \cdot \pi_{R,f}$) using patent market valuations. This proxy requires the assumption that the market can adequately identify the expected discounted lifetime payout that a given patented innovation will yield to a firm and that this value is reflected in the change of the assignee's stock market price upon patent grant announcement. The device payout distribution is generated by fitting a gamma distribution to the medical device patent market valuations for Class III devices. I then fit a lognormal distribution to my firm size data to generate a distribution of asset values across firms. Lastly, I fit a gamma distribution to my firm FDA experience data.

After sampling from these fitted distributions to form a set of representative firms, I model how flattening the learning curve affects the rate of new device inventions across these firms. To this end, I anchor the right tail of the learning curve to the approval delay of the firm with the highest regulatory experience in my data and iteratively reduce the learning parameter (γ) while solving for a $\beta(R_c)$ value that allows the new curve to pass through the anchored value. I then calculate the firms' decisions to innovate, given the approval times corresponding to the new learning curve, and calculate the difference between the ex-post investment decisions (i.e., after the learning curve is flattened) and the ex-ante investment decisions (i.e., at the baseline values of γ and β). I then sum these differences across each firm and calculate the percentage change in new device inventions relative to the baseline values. Figure A9 shows the iterative flattening of the learning curve, and table A21 provides the calculations of the percentage change in new device inventions.

A4 Patent Data Collection

In this appendix section, I describe the process for collecting patents by device type in more detail. I also evaluate the accuracy of the procedure and demonstrate that my results are

robust to intuitive restrictions to the generated patent sample.

A4.1 Procedure for Gathering Patents by Device Type

The patent collection process begins by gathering a set of FDA device type descriptions for over 5,000 medical device types. These descriptions consist of both a broad FDA regulation number description and a narrower FDA device name description. To prepare these descriptions for keyword searches, I remove stop words, punctuation marks, and duplicate words. For example, the regulation number description "Implantable pacemaker pulse generator" and device type description "Leadless Pacemaker" would be transformed into the search string "implantable pacemaker pulse generator leadless." Next, I search the full text of the universe of US patent documents and gather all patents that contain all of the keywords in the search string. This process is repeated for all device types.

In some instances, patents are included in more than one device type. In such cases, I drop the patent from all but one randomly chosen device type.

A4.2 Examining the Accuracy of the Procedure

Naturally, keyword searches that link patents to device types can sometimes lead to false positive and false negative errors. For example, one of the most common inclusion errors I encountered was when keyword searches mistakenly linked drug-related patents to medical device types, according to the Cooperative Patent Classification (CPC) system. However, these discrepancies between the CPC classifications and my linkages may not always be erroneous, as some drug technologies may be complementary to certain device types. Therefore, using keyword searches instead of the CPC system can be useful for capturing complementary technologies, but using both can provide a way to validate my data. Below, I present a few examples of patents I identified through random sampling of drug-related patents, which may or may not be inclusion errors.

First, the patent "US-10428030-B2" describes a compound that can be used as a diag-

nostic tool in combination with Nuclear Magnetic Resonance Imaging (NMRI). According to the Cooperative Patent Classification (CPC) system, this compound is classified as a drug rather than a medical device. However, when I searched patent texts using the medical device type keywords "nuclear magnetic resonance imaging diagnostic systems," the patent was included in my results. Even though the compound itself is not a device, it may be possible that innovation in these types of compounds increases when NMRI diagnostic systems (complementary technologies) are deregulated. The patent "US-10314846-B2" is another example of this technological complementarity. My keyword search technique includes these complementary technologies while relying on patent classifications alone would not, as the compound is labeled as a drug (i.e., A61P25/14–Drugs for disorders of the nervous system for treating abnormal movements, e.g., chorea, dyskinesia).

Another example of the benefits of using keyword searches is demonstrated when searching for patent documents containing the keywords "cyclosporine test system." In this case, the patent "US-10011612-B2" is included in the results. According to the Cooperative Patent Classification (CPC) system, this patent is classified as a drug (i.e., A61P1/16–Drugs for disorders of the alimentary tract or the digestive system for liver or gallbladder disorders, such as hepatoprotective agents, cholagogues, and lithophytic). As described in the patent, the drug is administered in combination with other agents, such as an anti-inflammatory drug, antimicrobial agent, anti-angiogenesis agent, immunosuppressant, antibody, steroid, an ocular antihypertensive drug, or a combination of these agents. Examples of these agents include cyclosporine. The administration of such drugs is typically monitored using cyclosporine tests to ensure that appropriate levels of the drug are in a patient's system. Therefore, it is plausible that increased innovation in and cheaper acquisition of cyclosporine test systems could lead to increases in innovation in cyclosporine immunosuppressants.

However, this type of sensitivity in keyword searches can also result in inclusion errors. For example, when I searched patent texts for the device type "soft contact lens daily wear," I included a patent for a drug that treats corneal ulcers (eye ulcers). This patent was included in my

results because it mentions that the drug can be administered as a contact lens or reservoir, among other methods. While there may be some technological complementarities between contact lenses and this type of drug, the connection is weaker. Nonetheless, this example demonstrates how keyword searches can sometimes include patents that may seem only tangentially related.

Although there may be valid reasons to include drug-related technologies and other non-medical-device technologies in my patent data, I also demonstrate that my results are not sensitive to restricting my patent data only to medical devices in the following section.

A4.3 Robustness of Procedure

To validate the results of my main specification that analyzes patent data collected using keyword searches, I use the CPC system to restrict my patent sample to only include medical devices and find that my results are robust. To restrict the sample, I only keep collected patents that fall under the "Medical or Veterinary Science Hygiene" CPC categories (i.e., include "A61"), but that exclude patents classified as drugs (i.e., not "A61P"). This restriction reduces the number of included patents from 1,248,289 to 239,315 patents. In the CSV file linked here, I provide the top three CPC labels for patents collected in each device type for all affected Class III devices used in my analysis. In another CSV file linked here, I provide the top three CPC labels for patents collected in each device type for all affected Class II devices used in my analysis. Notice that the descriptions of most top CPC codes correspond with the descriptions of medical device types.

Table A24 presents the estimates of equation 1.4.1 using the restricted patent sample for my patenting rate outcomes. The table reveals that the estimates remain large in magnitude and statistically significant. In fact, the percentage change in patenting rates relative to pre-event means is larger for both Class III to II and Class II to I events. However, the magnitude of the effects is reduced by approximately one third, signifying that approximately one-third of the effect on patenting in my main specification may be due to positive spillovers into complementary technologies. Figure A15 shows the estimates from an event-study analysis and suggests that the

results from my main specification are robust when using this restricted sample of patents.

Lastly, my estimates for the outcome defined as the number of new FDA device submissions (i.e., the "Device Submission Rate") also support my patenting results by showing similar increases in innovation.

A5 Additional Details

A5.1 FDA Decision Rule for Class II to I Events

All Class II to I down-classifications were determined using a "device priority score." These scores were calculated using the following linear combination of evaluation factors,

$$\text{DPM} = 0.38\text{D} + 0.3\text{S} + 0.12\text{LS} + .08\text{U} + .08\text{B} + 0.04\text{E}. \tag{A5.1}$$

In the model, D is the frequency of death, S is the frequency of serious injury, LS is the frequency of less serious injury, U is the frequency of use, B is the health benefit, and E is effectiveness. The FDA calculated the adverse event evaluation factor scores D, S, and LS with the following rule,

$$\text{Y} = \begin{cases} 100 \text{ if in "high" range,} \\ 50 \text{ if in "medium" range,} \\ 0 \text{ if in "low" range.} \end{cases} \tag{A5.2}$$

The FDA pre-determined the three different ranges and their respective cutoffs, given annual counts of the outcome Y. The evaluation factor scores for U, B, and E are given by

$$\text{Y} = \begin{cases} 0 \text{ if in "high" range,} \\ 50 \text{ if in "medium" range,} \\ 100 \text{ if in "low" range.} \end{cases} \tag{A5.3}$$

Intuitively, this means that given two devices with the same annual incidence of deaths and injuries, the device with the highest DPM score is the device that has the highest intrinsic risk per use, the lowest health benefit, and the least effectiveness. The FDA uses the resulting DPM score to flag marginal devices on the edge of their decision rule (see FDA (1995)). Other conditions for down-classification are uniformly satisfied across all down-classified types and would not affect the marginal decision.

I do not observe the pre-determined thresholds for D, S, and LS, and I do not observe B, U, and E. I proxy for the decision rule by taking a linear combination of the average yearly counts of deaths (D), serious events (S), and less-serious events (LS). This calculation is given by

$$\text{DPM} = 0.38\text{D} + 0.3\text{S} + 0.12\text{LS}. \tag{A5.4}$$

I then compare the DID estimates from the treated device types in the top decile of calculated DPM scores against treated device types from the 0–90th percentile. In practice, U, B, and E would not influence the ordering of calculated DPM scores as the average DPM score of the top decile of medical device types is four times higher than the average DPM value of the device type at the 89th percentile. Additionally, device types with a high D evaluation factor also tend to have high S and LS evaluation factors; Thus, the stepwise construction of D, S, and LS in the FDA's decision rule would not substantially affect ordering.

A5.2 FDA Decision Rule for Class III to II Events

Class III to II events are much less mechanical. When considering down-classifying a Class III device, the FDA analyzes the health risks of the device and whether Class II regulations will reasonably mitigate those risks. It makes these assessments by consulting the medical literature, internal data (i.e., premarket approval applications, equipment problems in the past resulting in recalls and adverse events), and clinical experiences with the device.

An illustrative example of a Class III to II event is the down-classification of daily-wear

soft contact lenses in 1994. In the minutes of the 1994 ophthalmic panel meeting in which the FDA announced this event, the FDA cites safety information contained in submitted PMAs as the reason for deregulation. However, the timing of this event is "as good as random." In this same document, the FDA cites that it had been "dealing with [the down-classification event] for about ten years" and that because "the data that were needed to support reclassification were contained in PMAs and were not publicly available," they could not act. Thus, bureaucratic hurdles make these policies difficult to predict, making the timing of the events unlikely to be correlated with changes in outcomes beyond the effects of deregulation. Upon reclassification, the number of unique daily-wear soft contact lenses submitted for approval rose sharply, as the number of new extended-wear contact lenses, which remained in Class III, remained steady (see figure A14).[10]

A5.3 Class I, II, and III Medical Device Regulations

Manufacturers of Class I devices (those perceived as low-risk) must simply abide by a standard set of safe marketing practices called "general controls."[11] A newly marketed medical device can be categorized as Class I if it is reasonably similar (i.e., same intended use and broad characteristics) to another device categorized as Class I. However, if a new medical device has distinct characteristics or intended use, the new device is given a new class III product code.[12]

Manufacturers of Class II devices are required to follow specific guidelines, called special controls, designed to mitigate device-specific risk and submit a 510(k) document, or "premarket

[10]Note that because I cannot observe the safety variables that drive Class III to II events, it is difficult for me to extrapolate the product safety results I find in these events to other Class III devices that were not down-classified. Because I do not observe these variables, I do not know what the "marginal" device type would be; thus, I cannot determine whether the average effects differ from the marginal effects.

[11]These devices are "low-risk" as they do not support or sustain human life and do not pose a potential unreasonable risk of illness or injury (e.g., a tongue depressor). 41% of all medical device types, or "product codes," fall under Class I. Of these, 90% are exempt from filing any documentation (aside from facility registration with the FDA).

[12]The FDA can then evaluate the safety and efficacy of new product codes and reclassify them, or a device manufacturer can submit a "De Novo" petition for the formal classification of a new device type. A new device can be classified as Class I or II if "the device has existing or reasonably foreseeable characteristics of commercially distributed devices within that generic type or...[The device requires a 510(k) (even if its generic type is Class I) if] the device is intended for a use different from the intended use of a legally marketed device in that generic type of device...[or if] the modified device operates using a different fundamental scientific technology" (FDA 2020a).

notification."[13] Through the 510(k) process, a manufacturer must demonstrate that their device is "substantially equivalent" to a previously marketed device for which a "premarket approval" (PMA) is not required. A device is substantially equivalent if it has the same intended use and technological characteristics as the predicate device. The 510(k) path is shorter and less costly than the more intensive PMA process described below. However, the 510(k) process can be expensive, with an average cost of $24 million (Makower et al. 2010). If the FDA finds that a device is not sufficiently similar to a predicate device, the manufacturer must file a PMA, which carries the most stringent requirements.

Manufacturers of Class III devices must perform clinical trials through the PMA process to ensure their new device is safe and effective before commercialization.[14] Class III device types are perceived as high-risk since not enough information exists to establish special controls that ensure safety and effectiveness (i.e., new device types) or if special controls do not adequately mitigate device risk.[15] The PMA process takes much longer than the 510(k) process, and costs, on average, $75 million (Makower et al. 2010). After a manufacturer has submitted a PMA document for their device, any small changes to their device that affect the device's safety or effectiveness require a PMA supplement submission. PMA supplements often do not require premarket clinical data and experience shorter review timelines (Johnson 2012).[16]

[13] 56% of medical device product codes fall under this category.

[14] Pre-amendment class III devices (those existing before 1976) only have to submit a 510(k) if the FDA has not issued a final order requiring PMA submission (Center for Devices and Radiological Health 2018). A small percentage of 510(k)s also require a small amount of clinical data to support marketing clearance by the FDA.

[15] Roughly 2% of product codes currently fall under this classification, although these product codes represent an outsized portion of U.S. medical device spending (Meier 2009).

[16] However, the requirements associated with PMA supplements are dependent on the degree to which the new device has changed, with small changes (like labeling changes) requiring no fee and design changes requiring preclinical testing. Most submitted class III documentation is from PMA supplements.

A6 Supplemental Figures and Tables

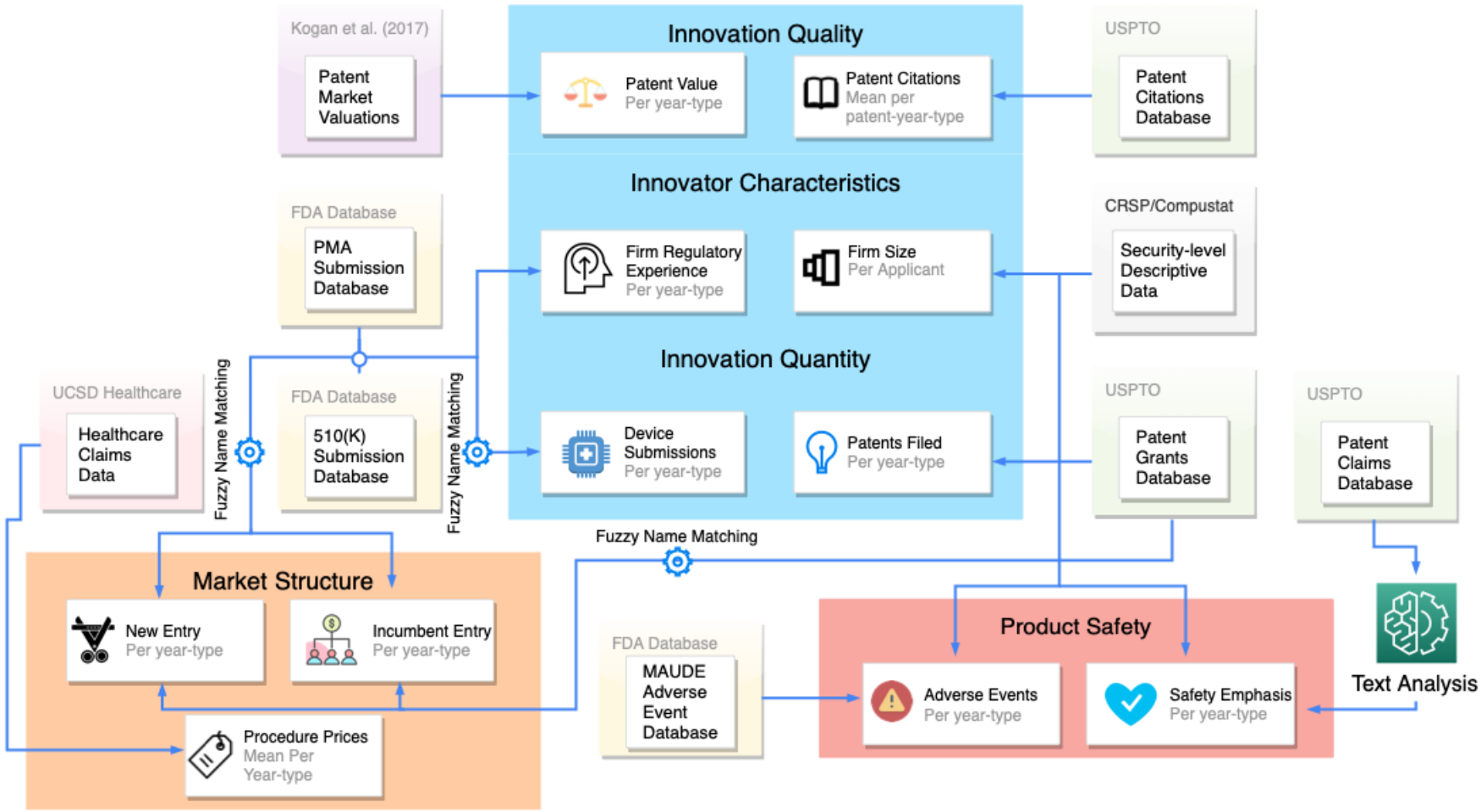


Figure A1. Data Catalog. Note: This figure presents a catalog of the various data sources used in this study. The three broad outcomes are represented by the three colored boxes: blue innovation, orange market structure, and red product safety. Each broad outcome contains various specific outcomes measured, in most cases, by two different data sources. Buttons on the exterior represent data sources. The blue arrows connect the data sources to outcome measures. The cogs indicate when algorithms were used to process the data into an outcome measure. The green "Text Analysis" cog represents the word2vec algorithm used to extract safety-related keywords from patent claims data.

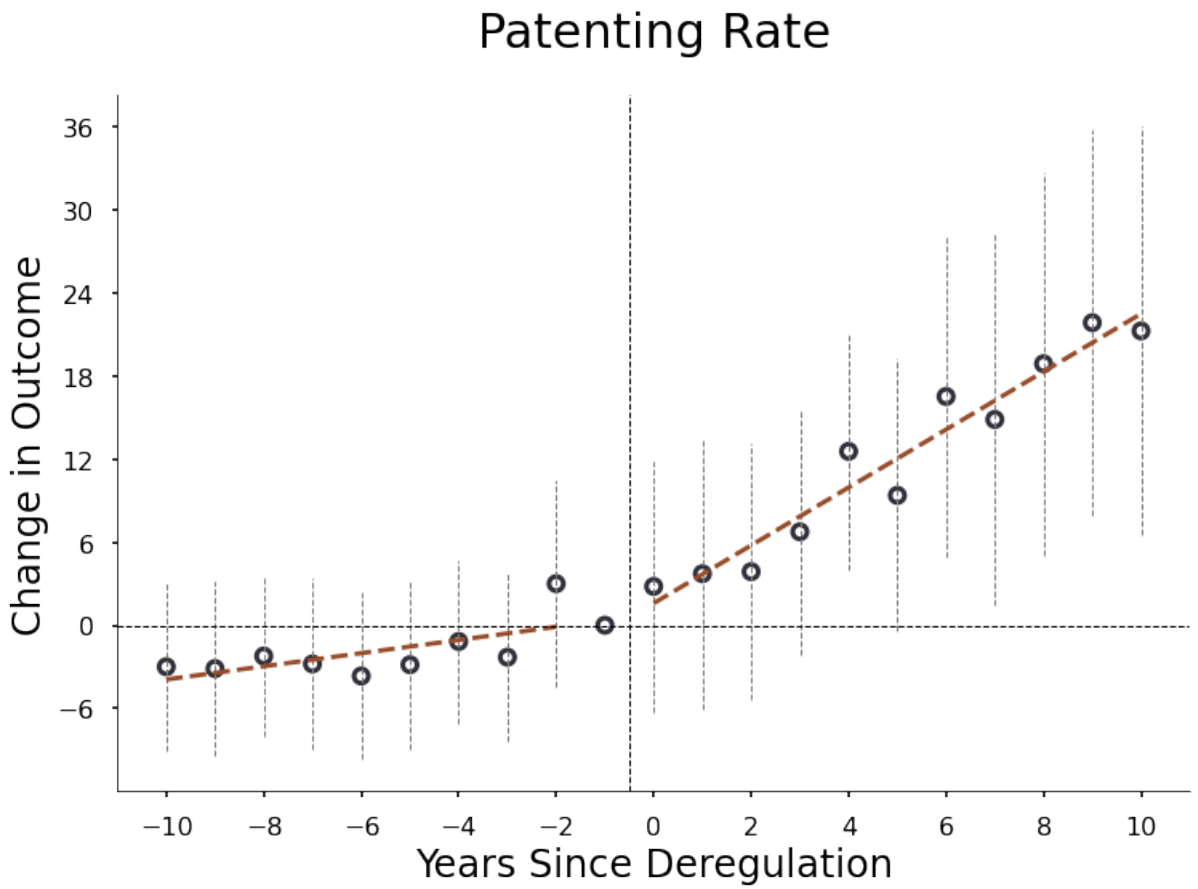


Figure A2. Petitioned Down-Classification Events (Not FDA-Initated). Note: This figure presents the estimates of the coefficients from event-study equation 1.4.2 for the patent filing rate measure and illustrates the potential biases that stem from industry petition of down-classification. Outcome data are derived from USPTO patent data. Only Class III to II down-classification events petitioned by industry (not by the FDA's own initiative) are considered. Controls are device types matched on baseline averages of the outcome. Data are analyzed at an annual frequency. 95% confidence intervals are calculated following Conley and Taber (2011).

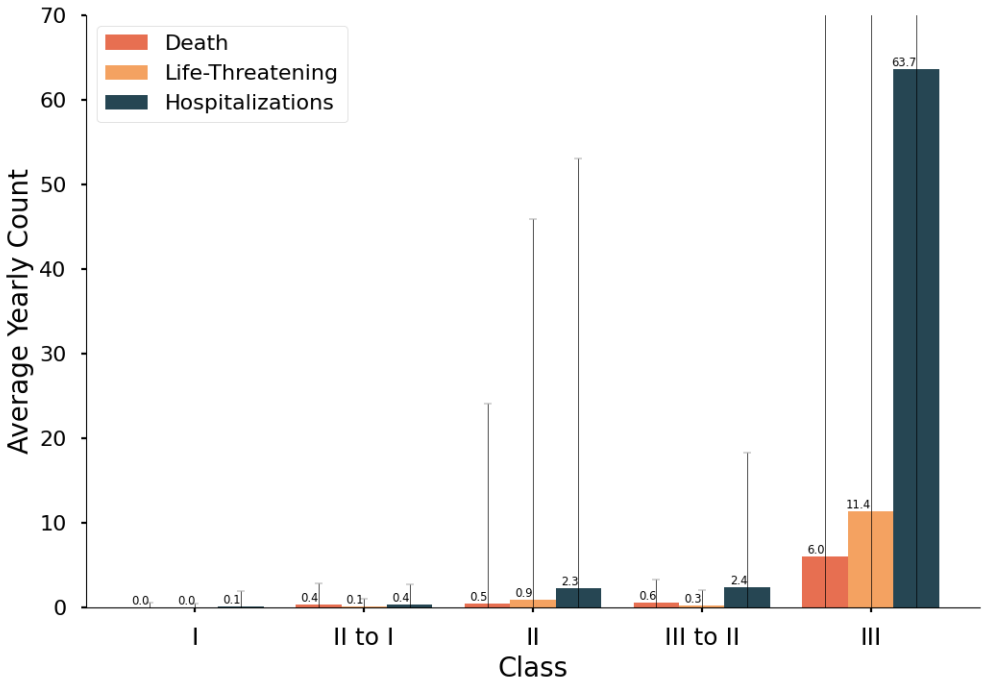


Figure A3. Mean Yearly Adverse Event Counts by Device Type Class. Note: This figure presents the annualized average counts of the specified adverse events for medical device types within the respective classification. The x-axis indicates the device type Class. The x-axis includes down-classified devices from Class III to II and Class II to I events separately. The y-axis details the average annualized count for a given class and adverse event type. The red bar represents the average number of yearly deaths across device types and years. The orange bar calculates a similar average for life-threatening events, and the blue bar calculates the average number of hospitalizations. These three variables are derived from the FDA MAUDE adverse event data. Standard error bands also overlay the average estimates.

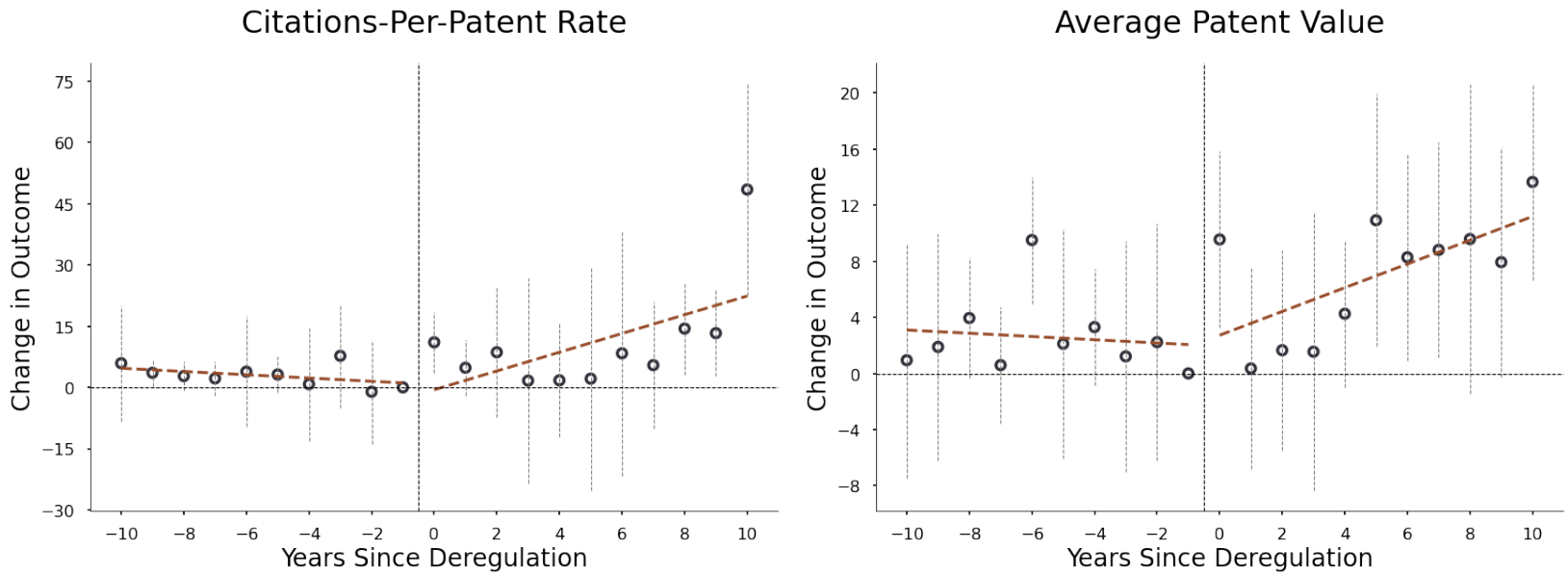


Figure A4. Innovation Quality Event Study Class III to II. Note: This figure presents the estimates of the coefficients from the event-study equation 1.4.2 for the innovation quality outcomes. Only Class III to II down-classification events are considered. Controls are device types matched on baseline averages of the outcome. Data are analyzed at an annual frequency. The left subfigure describes the evolution of the average citations-per-patent rate. When no patents are filed in a given year, the citations-per-patent rate is set to zero. The right subfigure presents the evolution of the average patent value in treated device types relative to controls. Patent values are derived from Kogan et al. (2017), who calculate the change in a firm's stock market valuation upon patent grant announcements to measure patent value. Standard errors are calculated following Conley and Taber (2011).

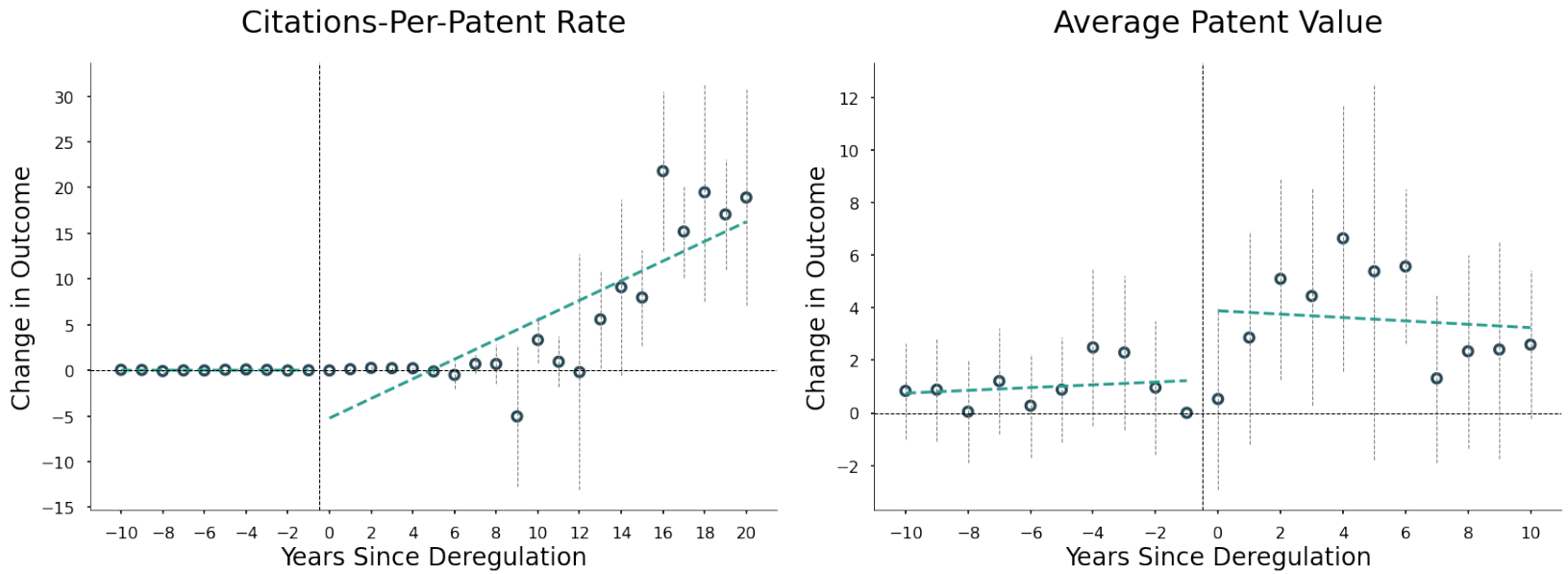


Figure A5. Innovation Quality Event Study Class II to I. Note: This figure presents the estimates of the coefficients from event-study equation 1.4.2 for my innovation quality measures. Only Class II to I down-classification events are considered. Controls are device types matched on baseline averages of the outcome. Data are analyzed at an annual frequency. The left subfigure describes the evolution of the average citations-per-patent rate. When no patents are filed in a given year, the citations-per-patent rate is set to zero. The right subfigure presents the evolution of the average patent value in treated device types relative to controls. Patent values are derived from Kogan et al. (2017), who calculate the change in a firm's stock market valuation upon patent grant announcements to measure patent value.

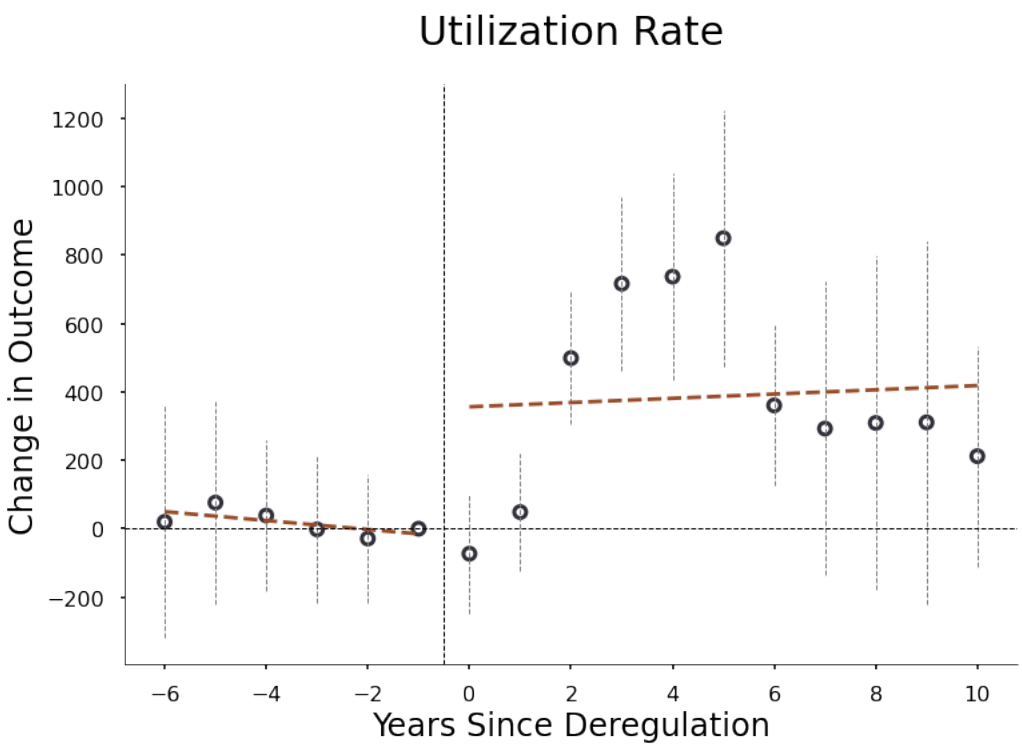


Figure A6. Utilization Rates Event Study. Note: This figure presents the estimates of the coefficients from event-study equation 1.4.2 for the utilization rates of procedures that use treated or control medical device types. I do not have claims data before 2005; Thus, I only consider post-2005 Class III to II down-classification events. Controls are device types matched on baseline average innovation rates. Data are analyzed at an annual frequency. Utilization is measured by the yearly number of paid claims for a given procedure. Claims data come from the UCSD healthcare system. Conley–Taber 95% confidence intervals are provided.

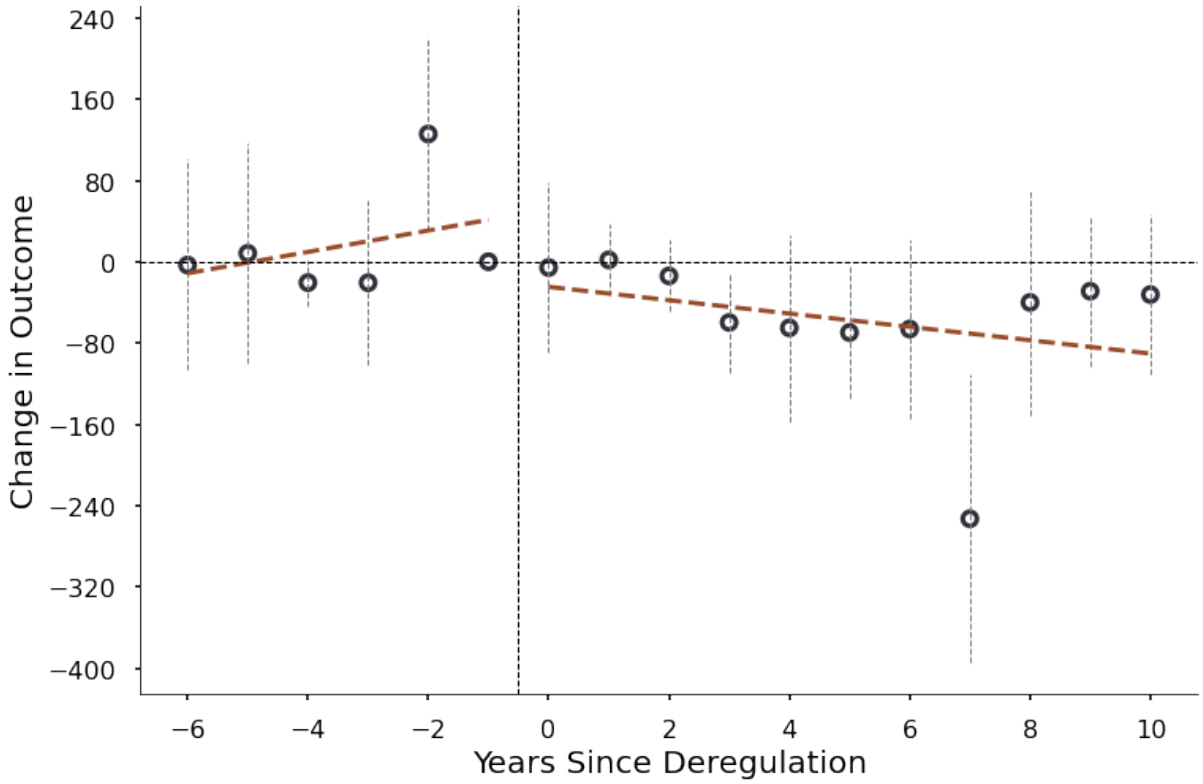


Figure A7. Procedure Price Event Study Class III to II. Note: This figure presents the estimates of the coefficients from event-study equation 1.4.2 for the price component of my market structure measures. Controls are device types matched on baseline outcome averages. I do not have UCSDH claims data before 2005; Thus, I only consider post-2005 Class III to II down-classification events. Data are analyzed at an annual frequency. The price is determined by the amount insurers paid for a given procedure. The figure describes the evolution of the prices of procedures that use treated device types relative to control groups matched using pre-event price averages. Conley–Taber 95% confidence intervals are provided.

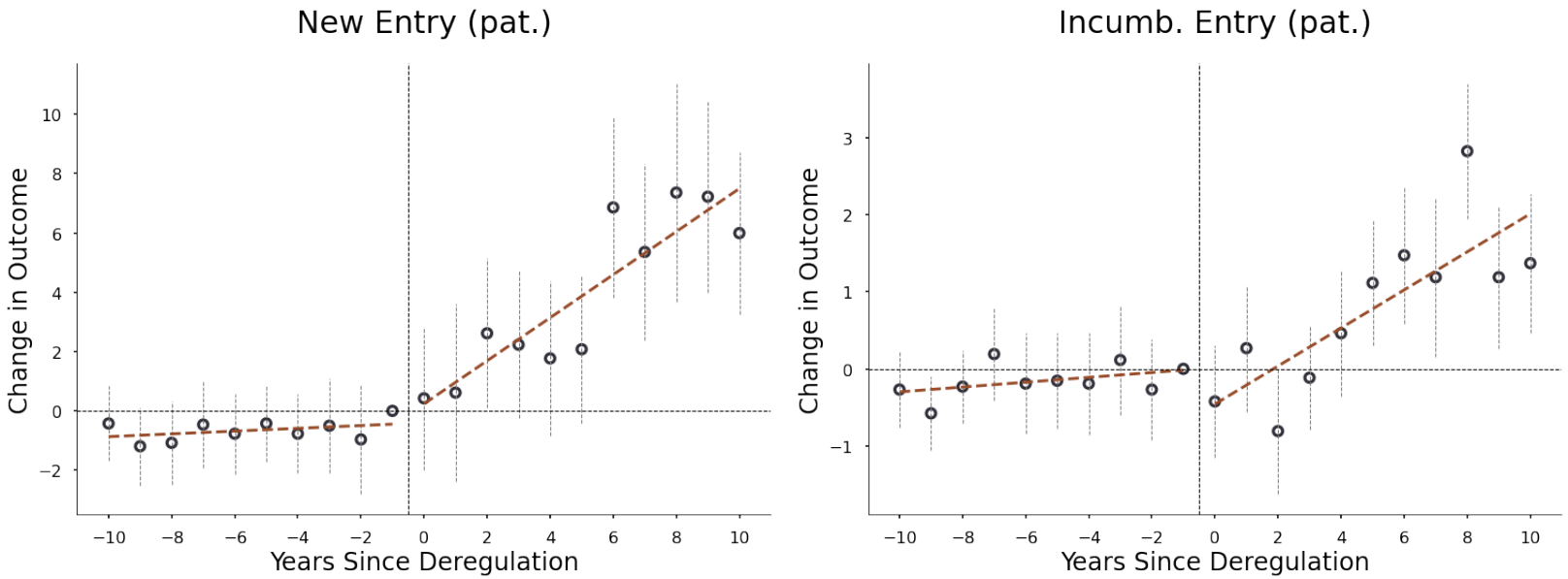


Figure A8. Market Structure Event Study Class III to II (Patent Measures). Note: This figure presents the estimates of the coefficients from event-study equation 1.4.2 for my market structure measures. Only Class III to II down-classification events are considered. Data are analyzed at an annual frequency. Controls are device types matched on baseline average innovation rates. The left subfigure describes the evolution of new entry of firms that have never before received a granted patent (counts per year), measured by patent data. The right subfigure presents the evolution of incumbent entry into treated device types relative to controls, measured by patent data. Conley–Taber 95% confidence intervals are provided.

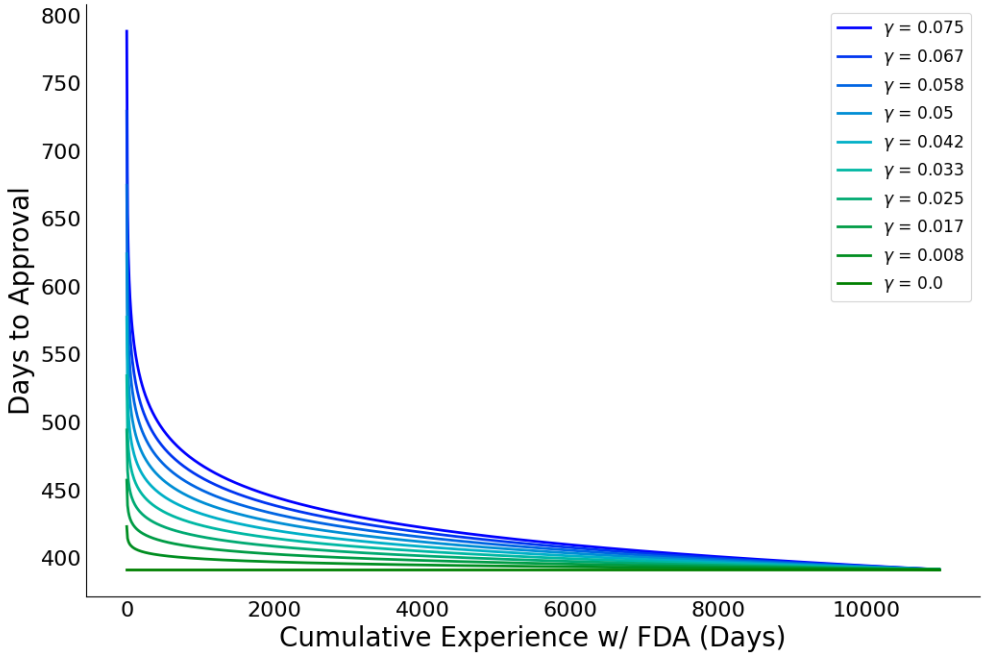


Figure A9. Flattening the Learning Curve Simulation. Note: This figure presents the simulation exercise of flattening the Class III learning curve estimated in equation A3.1. I flatten the learning curve relative to the most experienced firm. The results of this simulation are provided in table A21. Above, gamma begins at its initial starting point estimated in equation A3.1. Subsequent lines show the change in the learning curve as gamma is reduced while maintaining the approval time of the top quartile of experienced firms.

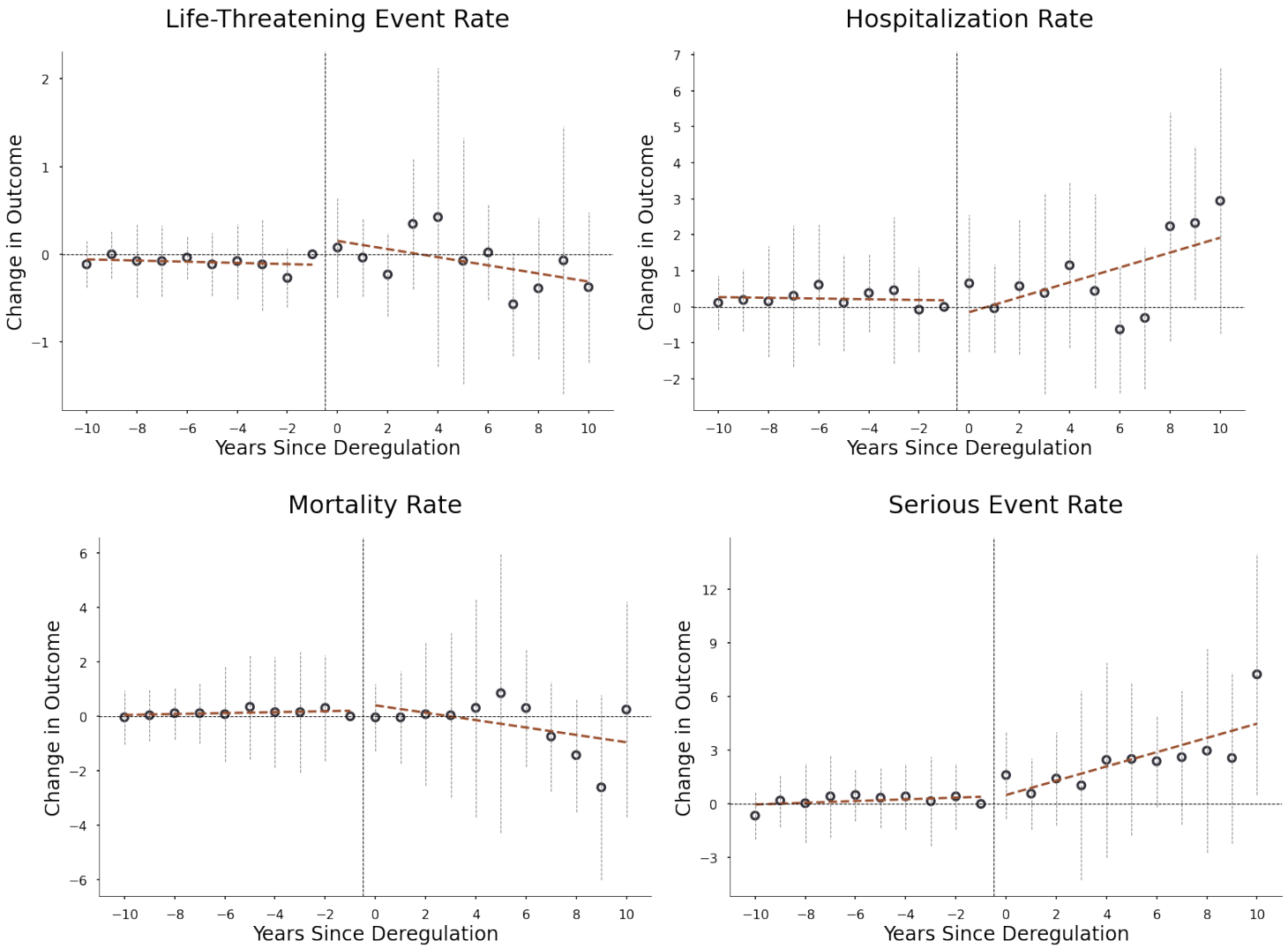


Figure A10. Adverse Event Event Study Class III to II. Note: This figure presents the estimates of the coefficients from event-study equation 1.4.2 for my adverse event measures. Only Class III to II down-classification events are considered. Data are analyzed at an annual frequency. Controls are device types matched on baseline outcome averages. The top-left subfigure describes the evolution of the rate of life-threatening events stemming from the use of treated device types relative to control groups matched using baseline averages. The top-right subfigure describes the evolution of the rate of hospitalizations of treated device types relative to control groups. The bottom-left subfigure describes the evolution of the death rate. The bottom-right subfigure presents the evolution of the sum of all serious adverse events (life-threatening, death, hospitalizations, and disability) in treated device types relative to controls. Adverse events are derived from the FDA MAUDE database. Conley–Taber 95% confidence intervals are provided.

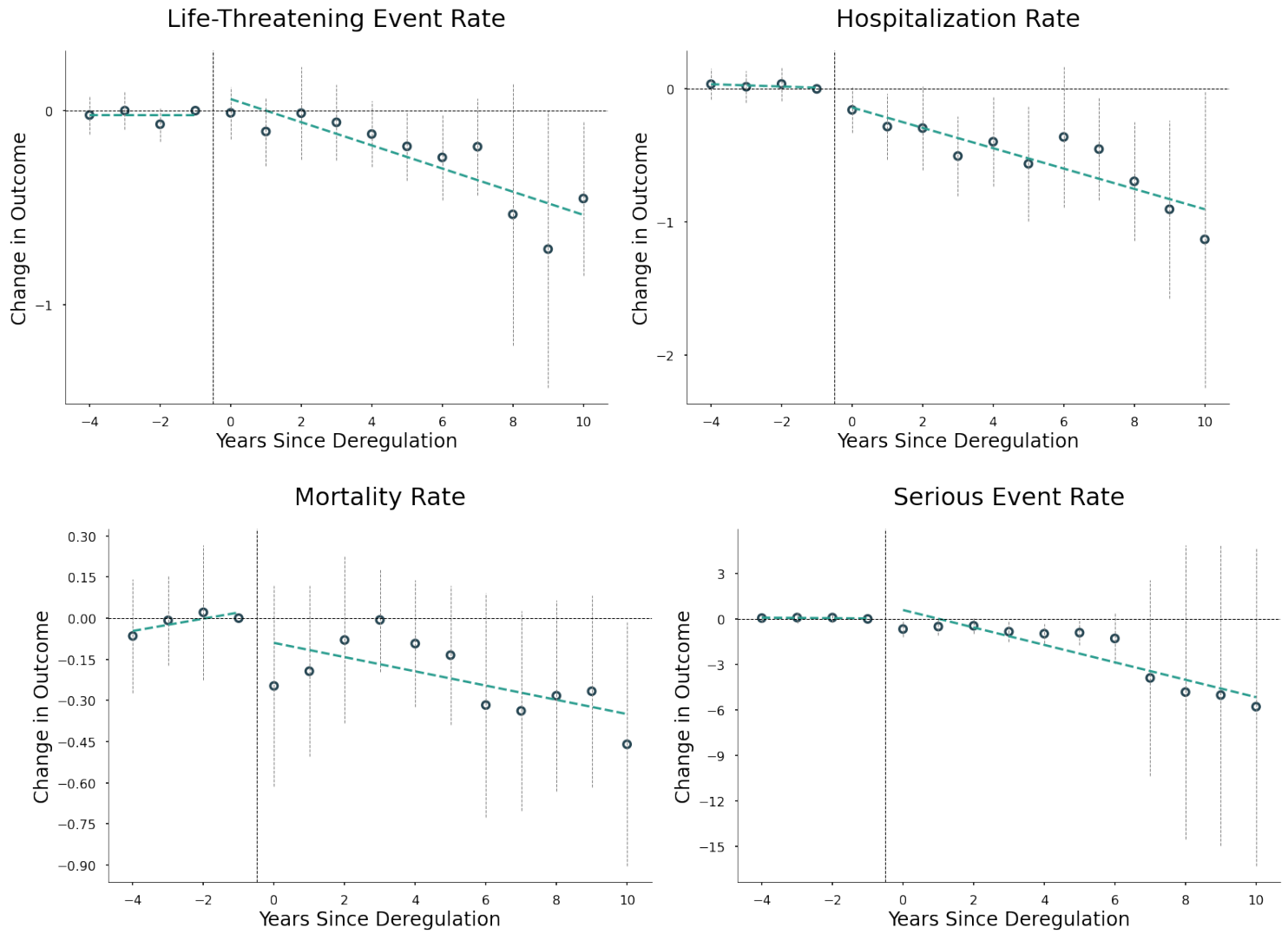


Figure A11. Adverse Event Event Study Class II to I. Note: This figure presents the estimates of the coefficients from event-study equation 1.4.2 for my adverse event measures. Only Class II to I down-classification events are considered. Only four pre-periods are included because there are no prior adverse event data. Data are analyzed at an annual frequency. Controls are device types matched on baseline outcome averages. The top-left subfigure describes the evolution of the rate of life-threatening events stemming from the use of treated device types relative to control groups matched using baseline averages. The top-right subfigure illustrates the evolution of the rate of hospitalizations of treated device types relative to matched control groups. The bottom-left subfigure describes the relative evolution of the death rate. The bottom-right subfigure presents the relative evolution of the sum of all serious adverse events (life-threatening, death, hospitalizations, and disability) in treated device types. Adverse events are derived from the FDA MAUDE database. 95% confidence intervals are provided.

Figure A12. Safety Emphasis Event Study Class II to I. Note: This figure presents the estimates of the coefficients from event-study equation 1.4.2 for inventors' emphases on safety. Only Class II to I down-classification events are estimated. Data are analyzed at an annual frequency. Controls are device types matched on baseline outcome averages. The figure describes the evolution of the proportion of patents that emphasize safety within patent texts. The volatility in the four years prior to the down-classification represents the congressional whiplash that occurred regarding whether to abolish the FDA. 95% confidence intervals are provided.

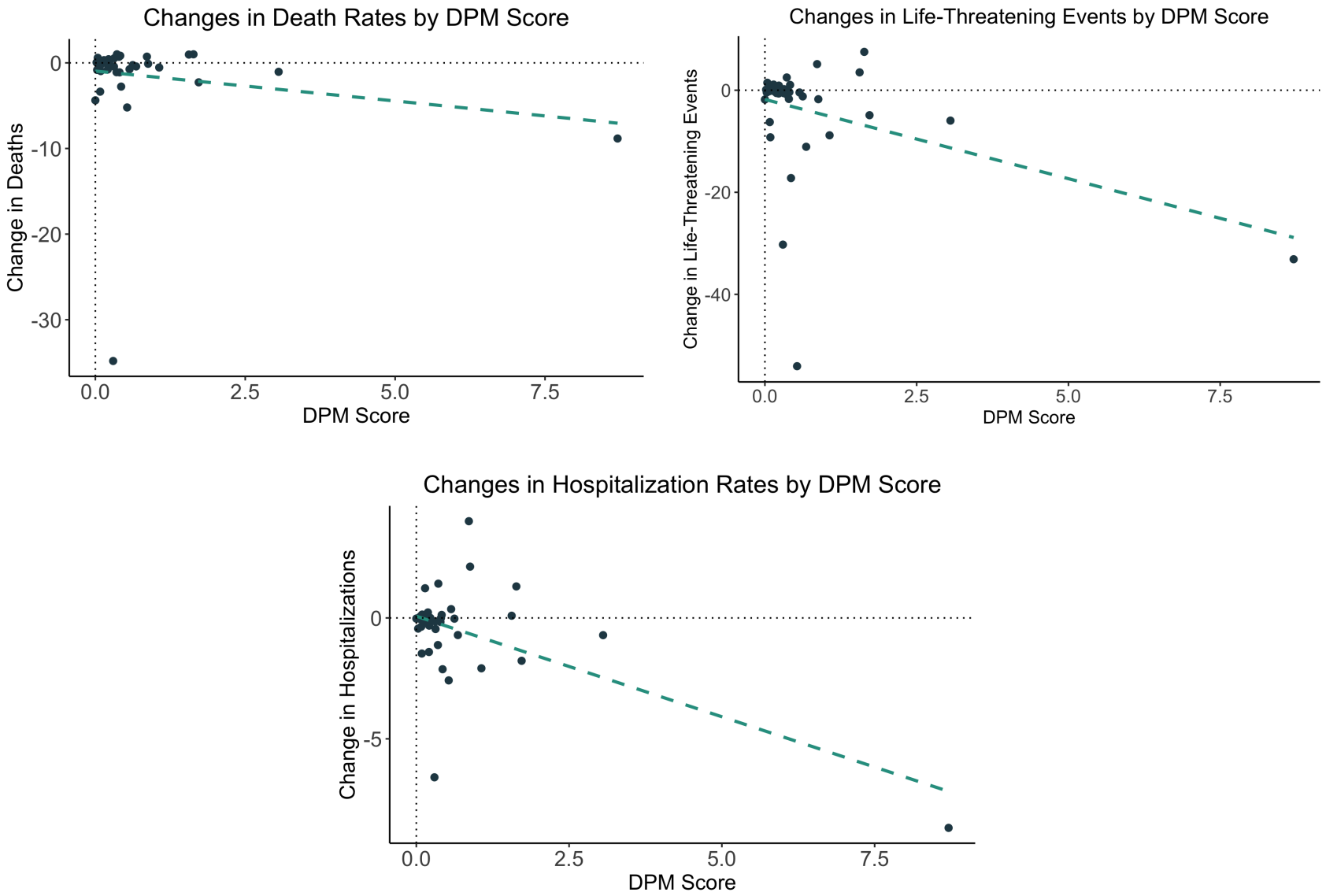


Figure A13. Class II to I Changes in Adverse Event Rates at Margin of Decision Rule. Note: This figure presents separate DID estimates of equation 1.4.1 for each adverse event measure and each treated device type with a given proxy DPM score relative to matched controls. The DPM score is primarily an increasing function of the baseline average annual incidence of adverse events before deregulation (see appendix A5.1). When the rightmost outlier is removed, the slopes of the fitted lines are still negative or zero. Controls for each treated device type are selected by matching based on DPM scores across both Class I and II devices that were not down-classified in the given period. The x-axes describe the same proxy DPM score across the three adverse event outcomes. The y-axes describe the change in the rate of the given adverse event type in the treated device type relative to matched control device type. The top-left figure shows the differences-in-differences estimates for the change in death rates across device types, the top-right figure shows the same for life-threatening events, and the bottom figure shows the same for hospitalizations. Adverse event data are from the FDA's MAUDE database. 95% confidence intervals are provided.

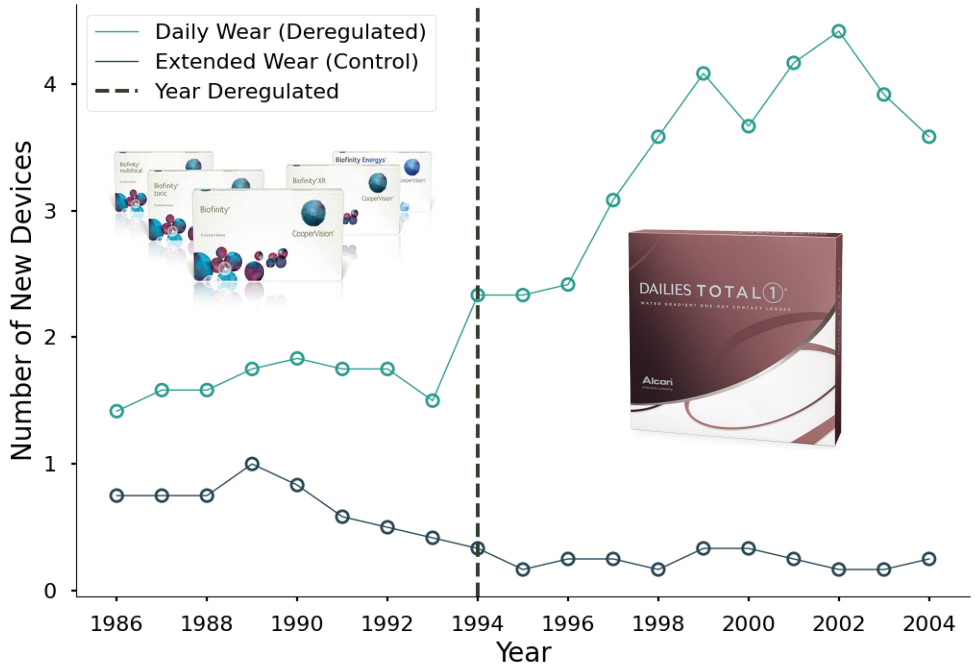


Figure A14. Contact Lens Use Case—III to II Down-Classification. Note: This figure presents an example of a Class III to II down-classification event. In 1994, the FDA down-classified daily-wear soft contact lenses to Class II but kept extended-wear soft contact lenses in Class III. The x-axis measures the year, and the y-axis measures the number of unique contact lens devices submitted to the FDA for approval in a given year. The green line represents daily-wear contact lenses submitted for approval (deregulated), and the blue line represents extended-wear soft contact lenses submitted for approval (remained in Class III). The vertical black line represents the year of reclassification. The left-imposed picture shows an example of a soft contact lens invented before reclassification. The right-imposed picture shows an example of a soft contact lens invented after reclassification.

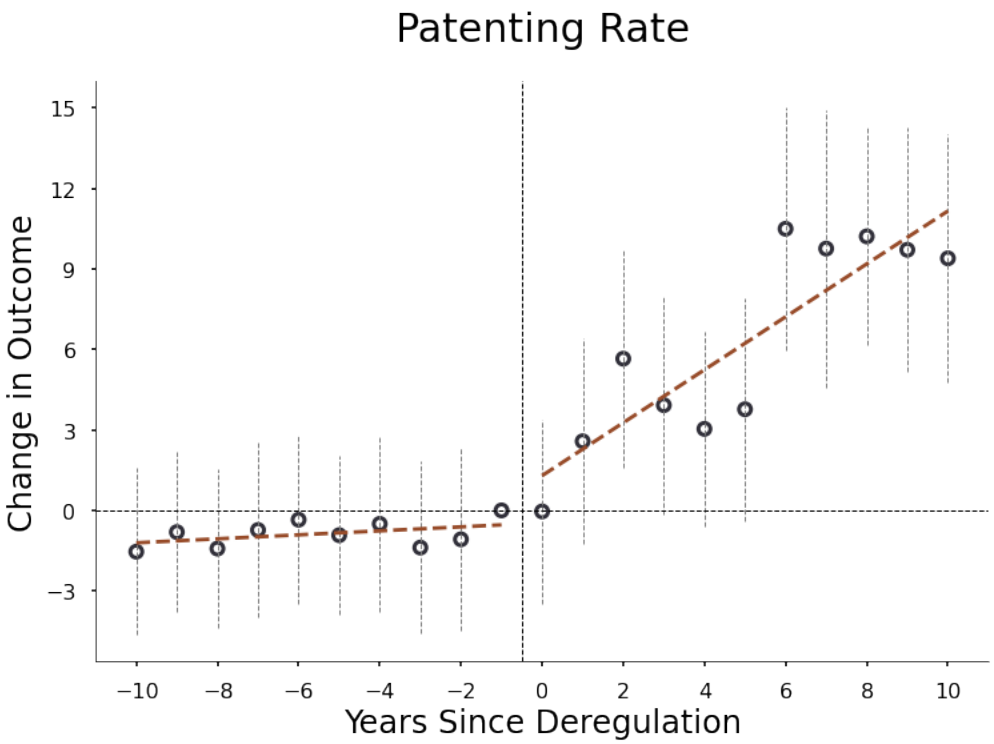


Figure A15. Effects of Class III to II Events on Patenting Rates: Restricted Patent Sample. Note: This figure presents the estimates of the coefficients from event-study equation 1.4.2 for patenting rates using the restricted patent sample described in appendix A4. Compare to the top-left subfigure of figure 1.3. Controls are device types matched on baseline average innovation rates. Data are analyzed at an annual frequency. The patenting rate is measured by the yearly number of patents filed in a given device type. Patent data comes from the USPTO patent database. Conley–Taber 95% confidence intervals are provided.

Table A7. Summary Statistics – Class I. Note: This table presents summary statistics only for Class I devices. See Kogan et al. (2017) for more information on the patent market valuation data, which was merged into my patent dataset. The CRSP/Compustat database was used to derive the total assets of the firms applying for patent protection and is a proxy for firm size. Market values and applicant assets are only available for patents filed by publicly traded firms, representing roughly 25% of the total sample of patents. *"Regulatory proficiency" indicates the total number of days a firm has experienced approval delays across all its submitted devices.

	N	Mean	SD	Range
FDA Admin. Data—Device Submissions (PMA and 510(k) Databases)				
Total	30,797	-	-	-
per Device Type	1,560 (Types)	19.7	78.1	[1, 1,927]
Total Submitting Firms	5,253	-	-	-
Firms per Device Type	1,560 (Types)	11.3	36.7	[1, 1,048]
Firm Regulatory Proficiency	1,554 (Types)	6.1yrs	18.2yrs	[0, 603.7yrs]*
FDA Admin. Data—Adverse Event Reports (MAUDE)				
Total	475,782	-	-	-
per Device Type	1,264 (Types)	376.4	2550.8	[1, 52,526]
Serious Events per Dev. type	612 (Types)	25.6	107.3	[1.0, 1,547]
Assets of Offending Firm	271,715	$3.2B	$12.7B	[0, $0.7T]
USPTO Device Patents				
Total	671,665	-	-	-
per Device Type	961 (Types)	698.9	2453.4	[1, 23,056]
Citations	671,665	10.6	56.4	[1, 5,067]
Market Valuation	201,638	$12.5M	$30M	[$40, $1.7B]
Applicant Assets	192,619	$26.1B	$53.5B	[$0.07M, $0.79T]

Table A8. Summary Statistics – Class II. Note: This table presents summary statistics only for Class II devices. See Kogan et al. (2017) for more information on the patent market valuation data, which was merged into my patent dataset. The CRSP/Compustat database was used to derive the total assets of the firms applying for patent protection and is a proxy for firm size. Market values and applicant assets are only available for patents filed by publicly traded firms, representing roughly 25% of the total sample of patents. *"Regulatory proficiency" indicates the total number of days a firm has experienced approval delays across all its submitted devices.

	N	Mean	SD	Range
FDA Admin. Data—Device Submissions (PMA and 510(k) Databases)				
Total	118,820	-	-	-
per Device Type	2,496 (Types)	47.6	131.2	[1, 2,457]
Total Submitting Firms	13,657	-	-	-
Firms per Device Type	2496 (Types)	20.7	44.2	[1, 747]
Firm Regulatory Proficiency	2,466 (Types)	11.9yrs	38.3yrs	[0, 669.3 yrs]*
FDA Admin. Data—Adverse Event Reports (MAUDE)				
Total	4,510,435	-	-	-
per Device Type	1,975 (Types)	2,283.8	162,560	[1, 0.41M]
Serious Events per Dev. type	1,238 (Types)	344.3	2,402	[1, 46,502]
Assets of Offending Firm	2,818,635	$3.3B	$6.3B	[$0, $0.7T]
USPTO Device Patents				
Total	567,204	-	-	-
per Device Type	1,100 (Types)	515.6	1,732.6	[1, 17,559]
Citations	567,213	19.2	115.8	[1, 5817]
Market Valuation	173,194	$13.8M	$31.5M	[0, $1.9B]
Applicant Assets	164,686	$27.5B	$56.6B	[$0.2M, $0.7T]

Table A9. Summary Statistics – Class III. Note: This table presents summary statistics only for Class III devices. See Kogan et al. (2017) for more information on the patent market valuation data, which was merged into my patent dataset. The CRSP/Compustat database was used to derive the total assets of the firms applying for patent protection and is a proxy for firm size. Market values and applicant assets are only available for patents filed by publicly traded firms, representing roughly 25% of the total sample of patents. *"Regulatory proficiency" indicates the total number of days a firm has experienced approval delays across all its submitted devices.

	N	Mean	SD	Range
FDA Admin. Data—Device Submissions (PMA and 510(k) Databases)				
Total	3,395	-	-	-
per Device Type	59 (Types)	57.5	148.1	[1, 795]
Total Submitting Firms	109	-	-	-
Firms per Device Type	59 (Types)	7.3	12.3	[1, 57]
Firm Regulatory Proficiency	3,184 (Types)	49.8yrs	74.7yrs	[0, 667.4yrs]*
FDA Admin. Data—Adverse Event Reports (MAUDE)				
Total	976,693	-	-	-
per Device Type	101 (Types)	9,670.2	32,432.6	[1, 0.2M]
Serious Events per Dev. type	78 (Types)	2,871	13,442.2	[1, 0.1M]
Assets of Offending Firm	786,010	$4.6B	$6.2B	[$0.6M, $0.7T]
USPTO Device Patents				
Total	9,423	-	-	-
per Device Type	52 (Types)	181.2	453.7	[1, 2536]
Citations	9,424	21.6	97.7	[1, 4265]
Market Valuation	2,633	$16.7M	$30.5M	[$0, $440M]
Applicant Assets	2,500	$15.5B	$33.6B	[$1.1M , $0.9T]

Table A10. Keywords Used in Text Analysis of Patent Claims. Note: The table presents the keywords related to product safety that were extracted using the Word2Vec algorithm. I label a patent as advancing safety if any of the above words are included in its claims section. Importantly, patent examiners heavily scrutinize the patent claims text for accuracy as the text codifies the right to singular ownership of the claimed advancement. Interestingly, some keywords indicate safety advancements in what the product is not: some inventors claim advancements in product safety by moving away from constructions that are "hazardous," "unsafe," or "dangerous." It is important to note that inventors would not reasonably claim a product advancement that would lead to more injuries. Thus, one can assume that these negative mentions can still be attributable to safety improvements.

Safety Advancement Keywords	
safety	hazard
safe	danger
safer	dangerous
endangering	harming
precautions	injuring
unsafe	injury
hazardous	jeopardizing
failsafe	risk
safely	complication
dangerous	jeopardizing

Table A11. Effect of Down-Classifications on Innovation (Using Borusyak et al. (2021) Estimator).
Note: The table presents estimates of equation 1.4.1, which is a difference-in-differences (DID) style OLS regression model. In this analysis, I drop all device types that do not exhibit any positive quantity of the given outcome. Column (1) presents the 5-year baseline average of treated device types for the outcomes listed on the left-hand side. Columns (2)–(5) present DID estimates for the listed outcomes using different control groups: namely, a matched control group, intuitively similar device types (treat similar diseases), "later-treated" device types (treated after sample window), and the full sample, respectively. Confidence intervals are calculated using Conley–Taber test statistics. +, *, **, and *** correspond with statistical significance at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

		DID Estimates			
Down-Classification	Pre-mean (1)	Matched (2)	Intuitive (3)	Later (4)	Full (5)
A. Class III to II:					
Patenting Rate	7.95	19.73*	27.70**	28.48**	22.11*
	(9.27)	(9.96)	(8.80)	(10.29)	(8.85)
Device Submission Rate	0.47	2.11***	1.85***	1.71***	1.76***
	(1.03)	(0.32)	(0.29)	(0.33)	(0.27)
Citations-Per-Patent Rate	9.06	17.60*	21.86*	17.07***	27.46***
	(20.65)	(7.61)	(8.76)	(4.90)	(7.15)
Average Patent Value	4.36	9.37***	11.72***	11.61***	11.82***
	(6.12)	(1.65)	(1.59)	(1.75)	(1.44)
Sample Size		1540	1056	920	60456
B. Class II to I:					
Patenting Rate	16.32	8.15	7.77	14.16**	31.04**
	(37.11)	(13.00)	(6.64)	(5.16)	(10.46)
Citations-Per-Patent Rate	0.64	6.84**	2.07+	4.01***	6.03***
	(0.48)	(2.09)	(1.18)	(0.94)	(1.42)
Average Patent Value	6.49	3.46***	0.86+	2.00***	5.00***
	(14.19)	(0.95)	(0.50)	(0.44)	(0.71)
Sample Size		15180	20592	27764	32472

Table A12. Effect of Down-Classifications on Market Structure (Using Borusyak et al. (2021) Estimator). Note: The table presents estimates of equation 1.4.1, which is a difference-in-differences (DID) style OLS regression model. In this analysis, I drop all device types that do not exhibit any positive quantity of the given outcome. Column (1) presents the 5-year baseline average of treated device types for the outcomes listed on the left-hand side. Columns (2)–(6) present DID estimates for a given outcome using different control groups: namely, a group matched on baseline prices, a group matched on baseline innovation and adverse event levels, an intuitively comparable group, a later-treated group, and the full sample of controls, respectively. Confidence intervals are calculated using Conley–Taber test statistics. +, *, **, and *** correspond with statistical significance at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

		DID Estimates				
Down-Classification	Pre-mean (1)	Price (2)	Matched (3)	Intuitive (4)	Later (5)	Full (6)
A. Class III to II:						
Amount Paid	95.68	-89.73***	-75.84*	-	-	-51.99***
	(123.78)	(25.35)	(34.42)	-	-	(10.85)
Sample Size		480	176	-	-	36240
Incumb. Entry (dev.)	0.40	-	1.17***	1.09***	1.02***	1.08***
	(0.91)	-	(0.11)	(0.11)	(0.12)	(0.09)
New Entry (dev.)	0.07	-	0.60***	0.61***	0.52**	0.55**
	(0.31)	-	(0.17)	(0.17)	(0.19)	(0.17)
Incumb. Entry (pat.)	1.47	-	2.36***	3.01***	3.69***	2.82***
	(1.78)	-	(0.59)	(0.56)	(0.69)	(0.53)
New Entry (pat.)	3.78	-	7.29+	11.54**	12.02**	10.04**
	(4.76)	-	(4.33)	(3.85)	(4.60)	(3.86)
Sample Size		-	1364	1056	920	60456
B. Class II to I:						
Incumb. Entry (pat.)	2.26	-	0.08	0.35	0.65*	1.43**
	(4.33)	-	(0.68)	(0.36)	(0.29)	(0.49)
New Entry (pat.)	7.27	-	4.24	2.82	5.11**	11.10***
	(16.87)	-	(3.87)	(2.05)	(1.61)	(3.07)
Sample Size		-	13552	20592	27764	32472

Table A13. Effect of Down-Classifications on Adverse Events (Using Borusyak et al. (2021) Estimator). Note: The table presents estimates of equation 1.4.1, which is a difference-in-differences (DID) style OLS regression model. In this analysis, I drop all device types that do not exhibit any positive quantity of the given outcome. Column (1) presents the 5-year baseline average of treated device types for the outcomes listed on the left-hand side. Columns (2)–(5) present DID estimates for the listed outcomes using different control groups: namely, a matched control group, intuitively similar device types (treat similar diseases), "later-treated" device types (treated after sample window), and the full sample, respectively. For column (4), Class III to II, control device types are treated after 2015; thus, all observations after 2015 are dropped. Confidence intervals are calculated using Conley–Taber test statistics. +, *, **, and *** correspond with statistical significance at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

		DID Estimates			
Down-Classification	Pre-mean (1)	Matched (2)	Intuitive (3)	Later (4)	Full Sample (5)
A. Class III to II:					
Emphasis on Safety	0.16	0.074+	-	-	-
	(0.21)	(0.038)	-	-	-
Life-Threatening Event Rate	0.07	0.59	0.81+	-0.58	-1.93
	(0.31)	(0.44)	(0.43)	(0.78)	(1.35)
Hospitalization Rate	0.25	3.36**	3.44**	2.27*	-2.21
	(0.84)	(1.14)	(1.14)	(0.93)	(1.97)
Mortality Rate	0.08	-0.50	1.08*	0.29	0.33
	(0.46)	(1.34)	(0.47)	(0.53)	(0.49)
Sample Size		588	644	528	38444
B. Class II to I:					
Emphasis on Safety	0.065	0.056***	-	-	-
	(0.218)	(0.012)	-	-	-
Life-Threatening Event Rate	0.07	-2.57	-0.36	-3.21	-3.16+
	(0.41)	(1.96)	(0.26)	(2.73)	(1.71)
Hospitalization Rate	0.15	-1.93**	-3.04	-4.84+	-5.44*
	(0.88)	(0.63)	(2.71)	(2.64)	(2.51)
Mortality Rate	0.23	-0.44*	-0.29	-0.47	-0.60***
	(1.98)	(0.17)	(0.29)	(0.29)	(0.17)
Sample Size		10332	13104	17668	20664

Table A14. Down-Classification Spillovers (Innovation). Note: The table presents estimates of equation 1.4.1, which is a difference-in-differences (DID) style OLS regression model for device types that are closely related to treated medical device types. Column (1) presents the 5-year baseline average of closely related device types for the outcomes listed on the left-hand side. Columns (2) and (3) present my OLS estimates of down-classifications on device types closely related to treated device types using different control criteria. Confidence intervals for my estimates in columns (2) and (3) are calculated using Conley–Taber test statistics. Column (2) presents the estimates when closely related groups are compared to matched control groups, whereas column (3) presents results from comparing against full sample controls. Standard errors allow for clusters at the PC level. +, *, **, and *** correspond with statistical significance at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

		DID Estimates	
Down-Classification	Pre-mean (1)	Matched (2)	Full Sample (3)
A. Class III to II:			
Patenting Rate	7.95	1.67	-3.91
	(9.27)	(2.56)	(3.89)
Device Approval Rate	0.47	0.06	-0.01
	(1.03)	(0.14)	(0.29)
Sample Size		792	179520
B. Class II to I:			
Patenting Rate	19.12	-1.49	1.72
	(39.50)	(3.41)	(4.63)
Sample Size		7656	179872

Table A15. Class III to II Device Types by Broad Device Category: Treated Group versus Intuitive Control Group. Note: The table presents the broad device types used in the treatment and intuitive control groups. No life-sustaining devices are considered in the treatment and control groups. When "Treatment" is 0, the description counts refer to the control group and refer to the treated group otherwise. The column"Implant" indicates the counts of device types that are implantable in the given broad device category.

Treatment	Category Description	Count	Implant
0	Anesthesiology devices—monitoring devices	1	0
	Cardiovascular devices—cardiovascular prosthetic devices	2	2
	Clinical chemistry—test systems	1	0
	Dental devices—therapeutic devices	1	0
	Gastroenterology-urology devices—therapeutic devices	1	0
	Immunology and microbiology devices—serological reagents	1	0
	Ophthalmic devices—therapeutic devices	2	0
	Orthopedic devices—prosthetic devices	1	1
	Radiology devices—diagnostic devices	2	0
1	Anesthesiology devices—monitoring devices	1	0
	Cardiovascular devices—cardiovascular prosthetic devices	2	2
	Clinical chemistry—test systems	1	0
	Dental devices—therapeutic devices	1	0
	Gastroenterology-urology devices—therapeutic devices	1	0
	Immunology and microbiology devices—serological reagents	1	0
	Ophthalmic devices—therapeutic devices	2	0
	Orthopedic devices—prosthetic devices	1	1
	Radiology devices—diagnostic devices	2	0

Table A16. Class II to I Treated Device Types by Broad Category. Note: The table presents the counts of broad device types used in the treatment group. No life-sustaining devices are considered. Implant counts are also provided.

Treatment	Category Description	Count	Implant
1	Anesthesiology devices—diagnostic devices	3	0
	Anesthesiology devices—miscellaneous	3	0
	Anesthesiology devices—monitoring devices	11	0
	Anesthesiology devices—therapeutic devices	23	0
	Cardiovascular devices—monitoring devices	5	0
	Cardiovascular devices—prosthetic devices	4	1
	Clinical chemistry—clinical chemistry test systems	6	0
	Clinical chemistry—clinical laboratory instruments	3	0
	Dental devices—diagnostic devices	2	0
	Dental devices—miscellaneous devices	1	0
	Dental devices—surgical devices	2	0
	Ear, nose, and throat devices—diagnostic devices	2	0
	Ear, nose, and throat devices—surgical devices	6	0
	Gastroenterology-urology devices—diagnostic devices	20	0
	Gastroenterology-urology devices—monitoring devices	1	0
	Gastroenterology-urology devices—surgical devices	10	0
	Gastroenterology-urology devices—therapeutic devices	19	1
	General and plastic surgery devices—surgical devices	1	0
	General hospital and personal use devices—miscellaneous devices	14	0
	General hospital and personal use devices—monitoring devices	5	0
	General hospital and personal use devices—therapeutic devices	7	0
	Hematology and pathology devices—manual hematology devices	4	0
	Hematology and pathology devices—used by blood manufacturer	4	0
	Immunology and microbiology devices—immunological test systems	14	0
	Immunology and microbiology devices—microbiology devices	1	0
	Immunology and microbiology devices—serological reagents	47	0
	Neurological devices—diagnostic devices	1	0
	Neurological devices—therapeutic devices	1	0
	Obstetrical and gynecological devices—diagnostic devices	1	0
	Obstetrical and gynecological devices—surgical devices	6	0
	Obstetrical and gynecological devices—therapeutic devices	2	0
	Ophthalmic devices—diagnostic devices	4	0
	Ophthalmic devices—prosthetic devices	7	4
	Orthopedic devices—diagnostic devices	1	0
	Orthopedic devices—surgical devices	1	0
	Physical medicine devices—diagnostic devices	5	0
	Physical medicine devices—prosthetic devices	6	0
	Physical medicine devices—	19	0
	Radiology devices—diagnostic devices	9	0
	Radiology devices—miscellaneous devices	11	0
	Radiology devices—therapeutic devices	1	0

Table A17. Class II to I Intuitive Control Device Types by Category. Note: The table presents the counts of broad device types used in the control group. No life-sustaining devices are considered. Implant counts are also provided.

Treatment	Category Description	Count	Implant
0	Anesthesiology devices—diagnostic devices	3	0
	Anesthesiology devices—miscellaneous	3	0
	Anesthesiology devices—monitoring devices	11	0
	Anesthesiology devices—therapeutic devices	23	0
	Cardiovascular devices—cardiovascular monitoring devices	5	0
	Cardiovascular devices—cardiovascular prosthetic devices	2	1
	Cardiovascular devices—cardiovascular surgical devices	2	0
	Clinical chemistry—clinical chemistry test systems	6	0
	Clinical chemistry—clinical laboratory instruments	3	0
	Dental devices—diagnostic devices	2	0
	Dental devices—miscellaneous devices	1	0
	Dental devices—surgical devices	2	0
	Ear, nose, and throat devices—diagnostic devices	2	0
	Ear, nose, and throat devices—surgical devices	6	0
	Gastroenterology-urology devices—diagnostic devices	20	0
	Gastroenterology-urology devices—monitoring devices	1	0
	Gastroenterology-urology devices—surgical devices	10	0
	Gastroenterology-urology devices—therapeutic devices	19	1
	General and plastic surgery devices—surgical devices	1	0
	General hospital and personal use devices—miscellaneous devices	14	0
	General hospital and personal use devices—monitoring devices	5	0
	General hospital and personal use devices—therapeutic devices	7	0
	Hematology and pathology devices—manual devices	4	0
	Hematology and pathology devices—used by blood manufacturer	4	0
	Immunology and microbiology devices—immunological test systems	14	0
	Immunology and microbiology devices—microbiology devices	1	0
	Immunology and microbiology devices—serological reagents	47	0
	Neurological devices—diagnostic devices	1	0
	Neurological devices—therapeutic devices	1	0
	Obstetrical and gynecological devices—diagnostic devices	1	0
	Obstetrical and gynecological devices—surgical devices	6	0
	Obstetrical and gynecological devices—therapeutic devices	2	0
	Ophthalmic devices—diagnostic devices	4	0
	Ophthalmic devices—prosthetic devices	4	4
	Ophthalmic devices—surgical devices	3	0
	Orthopedic devices—diagnostic devices	1	0
	Orthopedic devices—surgical devices	1	0
	Physical medicine devices—diagnostic devices	5	0
	Physical medicine devices—prosthetic devices	6	0
	Physical medicine devices—therapeutic devices	19	0
	Radiology devices—diagnostic devices	9	0
	Radiology devices—therapeutic devices	12	0

Table A18. Effect of Down-Classifications on Innovation (Drop No Counts). Note: The table presents estimates of equation 1.4.1, which is a difference-in-differences (DID) style OLS regression model. In this analysis, I drop all device types that do not exhibit any positive quantity of the given outcome. Column (1) presents the 5-year baseline average of treated device types for the outcomes listed on the left-hand side. Columns (2)–(5) present DID estimates for the listed outcomes using different control groups: namely, a matched control group, intuitively similar device types (treat similar diseases), "later-treated" device types (treated after sample window), and the full sample, respectively. Confidence intervals are calculated using Conley–Taber test statistics. +, *, **, and *** correspond with statistical significance at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

		DID Estimates			
Down-Classification	Pre-mean (1)	Matched (2)	Intuitive (3)	Later (4)	Full (5)
A. Class III to II:					
Patenting Rate	7.95	15.31**	23.68*	24.64*	7.77
	(9.27)	(5.58)	(10.20)	(10.94)	(25.79)
Device Submission Rate	0.47	2.69***	2.36**	2.27**	2.22***
	(1.03)	(0.59)	(0.76)	(0.72)	(0.34)
Citations-Per-Patent Rate	9.06	16.87*	-5.61	15.91*	20.13**
	(20.65)	(7.57)	(13.90)	(6.22)	(7.58)
Average Patent Value	4.36	8.56***	9.88**	10.45**	8.14***
	(6.12)	(1.67)	(3.49)	(3.41)	(2.32)
Sample Size		1452	660	680	21340
B. Class II to I:					
Patenting Rate	16.32	7.34	13.72	25.22**	29.17***
	(37.11)	(4.87)	(12.54)	(9.61)	(7.19)
Citations-Per-Patent Rate	0.64	6.85**	4.13*	7.52***	6.00***
	(0.48)	(2.28)	(1.84)	(1.49)	(1.38)
Average Patent Value	6.49	3.58***	2.06*	4.35***	4.47***
	(14.19)	(0.72)	(0.93)	(1.03)	(0.77)
Sample Size		14740	9328	9768	25784

Table A19. Effect of Down-Classifications on Market Structure (Drop No Counts). Note: The table presents estimates of equation 1.4.1, which is a difference-in-differences (DID) style OLS regression model. In this analysis, I drop all device types that do not exhibit any positive quantity of the given outcome. Column (1) presents the 5-year baseline average of treated device types for the outcomes listed on the left-hand side. Columns (2)–(6) present DID estimates for a given outcome using different control groups: namely, a group matched on baseline prices, a group matched on baseline innovation and adverse event levels, an intuitively comparable group, a later-treated group, and the full sample of controls, respectively. Confidence intervals are calculated using Conley–Taber test statistics. +, *, **, and *** correspond with statistical significance at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

		DID Estimates				
Down-Classification	Pre-mean (1)	Price (2)	Matched (3)	Intuitive (4)	Later (5)	Full (6)
A. Class III to II:						
Procedure Price	95.31	-58.25**	-43.54**	-	-	-27.50
	(123.95)	(21.16)	(15.66)	-	-	(144.11)
Sample Size		160	176	-	-	36240
Incumb. Entry (dev.)	0.40	-	1.58***	1.50**	1.49**	1.44***
	(0.91)	-	(0.35)	(0.54)	(0.54)	(0.21)
New Entry (dev.)	0.07	-	0.94***	0.98**	0.79**	0.88***
	(0.31)	-	(0.23)	(0.31)	(0.26)	(0.20)
Incumb. Entry (pat.)	1.47	-	1.96***	2.19+	3.33*	1.28
	(1.78)	-	(0.59)	(1.12)	(1.52)	(1.40)
New Entry (pat.)	3.78	-	6.14***	11.75*	12.65**	6.10
	(4.76)	-	(1.65)	(4.57)	(4.79)	(9.19)
Sample Size		-	1276	616	680	23848
B. Class II to I:						
Incumb. Entry (pat.)	2.26	-	0.02	0.59	1.09+	1.33**
	(4.33)	-	(0.47)	(0.69)	(0.59)	(0.44)
New Entry (pat.)	7.27	-	4.00+	5.18	9.26**	10.11***
	(16.87)	-	(2.07)	(4.17)	(3.29)	(2.26)
Sample Size		-	13288	9988	12672	28952

Table A20. Effect of Down-Classifications on Adverse Events (Drop No Counts). Note: The table presents estimates of equation 1.4.1, which is a difference-in-differences (DID) style OLS regression model. In this analysis, I drop all device types that do not exhibit any positive quantity of the given outcome. Column (1) presents the 5-year baseline average of treated device types for the outcomes listed on the left-hand side. Columns (2)–(5) present DID estimates for the listed outcomes using different control groups: namely, a matched control group, intuitively similar device types (treat similar diseases), "later-treated" device types (treated after sample window), and the full sample, respectively. For column (4), Class III to II, control device types are treated after 2015; thus, all observations after 2015 are dropped. Confidence intervals are calculated using Conley–Taber test statistics. +, *, **, and *** correspond with statistical significance at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

		DID Estimates			
Down-Classification	Pre-mean (1)	Matched (2)	Intuitive (3)	Later (4)	Full Sample (5)
A. Class III to II:					
Emphasis on Safety	0.16	0.073+	-	-	-
	(0.21)	(0.039)	-	-	-
Life-Threatening Event Rate	0.07	1.31	1.64	-1.96	-8.57
	(0.31)	(0.82)	(1.11)	(1.26)	(5.72)
Hospitalization Rate	0.25	4.30**	5.32*	2.38	-9.43
	(0.84)	(1.62)	(2.38)	(1.96)	(8.09)
Mortality Rate	0.08	-3.28	2.78*	-0.09	0.16
	(0.46)	(4.72)	(1.40)	(1.23)	(7.50)
Sample Size		336	196	216	11452
B. Class II to I:					
Emphasis on Safety	0.065	0.05***	-	-	-
	(0.218)	(0.012)	-	-	-
Life-Threatening Event Rate	0.07	-8.07	-1.51+	-15.92*	-9.17*
	(0.43)	(5.07)	(0.78)	(7.85)	(4.38)
Hospitalization Rate	0.17	-6.25***	-7.80+	-16.76*	-11.63*
	(0.94)	(1.24)	(3.98)	(7.62)	(5.32)
Mortality Rate	0.26	-1.72***	-1.03	-2.60+	-1.70*
	(2.13)	(0.39)	(0.77)	(1.37)	(0.75)
Sample Size		3612	3276	3752	7168

Table A21. Flattening the Learning Curve Simulation—Unique Devices Approved. Note: This table presents the results of the simulation exercise described in appendix A3.2, which simulates the effect of flattening the learning curve on the rate of unique devices approved at an annual frequency by asset quartiles. Figure A9 illustrates this flattening exercise. Standard errors generated from a Monte Carlo procedure are presented in parenthesis below the estimates. This procedure produces estimates across repeated random draws from the empirical distribution of firm characteristics to calculate confidence intervals. I express changes as percent changes relative to the gamma at a 0.075 baseline. I flatten the learning curve relative to the firm with the highest experience in the data. In the table, gamma begins at its initial starting point estimated in equation A3.1. Subsequent rows in the table show the percent change in the rate of unique device submissions as gamma, the learning rate, is reduced. These changes are presented for each experience quartile for Class III device manufacturers. The far-right column presents the total percent change in unique devices approved from a flattening of the learning curve relative to the baseline frequency of unique device submissions.

	Percent Changes				
γ	$T_{Sum,25}$	$T_{Sum,50}$	$T_{Sum,75}$	$T_{Sum,100}$	Total %Δ
0.075	0.0	0.0	0.0	0.0	0.0
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
0.067	13.3	10.2	4.8	2.49	6.19
	(10.17)	(8.12)	(3.84)	(2.95)	(5.29)
0.058	29.67	16.49	8.94	3.75	11.57
	(16.78)	(9.57)	(5.78)	(2.98)	(7.92)
0.05	59.0	25.64	14.07	6.47	19.32
	(26.66)	(12.88)	(6.0)	(4.82)	(11.38)
0.042	68.55	35.46	21.07	8.86	25.98
	(24.77)	(16.03)	(9.65)	(4.17)	(13.94)
0.033	85.34	46.38	23.74	9.35	31.75
	(31.51)	(20.97)	(10.22)	(4.66)	(17.95)
0.025	110.02	54.42	25.24	12.35	38.46
	(41.96)	(25.91)	(8.47)	(6.22)	(21.76)
0.017	150.65	64.74	36.93	14.69	48.77
	(61.78)	(22.15)	(12.04)	(7.11)	(25.41)
0.008	151.55	75.92	34.03	15.58	51.9
	(48.99)	(25.45)	(11.69)	(7.45)	(27.68)
0.0	186.41	88.62	43.45	19.13	63.32
	(74.03)	(29.59)	(11.61)	(7.67)	(33.3)

Table A22. Cross-Correlation Between Firm Size and FDA Experience. Note: The table presents the correlation coefficients between firm assets (size) and firm cumulative FDA experience. Data includes firms in the FDA database that were fuzzy matched to publicly traded firms in the CRSP database.

Variables	Cumulative FDA Experience	Firm Assets
Cumulative FDA Experience	1.00	
Firm Assets	-0.00	1.00
	(1.00)	

Table A23. Estimation of Learning Curve Parameters (in Days). Note: The table presents the estimates of equation A3.1, which estimates the learning coefficient gamma and the baseline time requirement, beta(Rc), for both Class III original PMA approvals (column 1) and Class II 510(k) approvals (column 2) of unique devices via OLS. The estimates for Class III devices are calculated by only considering the approval times of filed original PMAs by firms with at least one day of prior experience navigating FDA regulations. The estimates for Class II devices are calculated by only considering the approval times of 510(k) documents for unique devices that were submitted by firms with at least one day of prior experience navigating FDA regulations. Prior experience is calculated using approval times when filing any prior documentation type (510(k) or PMAs). Standard errors are clustered at the firm level. +, *, **, and *** correspond with statistical significance at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

	Class III Coeff./SE	Class II Coeff./SE
γ	0.075*	0.032***
	(0.033)	(0.004)
$log(\beta(R_c))$	6.678***	4.481***
	(0.326)	(0.031)
N	631	84,909
Clusters	94	9,067
Device Type Effects	Yes	No
Firm Effects	Yes	Yes
Device Type by Year Effects	No	Yes
SEs in Parentheses	Clustered	Clustered

Table A24. Effect of Down-Classifications on Innovation: Restricted Patent Sample Note: The table presents estimates of equation 1.4.1, which is a difference-in-differences (DID) style OLS regression model. This table differs from table 1.2 in that it presents estimates from an estimation that uses a restricted patent sample described in appendix A4, and only presents the patenting rate outcome. Simply put, patents in this analysis include only those labeled as health-related and non-drug by patent examiners. Patents are derived from the USPTO patent database. Column (1) presents the 5-year baseline average of treated device types for the outcomes listed on the left-hand side. Columns (2)–(5) present DID estimates for the listed outcomes using different control groups: namely, a matched control group, intuitively similar device types (treat similar diseases), "later-treated" device types (treated after sample window), and the full sample, respectively. For column (4), Class III to II, control device types are treated after 2015; thus, all observations after 2015 are dropped. Confidence intervals are calculated using Conley–Taber test statistics. +, *, **, and *** correspond with statistical significance at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

		DID Estimates			
Down-Classification	Pre-mean (1)	Matched (2)	Intuitive (3)	Later (4)	Full (5)
A. Class III to II:					
Patenting Rate	4.6	10.8**	15.81**	15.79*	14.07*
	(6.18)	(3.3)	(5.9)	(6.88)	(6.68)
Sample Size		1628	1056	920	60456
B. Class II to I:					
Patenting Rate	3.99	4.8*	1.91	3.01*	6.97**
	(13.74)	(2.26)	(1.61)	(1.48)	(2.5)
Sample Size		12540	20592	27764	32472

Appendix: Chapter 2

B1 Patent Trait Appendix: Examples, Illustrations, and Historical Narratives

This appendix provides descriptions and examples of patents that emphasize the traits used in our analysis. These traits include "cost," "simplicity," "adjustability," "appliances," "materials," "durability," "appearance," and "comfort." We connect these traits to specific instances of concrete technological change. We also detail the relative ease or difficulty of identifying each traits using a set of keywords.

B1.1 Cost

Technological advancements related to our "cost" trait led to a cheaper production process. One artificial limb patent emphasizing costs, for example, claims an advancement that improves the modularity of the device, allowing for uniform construction, by letting the manufacturer "replace or modify any one portion of [the limb] without altering the other portions and at very small expense" (U.S. 35,686; 1862). Another patent describes a new limb that facilitates cheaper, uniform construction by using parts that are adjustable to different users. It reads, "adjustment of the parts of an artificial limb...to adapt it to the length of the natural limb and conformation of the foot of the intending wearer, by which means the necessity of making a limb to suit each particular case is to a great extent obviated, and in consequence, the cost of manufacture is considerably reduced" (U.S. 37,282; 1863). These patents describe a more modular, uniform artificial limb design that leads to a cheaper production process.

Straightforward cost-oriented innovations are also present in other technological categories that form our control groups. A patent for a Civil War-era carriage cover, for example, emphasizes a "cheap, light and convenient covering from storms or the heat of the sun" (U.S. 32,477; 1861). Examples such as this, from technology classes other than prosthetic devices, lead us to designate "cost" as a trait for which control groups can reasonably be identified.

See Figure B1 for keywords we used to identify "cost" innovations, including words like "cheap" and "economical." We identify "cost" as a trait for which text analysis methods can be implemented effectively.

B1.2 Adjustability

New technologies that allow a product to be adjusted to user specifications are labeled as having the "adjustability" trait. Adjustable products enhance mass producibility by bypassing the need to tailor-make a product to accommodate the needs of a specific individual.

An example of an artificial limb patent that describes this type of advancement reads, "The improved artificial leg ... is so constructed that its length may be easily and nicely adjusted to suit the wearer" (U.S. 35,937; 1862). As mentioned above, U.S. patent 37,282 (1863) also advances mass producibility through the use of adjustable parts to adapt the limb "to the length of the natural limb...of the intending wearer...by which means the necessity of making a limb to suit each particular case is to a great extent obviated" (U.S. 37,282; 1863). These examples highlight advancements in artificial limbs that allow for uniform construction by enabling limbs to fit the user through adjustable parts.

Adjustability is broadly applicable to many types of technologies. For example, this trait is relevant when describing advancements in machinery that eliminate the need for additional parts to adapt to user specifications. These advancements simplify the production process by shedding extraneous components. One such patent describes a machine that can be "made adjustable in inclination" to suit the needs of multiple users (US 10,687; 1854). Although this trait is straightforward for wearable products, it is slightly more difficult to identify in machinery

technologies using simple keyword searches. This leads us to identify "adjustability" as a trait for which control groups can reasonably be constructed using other technological classes. However, it is not quite as straightforward as the "cost" trait.

See Figure B1 for keywords we used to identify "adjustability" innovations, including words like "adjust" and "adjustability." We note, however, that the concept of "adjustability" that we have in mind is more cleanly identified through close readings than through keywords. The keywords alone, for example, sometimes captured patents simply referring to the process of "adjusting" a screw to build the product. This instruction is obviously not an advancement in mass production. Thus, close readings can better identify patents for which the emphasis is on the product's mass producibility. This leads us to identify "adjustability" as a trait for which we rate the trait's ease of interpretability as weak, despite the clarity of its economic content.

B1.3 Simplicity

The trait "simplicity," as used in 19th-century artificial limb and mechanical patents, describes advancements that simplify the design and fabrication of new technologies. For example, one artificial limb patent states the use of a knee joint that mimics the natural simplicity of the human knee joint, avoiding unnecessary parts and ensuring "great simplicity, and therefore cheapness" (U.S. 37,087; 1862). Figure B12 shows a diagram of the knee joint with comparisons to the simplicity of the natural human knee joint. This patent emphasizes an advancement that leads to simple construction and lower production costs.

Like the previous two traits, the language that connects "simplicity" to a streamlined production process is not unique to artificial limbs. For example, a mechanical patent from the same era describes an advancement in a water pump as being "simple and cheap" (U.S. 15,221; 1856). Together, these examples illustrate the consistency of the language linked to "simplicity" across technology classes and highlight a trait whose meaning is easily derived in text analysis. This leads us to identify "simplicity" as a trait for which the control groups can reasonably be constructed using other technological classes.

Figure B1 provides the list of keywords we used to identify technological advancements in "simplicity." The keywords used include "simple," "difficult," and "complex." The straightforward meanings of the relevant keywords help illustrate why we identify "simplicity" as a trait for which text analysis methods can be implemented effectively.

B1.4 Appliances

Patents emphasizing improvements in tool attachments for artificial limbs are deemed as having the "appliances" trait. Such tools allow artificial limb wearers to operate machinery and perform a trade or skill, facilitating integration into the post-war workforce. "Appliances" is an example of a trait that is highly specific to artificial limbs as an applied technology. We thus identify "appliances" as a trait for which it is not particularly useful to construct control groups using other technological classes.

Despite being highly-specific, the associated economic content of our "appliances" trait is clearly defined. An example of an "appliances" innovation from a U.S. inventor during the World War I era states, "other appliances may be readily fastened in the arm end and tightly gripped there-by" (U.S. 1,213,222; 1917). A similar emphasis on attachable tools was seen in British patents during WWI. One such patent emphasizes that, "the invention has for its object to provide a mechanically worked elbow joint to which may be fitted a lower forearm member with or without a hand or an extension piece for appliances and other fitments" (GB113329A; 1917). Figures B3, B4, B5, and B6 show examples of these new appliance technologies, including a hand for writing, for soldering, and for hammering.

See Figure B1 for keywords we used to identify "appliances" innovations, including words like "appliances" and "fittings." This terminology highlights that "appliances" is a trait for which domain-specific knowledge is essential for connecting text to the relevant economic concept. Once that domain-specific knowledge has been obtained, however, the nature of the technological advance is very clear, as illustrated in Figures B3, B4, B5, and B6.

B1.5 Materials

Technological advancements in "materials" signify new materials, substances, compounds, or compositions used in the production process. Such advancements may lead to more efficient production processes and increased functionality.

An artificial limb patent describes one such advancement stating, "The socket...is composed of hard or vulcanized India-rubber...the rubber socket is simply tightened down upon the stump by means of the leather straps, and a perfect fit is secured at all times" (U.S. 38,550; 1863). The new use of vulcanized rubber improved the fit of artificial limbs. Another patent emphasizes a material advancement that leads to a more efficient production process, saying, "the foot and hand...[are] a composition of 'sponge rubber'...by this means I avoid the use of springs, pivots, joints... and also avoid the great expense and wear, making the limbs cheaper and more durable" (US 40,763; 1863). This trait can be complicated to encode as certain materials may only be relevant for a given technological class and may only be "innovative" for a limited time.

See Figure B1 for keywords we used to identify "materials" innovations, including words like "vulcanized" and "duralumin," both of which were new materials in the 19th and 20th centuries, respectively. These keywords help to illustrate that highly specialized knowledge may thus be necessary to capture materials innovations using text. Additionally, the materials associated with innovative designs will vary across technological classes, which complicates the construction of control groups. We thus identify "materials" as a trait for which it is not particularly useful to construct control groups using other technological classes.

B1.6 Durability

Improvements in "durability" signify inventions that aim to prevent the deterioration of an artificial limb over time. These improvements often utilized new materials or methods to create artificial limbs that lasted longer and required less-frequent replacement or repair.

An artificial limb inventor during the Civil War describes the new design of an artificial

leg by which "a strong and durable leg can be made" (U.S. 46,687; 1865). To achieve this level of durability, the inventor utilizes an innovative pear-shaped button to secure the movement of the artificial leg even when bent. Durability is fairly encodable in control classes as well. In one example, an inventor emphasizes a sounder construction of a wood boring machine for which the cogwheels within the frame "are arranged in a convenient and durable manner" (U.S. 3,645; 1844).

See Figure B1 for keywords we used to identify "durability" innovations, including words like "rot" and "burst." As with our materials trait, these keywords help to illustrate that specialized knowledge may be necessary to capture durability innovations using text. These keywords are mostly related to durability innovations for technologies made of wood, a central material of Civil War limb manufacturing. Additionally, the durability associated with innovative designs may vary across technological classes, which includes aspects of unique materials used during construction to improve durability. These insights complicate the construction of control groups. We thus identify "durability" as a trait for which it is difficult to encode in prosthetic limbs and control technological classes.

B1.7 Appearance

Artificial limb patents emphasizing a natural, life-like, tasteful, and neat appearance are labeled as having our "appearance" trait. These limbs are more discrete and make the artificial limb less obvious.

One such patent emphasizing "appearance" illustrates that "[this construction]...gives the limb a more natural appearance" (US55,645; 1866). Another patent describes the construction of an artificial hand and emphasizes its "most natural appearance" due to a "substantially smooth and continuous surface" (US 1,173,219; 1915). The top panel of Figure B7 illustrates this new technology with a more natural appearance relative to the predominant "Carnes hand" in the lower panel (US 999,484; 1910). Notice the continuous and smooth surface of the natural hand, especially at the joints, when compared to the more mechanical and rigid joints of The Carne's

Hand.

Appearance is also relevant for certain user-oriented mechanical innovations. For example, an advancement in cotton gins aims " to produce the finest sample or make the best and most presentable appearance " (U.S. 418,084; 1889). In this case, the quality of the output (cotton) depends on its presentability. For some mechanical innovations, however, appearance is not as relevant. This is an example of a trait for which control technologies must be selected carefully to ensure the trait's relevance, and where estimation using a simple time series changes may be preferable to using other classes of technologies to construct a control group.

See Figure B1 for keywords we used to identify "appearance" innovations, including words like "neat" and "tasteful." We identify "appearance" as a trait for which ease of interpretability is relatively strong.

B1.8 Comfort

Many 19th-century artificial limbs were quite uncomfortable, noisy, and smelly. Advancements to improve circulation and make limbs more comfortable are labeled as having the "comfort" trait.

An example of a patent that claims an artificial limb that is more comfortable is given in U.S. patent 53,206 (1866). The inventor emphasizes a novel way of constructing the inner lining of artificial limbs using cork sheets instead of traditional hard leather or rubber materials. He describes the invention as having a "smooth, soft surface, that is not materially affected [by] perspiration, because the pores in the cork allow said perspiration to escape, and said cork affords a pleasant, smooth surface to the tender stump." Figure B16 illustrates the construction of this cork lining. Some mechanical patents also emphasize comfort by, for example, suggesting that the sitting apparatus in the machine is made more comfortable for the user (U.S. 44,198; 1864).

Although some mechanical patents emphasize comfort, this trait stands in contrast with "simplicity" as a relatively complex trait. Difficulties arose as the language used to indicate a product's "comfort" was often ambiguous. For example, the word "disturbing" often connotes

bodily discomfort in prosthetic device patents. In mechanical classes, by contrast, the word "disturbing" tends to have meanings connected to the device's functionality (e.g., "disconnecting or disturbing the pump"). Thus, machine learning algorithms helped improve the accuracy of our "comfort" labels in the control group by overcoming these ambiguities. However, "comfort" is another example of a trait for which care should be taken when selecting control technologies, and where estimation using a simple time series methods may be preferable.

See Figure B1 for keywords we used to identify "comfort" innovations, including words like "circulation" and "pain." We identify "comfort" as a trait for which the ease of interpretability is moderate. Despite the clarity of the economic content itself, the semantic complexity of the trait is non-trivial, in part because of variations in how comfort might be described across technology classes.

B1.9 How Traits Relate to Technologies Influenced by Procurement

In this section, we detail how the traits we analyze capture technological changes as influenced by the desires of wartime procurers. First, we describe a set of traits related to the reintegration of veterans with amputated limbs into the workforce. We supplement this discussion with historical evidence on the demands of World War I era procurers. Then, we highlight traits related to advancements in mass production driven by the need to provide an unprecedented demand for artificial limbs associated with both the Civil War and World War I. Lastly, we detail how competitive pressures from consumer-directed limb purchases steered inventors to entice veterans with more desirable limbs.

Technologies for Employment and Social Reintegration (Appliances, Appearance)

Before World War I, the cost of the U.S. Civil War pension system outpaced the cost of the Civil War itself. In response, the U.S. government implemented a rehabilitation system focused on reintegrating veterans with amputated limbs into the workplace. These veterans "were expected to become citizen-workers...not to languish at the expense of the U.S. Treasury.

In a real sense, they were expected to be the opposite of the Civil War veteran" Linker (2011; p. 13). The British, too, learned from the American Civil War experience and focused on providing limbs geared to improve the employment prospects of veterans with amputated limbs. To accomplish this, both governments launched new initiatives to train veterans with amputated limbs to use artificial limbs in a new skill or trade before returning home. They contracted with limb manufacturers to compete against one another to invent artificial limb attachments for these trades (Kowalsky 2007).

Together, these forces led inventors to focus on technologies that improved the utility of artificial limbs (see Figure 2.3 and Table 2.5), with the increase being particularly strong in Britain. We measure changes in these technologies using the trait "appliances." Figures B3, B4, B5, and B6 show improvements in the utility of artificial limbs during World War I. Figure B3 illustrates a case of a soldier fitted with artificial arms that facilitate writing. Figure B4, taken from Linker (2011), shows a veteran with an amputated arm using a "utility arm" with a welding attachment. Figure B5 shows a diagram from U.S. patent 1,213,222 (1917), which illustrates a new artificial arm with an attachable hammer. Figure B6 displays a photo taken at Roehampton (a British army-training facility during World War I) that shows soldiers using various interchangeable terminal devices designed for specific trades. These artificial limb innovations facilitated reintegration into employment upon returning home.

The emphasis on the utility of limbs was coupled with a focus on improving limb appearance (see the bottom-right panel of Figure B2). Institutions strove to disguise the disability of veterans with amputated limbs. A War Risk Insurance Bureau chief noted that "one of the most useful and necessary duties of this department will be to prescribe and furnish medical and surgical treatment in order that disabilities may be reduced or caused to disappear entirely" (Linker 2011; p. 100).

Inventors responded to these desires by creating more life-like artificial limbs. These technologies are captured by our "appearance" trait. Figure B7 illustrates a new technology displayed in U.S. patent 1,173,219 (1915), which emphasizes a more natural-looking hand

through the use of continuous and smooth surfaces.

Mass Production During War

The Civil War brought an unprecedented demand shock to the U.S. artificial limb industry. The surge in demand led manufacturers to increase the mass producibility of their limbs. Manufacturers brought new materials advancements that made artificial limbs cheaper, simpler, and adjustable to user specifications (see Figure B27). For example, Amasa Marks, a prominent limb manufacturer whose firm persisted into World War I, filed U.S. patent 40,763 (1863), which details the construction of limb appendages using one such new material. The patent reads, "making the wearing parts of the limbs...of a composition of 'Sponge rubber'... [giving] the requisite degree of elasticity...making the limbs cheaper and more durable." Marks' use of vulcanized rubber allowed his limbs to be mass-producible through cheaper components that adjust to different stump sizes (elastic). Figure B8 shows a diagram from the patent illustrating the new materials technology.

Inventors also emphasized adjustability as a way to mass-produce limbs to meet pressing demand. U.S. patent 66,728 (1867) emphasizes an adjustable lacer for artificial limbs allowing a close fit to knee joints of different sizes. Figure B10 shows a diagram of this invention. U.S. patent 35,937 (1862) highlights the use of a spindle in the knee joint that allows the limb to adjust to the height of any wearer. Figure B11 shows the construction of this limb, with part D showing the adjustable spindle at the knee joint.

Simpler limbs also made for a quicker and less labor-intensive production process. U.S. patent 37,087 (1862) states the use of an artificial knee joint that mimics the natural simplicity of the human knee joint, avoiding unnecessary parts and ensuring "great simplicity, and therefore cheapness." The inventor describes a hinge joint of the artificial knee as one constructed of only "two principal parts, the upper part, representing the femur...and the lower part, representing the tibia." In contrast to Figure B13, which shows a more complex knee joint, Figure B12 shows a diagram of the described knee joint, illustrating the simplicity of the invention.

Although a strong domestic manufacturing presence was established during the Civil War, U.S. manufacturers were enlisted to meet global artificial limb needs during WWI. This led U.S. artificial limb manufacturers to invest further in standardization (Guyatt 2001; p. 313).

Inventors during World War I used modular construction to keep up with global demand. The "E-Z limb" was a standard-issue, temporary limb for acclimating veterans with amputated limbs to the use of an artificial limb before being discharged from military service. These limbs were modular and lightweight to facilitate mass production and showcased a smooth flesh-colored exterior that resembled the "shape of a real-life human leg" (Linker 2011; p. 109). Figure B9 illustrates the features of "E-Z limb." An increased emphasis on adjustability was thus common to prosthetic device patents during both World War I and the Civil War (see Table 2.5 and Figures B27 and B28). By contrast, inventors exhibited a much smaller increase in their emphasis on cost and simplicity during World War I than during the Civil War (again, see Table 2.5 and Figures B27 and B28).

Cost and Comfort Oriented Innovation During the Civil War

Lastly, two features of Civil War-era procurement contributed to increases in inventors' emphasis on cost and comfort. With respect to cost, the government's modest, fixed price reimbursement rates gave Civil War-era limb manufacturers a strong incentive to reduce production costs. Civil War-era inventors responded by increasing their emphasis on making limbs inexpensive (see Figure B27). One such artificial limb patent detailing a cost innovation states, "[The artificial limb] is simple, cheap..." (U.S. 37,637; 1863). To achieve cheapness, the inventor sheds "the use of straps around the waist or shoulder" and obviates "tedious fitting" by using a "bucket or socket to receive the stump of the amputated limb," which can secure the limb to the stump. This new technology is shown in Figure B14. Another inventor claims a new artificial arm design constructed entirely out of metal, "avoiding the use of catgut, whalebone, wood, or any other organic substance" and thus leading to "cheapness" (U.S. 40,397; 1863). Figure B15 shows the design of this metallic artificial arm.

With respect to comfort, Civil War veterans with amputated limbs received government limb allowances to finance the purchase of an artificial limb of their choosing. In addition, limb purchase was not required. This appears, in practice, to have led at least some manufacturers to strive to produce more comfortable artificial limbs (see the top-left panel of Figure B2). In one example, an inventor details a new way to construct the inner lining of artificial limbs by using cork sheets. This construction is described as having a "smooth, soft surface, that is not materially affected [by] perspiration, because the pores in the cork allow said perspiration to escape, and said cork affords a pleasant, smooth surface to the tender stump"(U.S 53,206; 1866). Figure B16 illustrates the construction of this cork lining.

By contrast, during World War I, the U.S. government de-emphasized the comfort of veterans with amputated limbs in favor of a strict rehabilitation program. Indeed, this program incorporated regiments of intentionally inflicting pain out of fear that a less severe approach would hinder rehabilitation. One source notes, "By eliciting pain from disabled soldiers, then, physiotherapists complied with the greater vision of the rehabilitation project...a vision fueled by the fear that overly sympathetic women would ruin a man's prospect of successful rehabilitation" (Linker 2011; p. 75). Soldiers often complained that government-provided limbs were painful (Linker 2011; p. 114). In the prosthetic device patents, we see this reflected in our analysis of the "comfort" trait, with inventors de-emphasizing comfort during World War I (see the top-right panel of Figure B2).

Adjustability	Cost	Materials	Simplicity	Appearance	Comfort	Appliances	Durability
adjust (I/W)	cheap (I/W)	substances	simple (I/W)	conceal (I/W)	unpleas (I/W)	appliance	durability
adjusting (E)	expens (I/W)	materials	simpli (I/W)	appearance (I/W)	circulation	appliances	strength
	inexpen (I/W)	compounds	simplif (I/W)	finish	noise	fittings	durabl (I/W)
	cost (I/W)	compositions	difficult	life (I/W) (P)	noisy	tools	strength
	econom (I/W)	vulcanized	complex	unsight (I/W)	noiseless	fitments	strengthen
		duralumin	complicat (I/W)	sightly (I/W)	noiselessly		strong
		celluloid	simplicity	beautif (I/W)	perspiration		dirt
		laminated		beauty	comfort (I/W)		waterproof
		polymer		hides (I/W)	rattle		friction
		certalmid		neat	soft (P)		preserv (I/W)
		vulcanite		neatness	ventila (I/W)		break
		filaments		ugly	pain		rot
		resisting		ugliness (I/W)	painful		tougher
				neater	chafe		leakage
				handsome	chafing		leak
				tasteful	odor		corrosion
				life-like	offensive		corrosive
				resembl (I/W)	rattling		burst
				wrinkle (I/W) (P)	rattles		weak
				embarrassment	clicking		
				ornamentation	creak		
				sight (P)	creaking		

Figure B1. Trait Keyword List. Note: The figure presents the keywords we used to define our traits of interest. The accronymn"I/W" means "in word", which denotes that we use all words that contain the given keyword. The letter "E" means we exclude any word containing that keyword. The letter "P" means that the machine learning algorithm learned to avoid using these keywords beyond the context of prosthetic limbs.

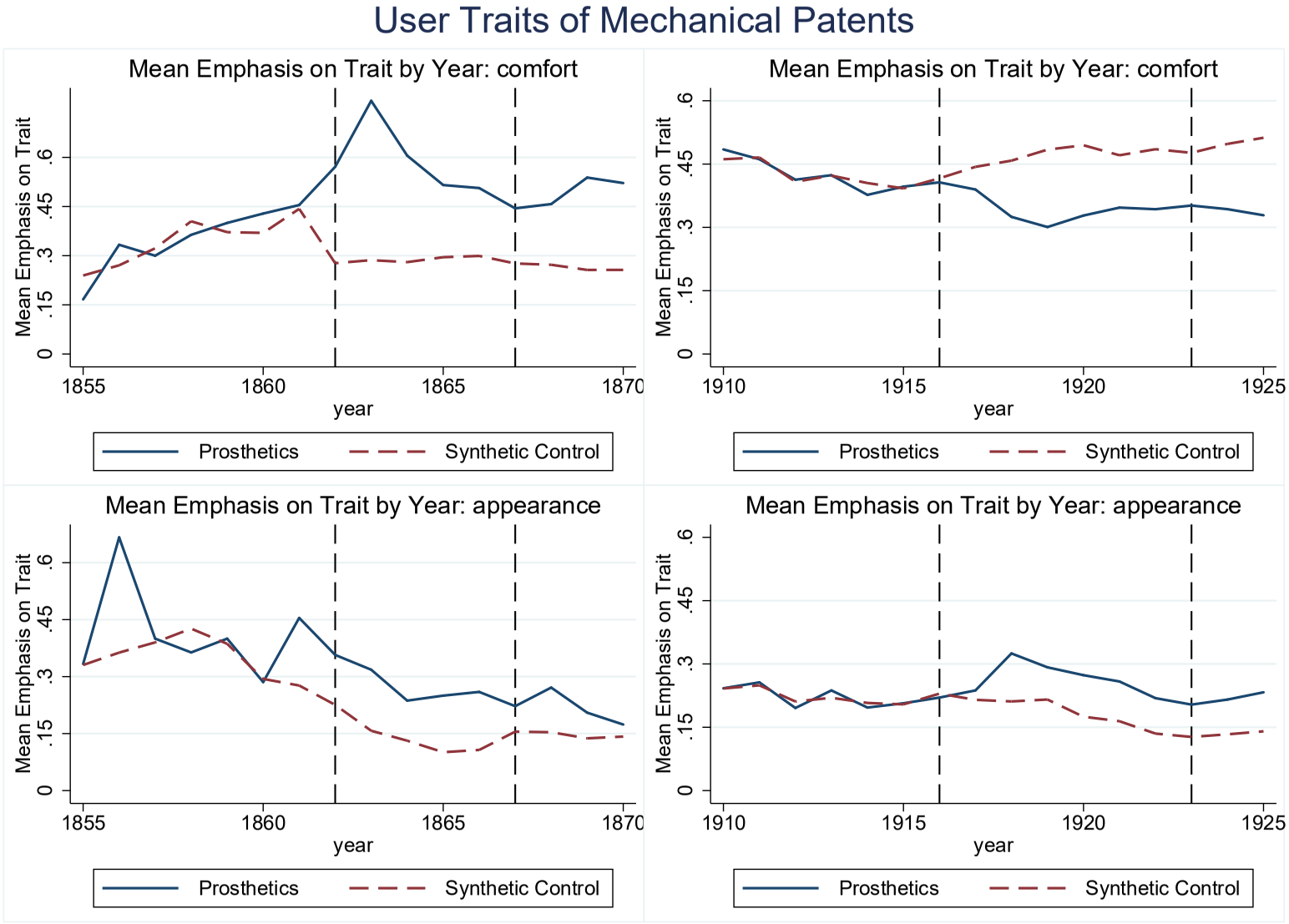


Figure B2. Quality-Oriented Traits: Civil War and World War I Synthetic Controls. Note: The figure presents "treatment" and "synthetic control" series that describe the evolution of patents' emphases on the traits we term "comfort" and "appearance." Further information on the definitions of each trait can be found in table 2.3 as well as in the main text. All series in the figure are calculated as 4-year moving averages. The series plot the share of patents in a given class ("Prosthesis" or the "Synthetic Control") that emphasize a given trait. We generate the synthetic control group using the "synth" package written by Abadie and Hainmueller (2010). For the panels in column 1, "Donor weights" are chosen to match the treatment group on values extending from 1855 to 1861. For the panels in column 2, "Donor weights" are chosen to match the treatment group on values extending from 1910 to 1915.

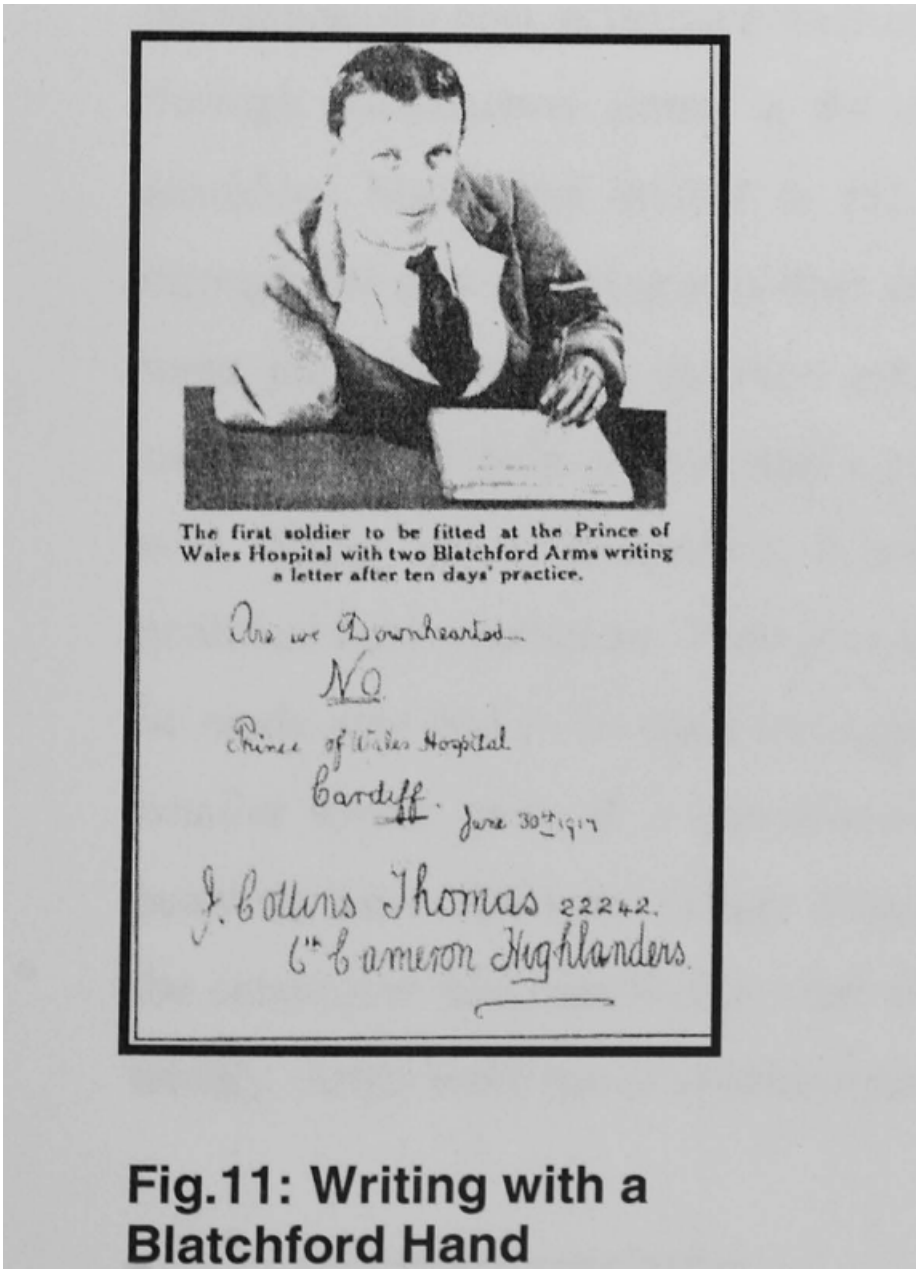


Figure B3. Regaining Writing Ability. Note: The diagram was taken from the article "Enabling the Great War: Ex-Servicemen, the Mixed Economy of Welfare and the Social Construction of Disability, 1899-1930" by Meaghan Melissa Marie Kowalsky. The figure shows an example of a prosthetic arm appliance attachment for writing.

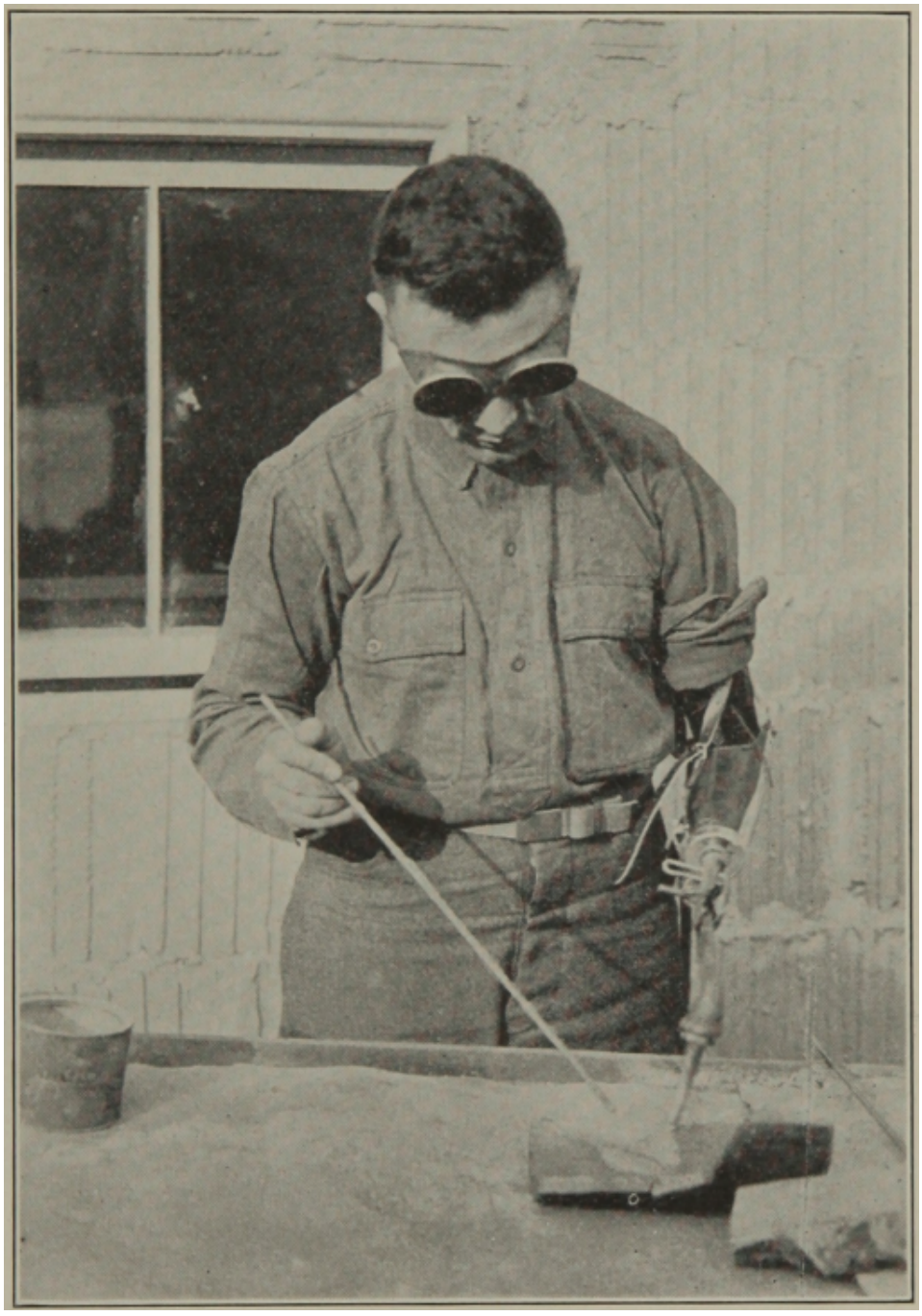


Figure B4. Rehabilitated to Work. Note: The figure shows an example of a prosthetic arm appliance attachment for welding. Source: The United States Army Surgeon General's Office, *The Medical Department of the United States Army in the World War*, Washington, DC: GPO, 1927, volume 13, page 107.

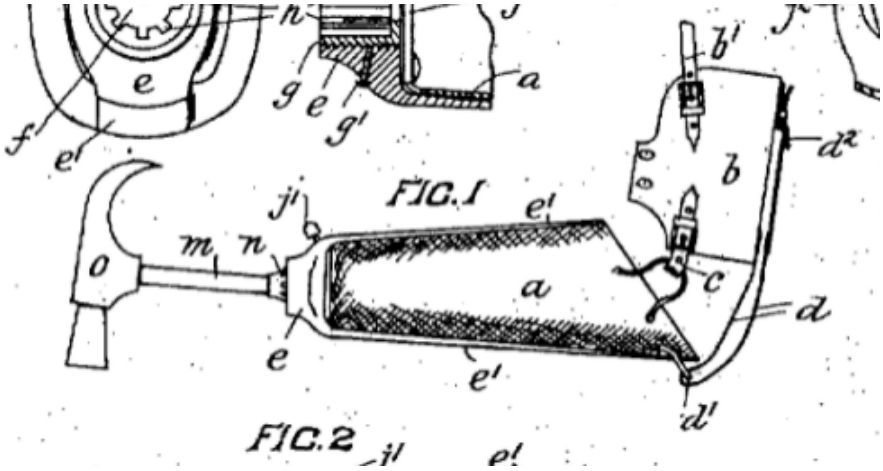


Figure B5. Rehabilitated to Work (Part II). Note: The diagram was taken from U.S. patent 1,213,222 (1917). The figure shows a limb with attachable appliances for use in various trades. This diagram presents an attachable hammer called the "hammer arm." Source: United States Patent and Trademark Office.

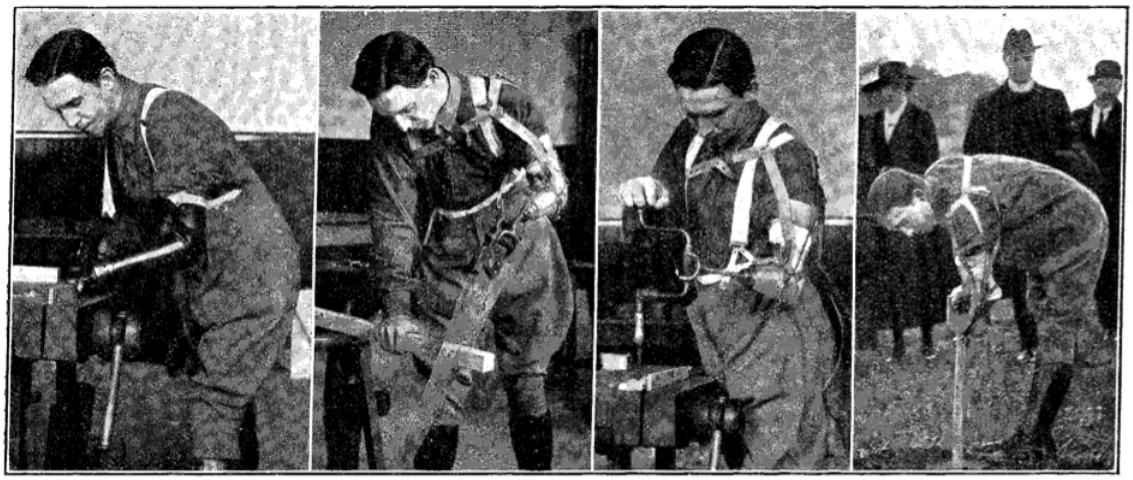
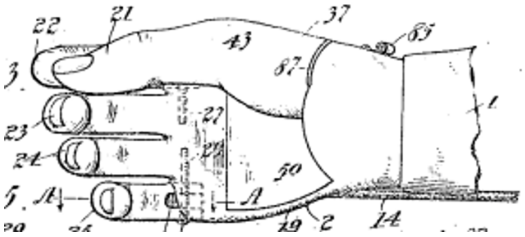
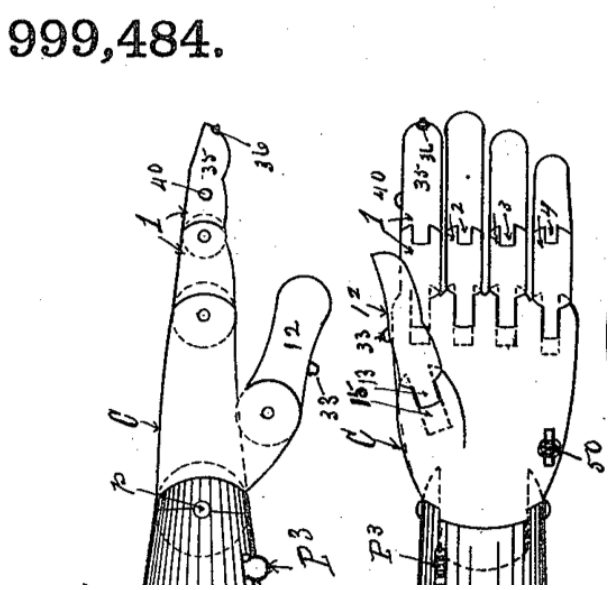


Figure B6. Rehabilitated to Work (Part III). Note: The figure shows interchangeable appliances that equip wearers to perform various trades. Image included with permission from Elsevier: Marshall CJ. Modern artificial limbs: The work of the arm-training centre at Roehampton. Lancet. 25 June 1921.



(a) The Natural Hand



(b) The Carnes Hand

Figure B7. The Natural Hand vs Predominant "Carnes Hand" Note: The top diagram was taken from U.S. patent 1,173,219 (1915), and the bottom diagram was taken from U.S. patent 999,484 (1910). The figure contrasts a more naturally designed hand emphasizing "appearance" (top subfigure) against a more mechanical and modular hand (bottom subfigure). Source: United States Patent and Trademark Office.

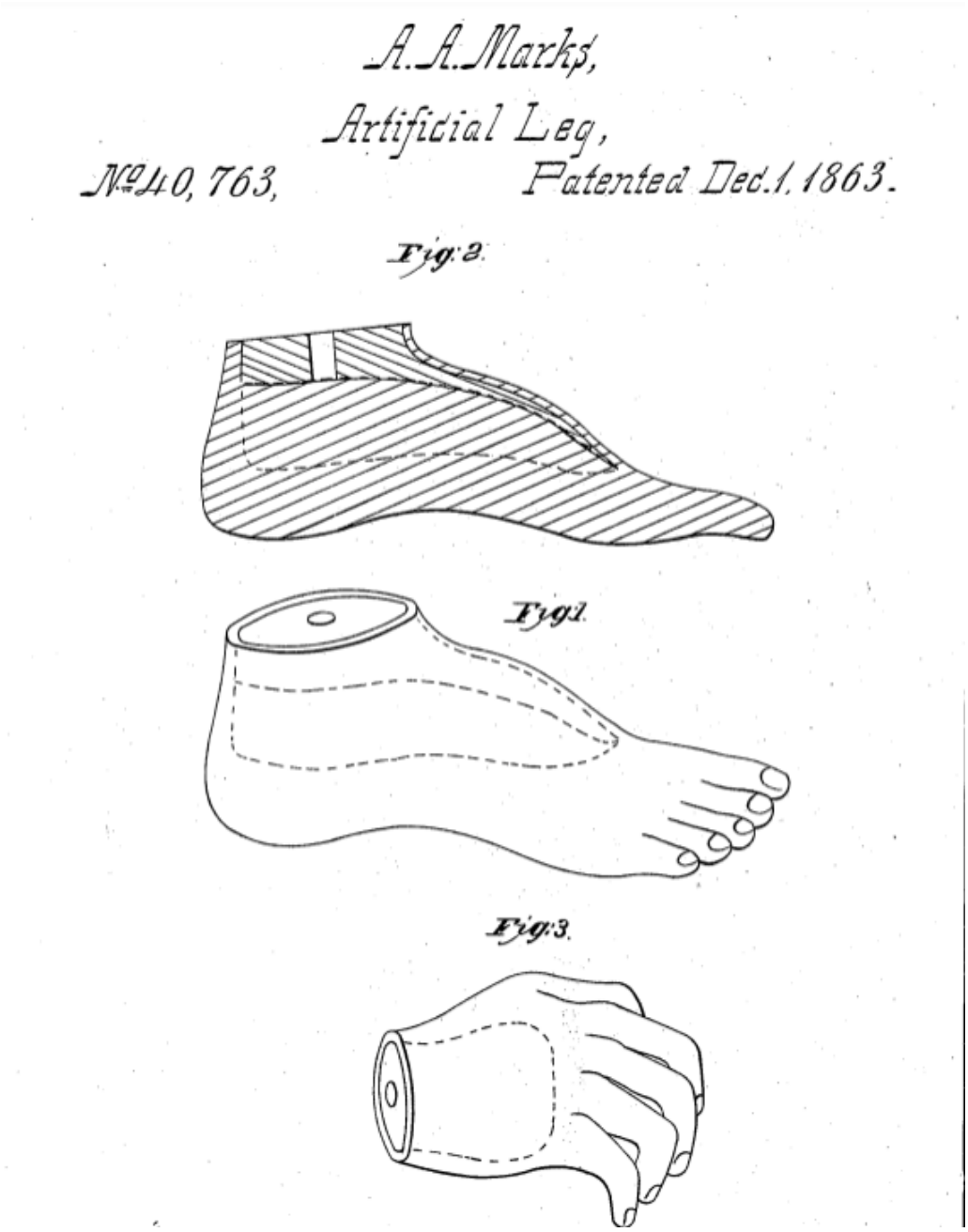


Figure B8. New Cheap Material. Note: The diagram was taken from U.S. patent 40,763 (1863). The figure shows a series of limb pieces constructed from a new, cheap material called vulcanized rubber. This allowed for the cheap construction of a variety of limb components. Source: United States Patent and Trademark Office.

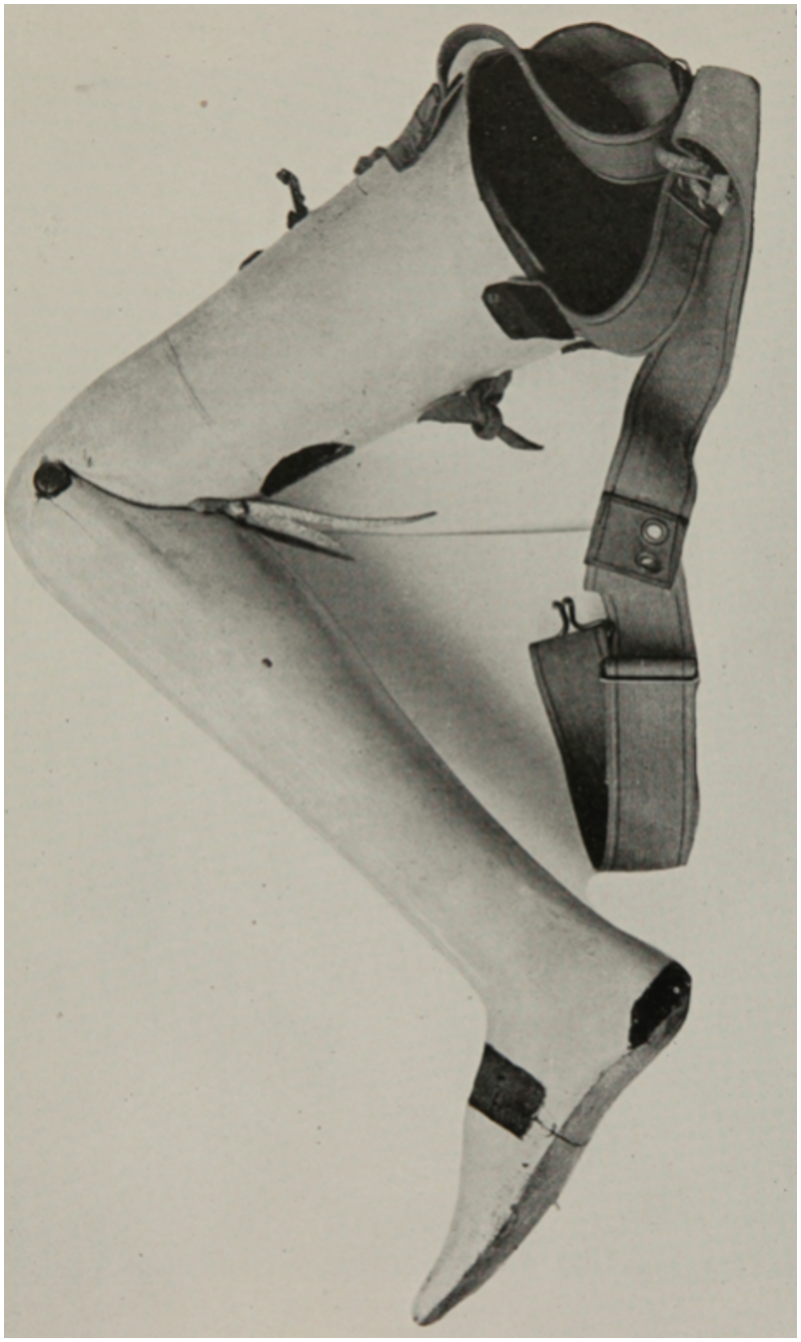


Figure B9. Cheap, Modular, and Life-Like Material. Note: This figure presents the "liberty limb," an artificial leg constructed with a fleshy-colored material and was modular in nature. Source: The United States Army Surgeon General's Office, *The Medical Department of the United States Army in the World War*, Washington, DC: GPO, 1927, volume 11, page 741.

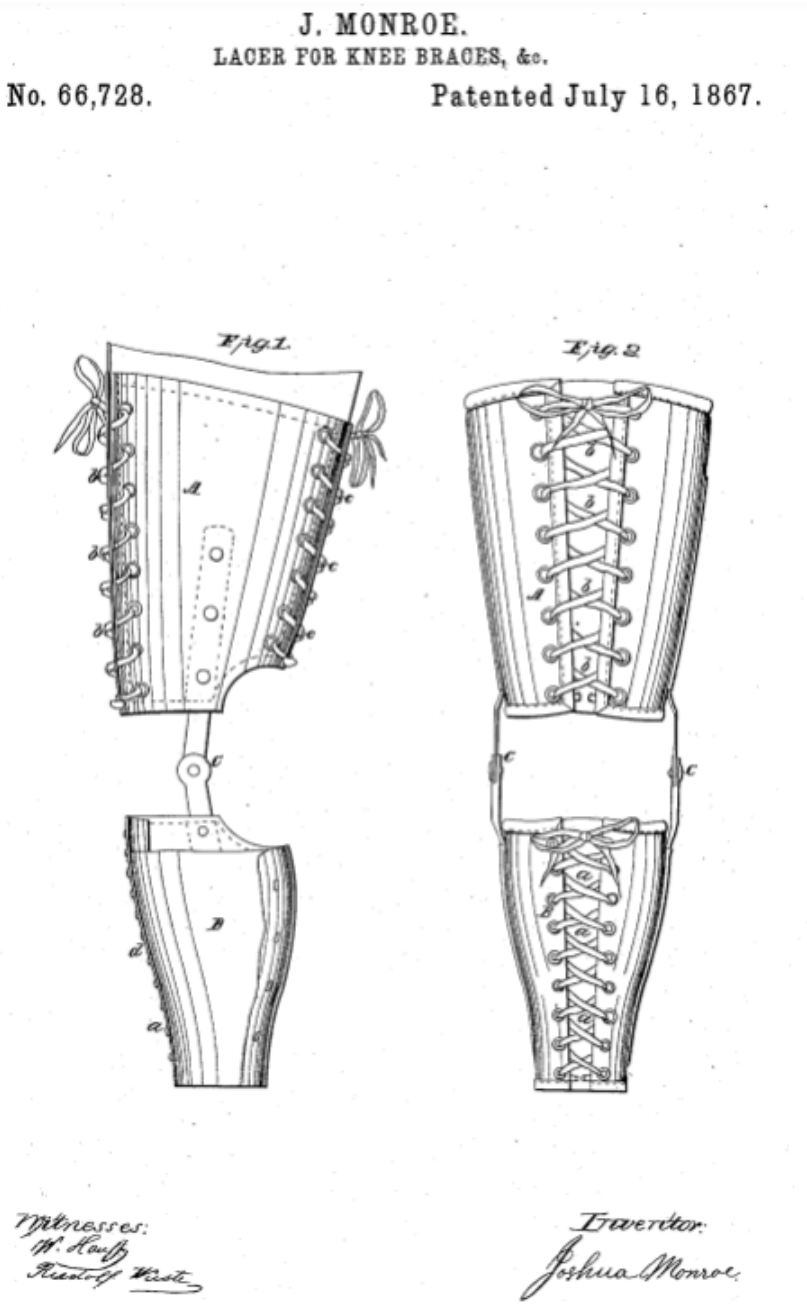


Figure B10. Adjustable Limb. Note: The diagram was taken from U.S. patent 366,728 (1867). The figure shows a lacer device that allows users to adjust knee braces to their unique specifications, lending to cheaper, uniform limb construction. Source: United States Patent and Trademark Office.

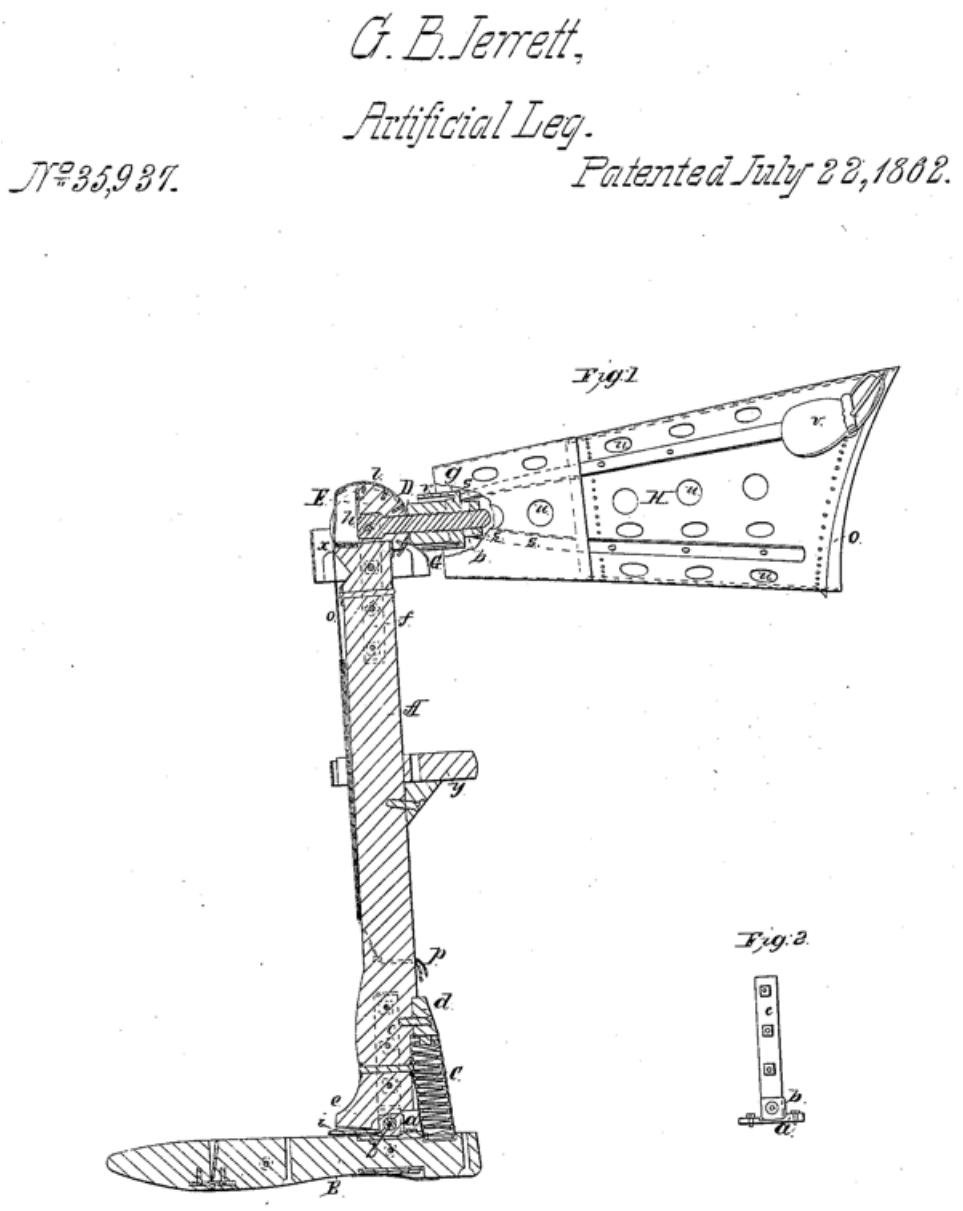


Figure B11. Adjustable Limb II. Note: The diagram was taken from U.S. patent 35,937 (1862). The figure shows an artificial leg with an adjustable height, which relies on an extending spindle in the knee joint. Such a design allows cheaper, uniform construction of limbs, avoiding more expensive, tailored construction. Source: United States Patent and Trademark Office.

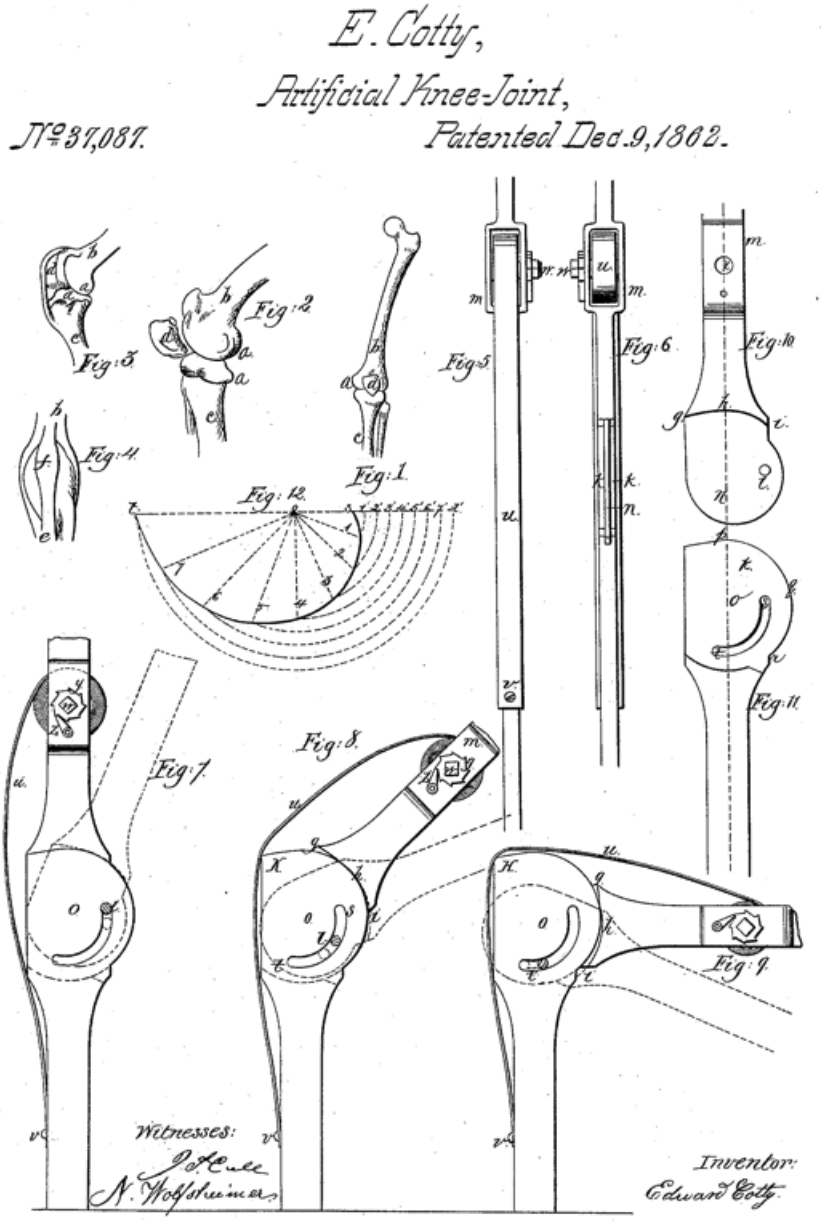


Figure B12. Naturally Simple Limb. Note: The diagram was taken from U.S. patent 37,087 (1862). The figure shows a knee joint constructed of only two primary components, with a simple hinge component at the knee. More complex knee joints, such as the one shown in figure B13, use more intricate mechanisms. Such simplicity allowed for ease of mass production. Source: United States Patent and Trademark Office.

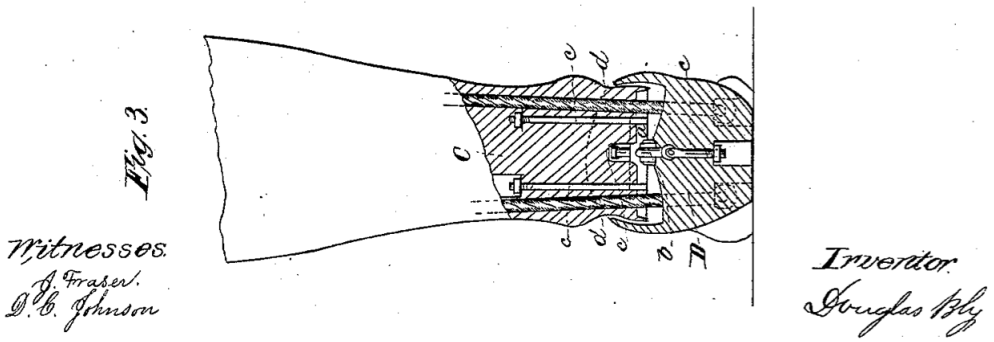


Figure B13. More Complex Knee Joint. Note: The diagram was taken from U.S. patent 38,549 (1863). The figure shows the internal workings of a more complex knee joint invention that emphasized appearance and comfort. Source: United States Patent and Trademark Office.

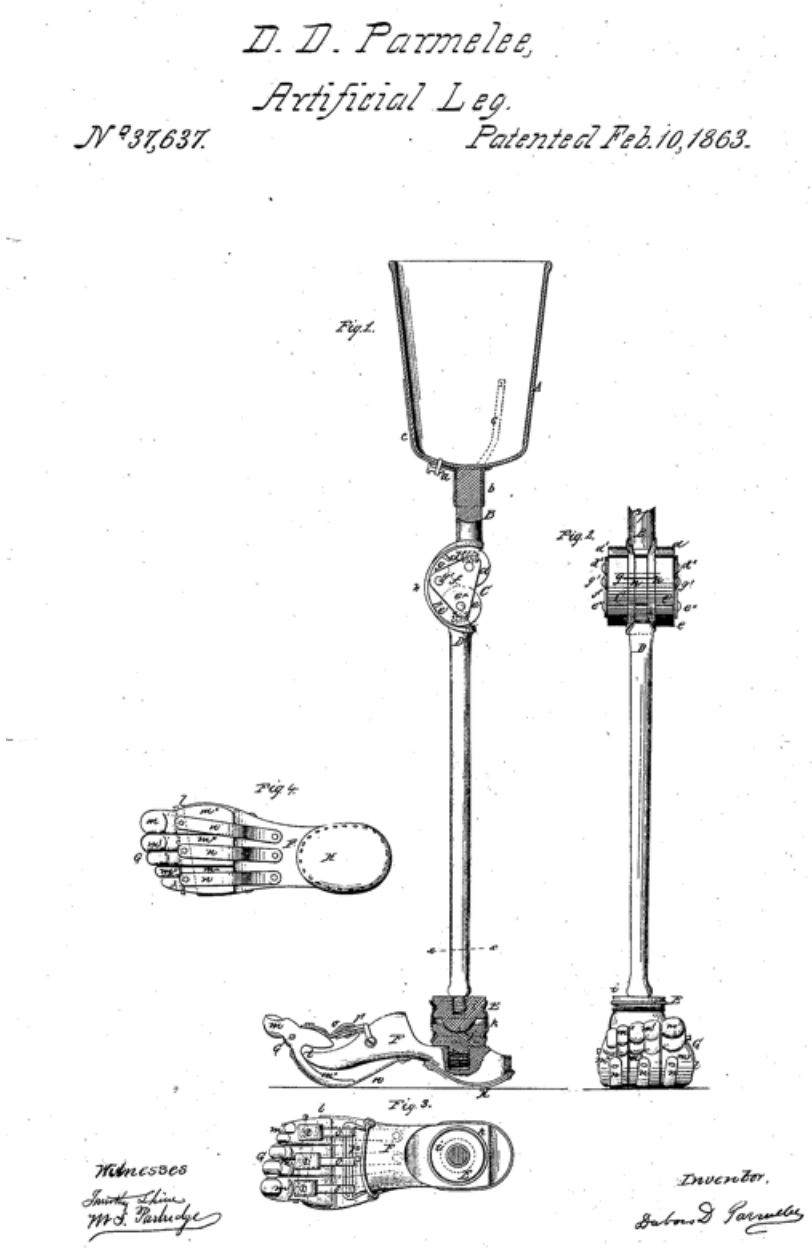


Figure B14. Cheap Bucket Limb. Note: This diagram was taken from U.S. patent 37,637 (1863). The figure shows a unique bucket design for the apparatus into which the stump is inserted. The bucket construction allowed the limb to be adjusted to different user specifications allowing for cheap, uniform construction. Source: United States Patent and Trademark Office.

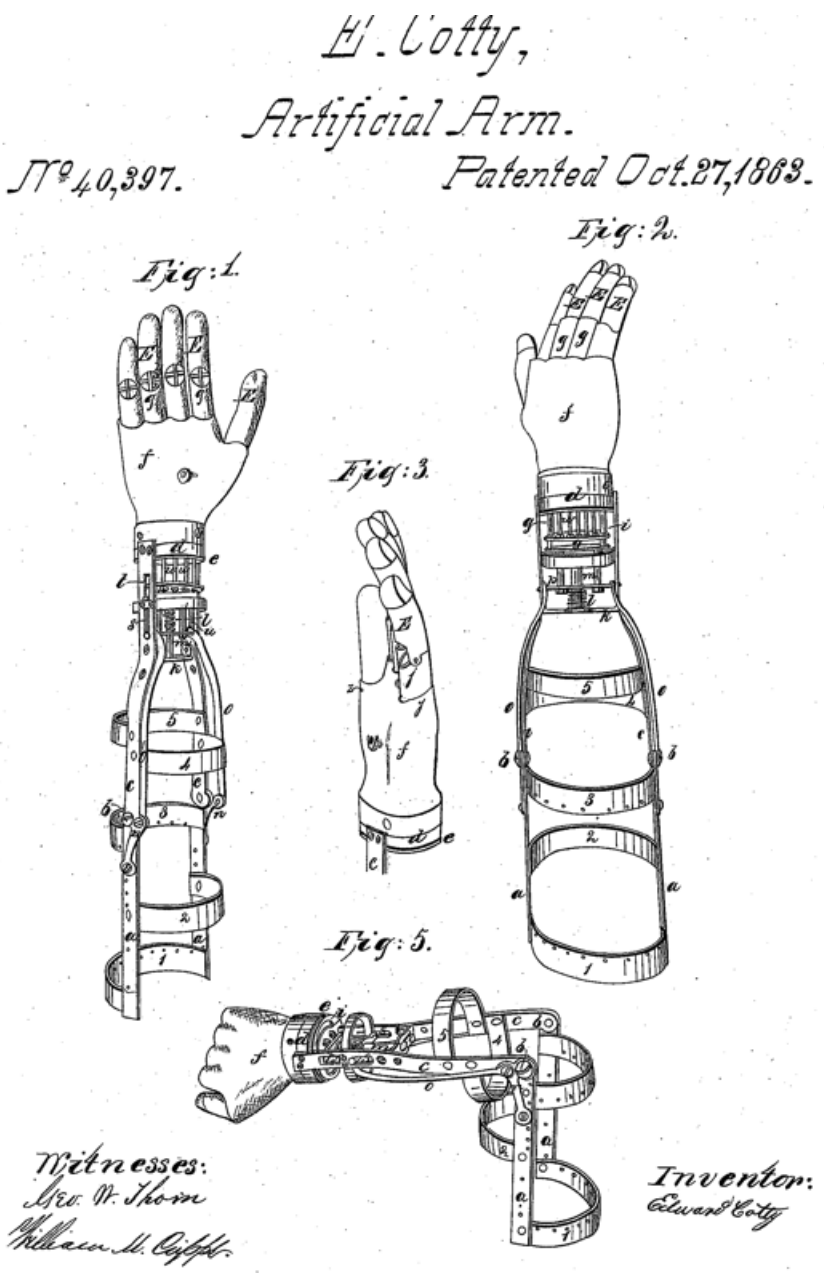


Figure B15. Cheap Metallic Limb. Note: This diagram was taken from U.S. patent 40,397 (1863). The figure shows the use of metal materials when constructing the forearm section of the prosthetic arm in an effort to reduce production costs. Source: United States Patent and Trademark Office.

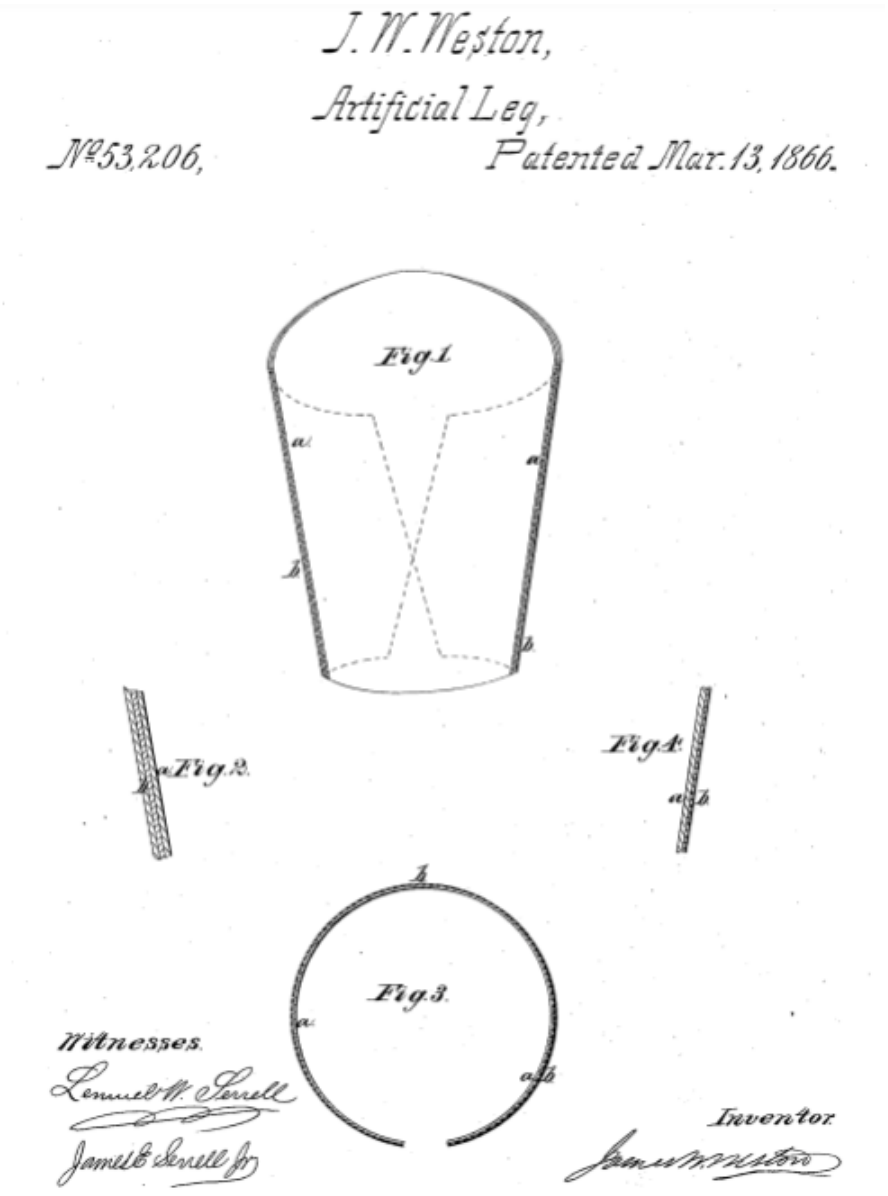


Figure B16. Comfortable Limb Casing. Note: This diagram was taken from U.S. patent 53,206 (1866). The figure shows the construction of a cork limb casing designed to wick away moisture and perspiration. Source: United States Patent and Trademark Office.

B2 Text Analysis Appendix

In this appendix we discuss our approach to designing, evaluating, and selecting our preferred machine learning algorithm for analyzing the texts of patent documents. We begin by describing our objective and comparing our setting with other uses of text analysis in economics research. We then define key terms and discuss examples of the key threats to successful text analysis, along with our approach to addressing them. Finally, we discuss several dimensions of best practice text analysis.

B2.1 Generating Economic Data through Text Analysis

Our goal in conducting text analysis is to create variables that describe the economic content of patent texts. Specifically, we analyze the texts of prosthetic device patents, other medical patents, and mechanical patents to determine whether they emphasize traits we term simplicity, cost, adjustability, materials, comfort, and appearance. We code these traits as binary variables, which are our text analysis outputs.

Our text analysis task shares several key commonalities with recent "sentiment" and "partisanship" analyses, where the objective is to rate the sentiment or the degree of partisanship of a publication, writer, or speaker (Shapiro et al. 2018, Shapiro and Wilson 2019, Garcia 2013, Gentzkow et al. 2019, Gentzkow and Shapiro 2010).[17] Key commonalities are as follows. First, the researcher must either obtain or create a data set containing a set of outputs (the "true values" for the variables of interest) corresponding to a set of text inputs (a subset of the texts of interest). A machine learning algorithm then learns a function, or model, that relates these input-output pairs. Cross-validation is used to evaluate the model's performance by splitting the manually coded input-output pairs into two sets: one on which the model will be trained and another on which the model's performance will be tested. The train-test split is crucial for reliably evaluating

[17] Similarly motivated text analysis exercises have also been used quite recently to study patents. Dechezlepretre et al. (2019), for example, use a keyword search approach to code patents based on whether they relate to "automation." Cockburn et al. (2018) similarly use a keyword search approach to track the advance of artificial intelligence through references within patent texts and journal articles.

performance, as testing on the same data used for training will tend to produce overly optimistic results due to over-fitting.[18] The selected predictive model is then used to assign values for the output variables of interest to the full set of text inputs. Note that these methods are typically used because resource limitations prevent researchers from closely reading and manually coding true values for the broader set of texts. In our case, for example, the broader set of texts consists of more than 700,000 patent documents.

Our preferred algorithm can be described as a modified supervised machine learning algorithm. Our algorithm is somewhat analogous to algorithms used for sentiment analysis by Shapiro et al. (2018). Straightforward algorithms for sentiment analyses make use of "lexicons" that assign positive and negative values to the sentiment associated with extensive lists of words. A simple "Lexical Methodology," for example, is to assign a document a sentiment score based on the sum or mean of the values assigned to the words in its text by the lexicon. In our setting, this is analogous to determining that a patent emphasizes a particular economic trait if its text contains a keyword with which we associate that trait. Shapiro et al. (2018) discuss how this basic approach can be improved upon through tools that account for context (e.g., "negation rules"). While the word "happy" conveys positive sentiment, for example, the phrase "not happy" conveys the opposite. A similar concern motivates the tool we design, which incorporates a neighborhood of contextual clues to root out false-positive errors.

B2.2 The Central Problems of "Polysemy" and "Synonymy"

When using algorithms to extract economic information from text, researchers must overcome errors driven by the complexity of language. In particular, errors can be generated by variations in a word's meanings across contexts and by similarities in the meanings of multiple words. These issues are commonly termed "polysemy" and "synonymy," respectively (Scott Deerwester 1990, Magerman et al. 2011).

[18] Testing on the left-out data gives insight regarding how generalizable a model will be to new data. Further, repeating cross-validation using randomized train-test splits decreases the likelihood that high performance is simply a result of an opportunistic split.

Synonymy (multiple words having the same meaning) can lead to false negatives, as an algorithm may fail to account for words that are similar in meaning to an attribute's most intuitive keywords. By contrast, polysemy (when words have multiple, context-dependent meanings) elicits false positives. If an algorithm does not detect a word's distinct contextual meaning, it may falsely connect a text input with the concept of interest (Turney and Pantel 2010). Polysemy can take multiple forms. In some cases, a word's meaning is straightforwardly negated by the words around it (e.g., the aforementioned difference between "happy" and "not happy"). In other cases, a word's meaning may differ with the subject matter contained in the full text or in a particular sentence (e.g., the meaning of "fork" in the phrases "fork in the road" versus "knife and fork"). The difficulties posed by polysemy and synonymy can be closely related, as a keyword's contextual meaning cannot be learned if the keyword itself is not initially detected.

B2.3 Illustrative Examples from Patent Texts

The attributes we analyze exhibit varying degrees of "polysemy" and "synonymy." The attribute we term "simplicity," for example, was relatively straightforward. This is because the language linked to "simplicity" is relatively common across texts; it is unlikely to have ambiguous meaning or numerous synonyms. One prosthetic device patent, for example, quite explicitly stated that "The object of my invention is to imitate this eccentric motion of the knee-joint in the simplest manner." Another states, "The advantages of my invention are as follows: ... great simplicity, and therefore cheapness." The meaning of simplicity extended quite well to patents in our control classes. One such patent highlights, for example, "that the machinery which we use, as hereinafter described, is simple in construction." The relative ease of classifying simplicity is shown in the high performance, which we define more precisely below, we obtain when training the models we consider. Notably, our preferred model performed quite well in predicting "simplicity" even when the training set contained as few as 100 observations.

By contrast, the attribute we term "comfort" was relatively difficult to work with. Difficulties arose because the language used to indicate a product's "comfort" regularly suffered

from ambiguity. Sometimes, the meaning of comfort was quite clear. A straightforward example from prosthetics states "My present invention has for its object the production of an artificial leg constructed on such principles that it will give more strength and durability to the limb, and also ease and comfort to the wearer." A straightforward true positive from a different mechanical class states that "Until the external pressure becomes too great... air [is] allowed to enter the box A, until the person sitting in it feels comfortable." Difficulties arose, however, from polysemous words used to describe discomfort. For example, the word "disturbing" often connotes bodily discomfort in prosthetic device patents. In mechanical classes, by contrast, the word "disturbing" tends to have meanings connected to the device's functionality (e.g., "disconnecting or disturbing the pump"). The difficulties created by such cases translated into poor predictive accuracy when we attempted to train our preferred model on relatively small training sets.[19]

B2.4 Assessing a Model's Accuracy

A model's accuracy in a binary classification problem can be well described by the evaluation metrics of "sensitivity" and "specificity." Sensitivity refers to the rate of true positives as a share of all positives, while specificity refers to the rate of true negatives as a share of all negatives. These metrics were particularly well suited for our study as they directly ascertain an algorithm's ability to confront the issues of polysemy and synonymy.

Sensitivity and specificity are related. When specificity is reasonably high, sensitivity measures how well an algorithm addresses synonymy by directly revealing the algorithm's ability to correctly detect the desired characteristics: If included keywords inadequately detect patent characteristics due to excluded synonymous keywords, sensitivity would be low. Whereas, when sensitivity is reasonably high, specificity measures the algorithm's ability to ascertain a keyword's context-specific meaning: If the algorithm correctly detects the absence of a given characteristic in the presence of a keyword, it is identifying contextual cues that nullify a keyword's relevance,

[19] As discussed below, comfort is a trait for which accuracy experienced substantial gains as the size of our training data set increased.

causing specificity to increase. If either sensitivity or specificity is very low, however, then the algorithm may arbitrarily assign positive or negative outcomes depending on which outcome occurs most frequently in the training data.

The simple average of sensitivity and specificity is commonly termed the "balanced accuracy score." The balanced accuracy score, averaged across "repeated 10-fold cross-validations," is the criterion we use for model evaluation. We used balanced accuracy, as opposed to other evaluation metrics, as it accounts for class imbalance in the dependent variable—a potential issue common in binary classification tasks.[20] As a rough rule of thumb, we targeted balanced accuracy scores of at least 90 percent.[21] As shown below, however, incremental improvements in an algorithm's accuracy can have meaningful implications for a research project's estimates of primary interest.

We contrast the performance of our preferred model with models generated by a variety of alternative algorithmic techniques. In cases where text classification is well defined by a set of important words, a natural benchmark for assessing alternative tools is a keyword search. A keyword search algorithm codes patents as emphasizing a particular trait if the document contains any words that are strong markers for the trait. As highlighted below, a keyword search is highly effective at identifying positive outcomes for tasks like ours. It may produce false positives, however, by ignoring contextual cues that nullify a keyword's relevance. Whether this shortcoming outweighs a keyword search's ability to detect positive outcomes depends on the degree of polysemy in a researcher's particular task.

[20] In the context of a binary classification problem, class "imbalance" means that there are more/fewer negative outcomes compared to positive outcomes. See Brodersen et al. (2010) for a widely cited discussion of the balanced accuracy score's attractive properties in settings where this holds.

[21] Another common measure of model performance in binary classification tasks is AUC, the area under the receiver operating characteristic curve. For our "comfort" trait we achieve an AUC score of 0.92 and for our "simplicity" variable we attain an AUC score of 0.95. These scores are quite high, suggesting that positive and negative outcomes are quite distinctly separated as the majority of outcomes are simply determined by the presence of a keyword.

B2.5 Our Preferred Algorithm: A Novel Modified ML Approach

We considered several classes of algorithms as potential tools for constructing our data set. These included "unsupervised" machine learning algorithms, "supervised" machine learning algorithms, modified supervised learning algorithms, and simple keyword searches. Our preferred algorithm can be described as a modified supervised learning algorithm. The key modification, which involves constraining the feature space from which the algorithm learns, generated advantages with respect to both accuracy and computing requirements.

Unsupervised learning tools are meant to form meaningful groupings of input data based on some predefined metric (Athey 2018). In our context, we found that such tools struggled to form groupings that coalesced around the economic attributes we sought to analyze. This problem cannot be resolved through the analysis of larger samples.

Standard supervised machine learning tools take as inputs a feature space generated from the entirety of each document's text. We find that these tools struggled to overcome the problems of synonymy and polysemy.[22] For supervised machine learning tools, we find that the performance of existing algorithms improved, to varying degrees, as we expanded the size of our training set. It is thus possible that these algorithms would reach tolerable accuracy thresholds on training samples of sufficient size. Our analysis is suggestive, however, that generating training samples of sufficient size may be beyond many research projects' scope. Closely reading thousands of patent texts or other context-relevant documents is a resource-intensive process.

We find that simple keyword searches performed quite well in our setting. Notably, the development of our lists of keywords benefited from our experimentation with machine learning. In our project's early stages, we attempted keyword searches based on a combination of intuition

[22] This may stem from the fact that even after processing the text data (removing stop words, word fragments, etc.), the full sample of patent texts contained over 18,000 features. In a simulation analysis using synthetic data, Hua et al. (2004) simulate error rates across alternative feature space sizes, sample sizes, and algorithms. In their context, they find that the optimal feature size is $N-1$ for uncorrelated features (where N is the sample size) and that the optimal feature size becomes proportional to $\sqrt{N}$ for highly correlated features. Although these findings are not necessarily generalizable, in our case the number of features (when using the full processed patent texts) was $15N$, suggesting that the relatively high number of features is plausibly linked to suboptimal performance.

and close readings of a small set of patents. This "procedure" performed poorly. The accuracy of our keyword searches increased substantially as we learned more about our domain through close readings of 1,200 patent documents in total. Success with either keyword searches or our modified machine learning approach will tend to require substantial knowledge of the domain one is attempting to analyze.[23] Both sets of approaches provide ample evidence of the idiom "garbage in, garbage out."

Although keyword searches ultimately performed quite well for our task, their general limitations are worth emphasizing. A keyword search does not, by construction, allow context to inform a word's meaning. This can lead to false-positive errors. In general, it should thus be possible to improve upon keyword searches by allowing contextual clues to inform a word's true meaning within each text.

Our preferred, modified approach connects the knowledge we obtained reading patent documents to the Gradient Boosted Machines algorithm (Friedman 2001).[24] When constructing this model we directly targeted the issues of synonymy and polysemy. First, while reading 1,200 patent documents, we compiled a non-comprehensive list of keywords that indicate each characteristic. To gather each keyword's synonyms, we mapped all our considered patent text corpora to a vector space.[25] This allows us to model the degree of contextual similarity between words using spatial word proximity, resulting in spatial groupings of keywords and their most relevant synonyms. After adding keywords and their synonyms into the feature space, we then include a flexible neighborhood of text surrounding these words to provide contextualization.[26]

[23]The success of our modified machine learning tool depended on a combination of manually gathered keywords through close readings and data-driven synonym determination. Although this form of feature selection required extensive domain knowledge, feature selection can be effectively executed using entirely data-driven algorithms (see Guyon et al. (2002) and Guyon and Elisseeff (2003)). In our case, however, these purely data-driven approaches selected features that induced worse performance than simply using the full patent text. Accuracy gains only occurred when we used a combination of hand-picked and data-driven feature selection.

[24]This is a "boosted" version of Random Forests (Breiman 2001) where error terms from previous decision tree predictions inform the construction of subsequent trees.

[25]We use Word2Vec (Mikolov et al. 2013) to construct these word embeddings. Word2vec uses shallow neural networks to map words within text documents to a vector space that captures word relationships through a distance metric. Words within this space are mapped as being close together if they occur in similar contexts in the text corpora.

[26]These steps are well described as a type of "feature selection." Feature selection has been shown to help at

We then train the machine learning algorithm with this reduced feature space to obtain more accurate and efficient results.[27]

Relative to alternative machine learning methods, our modified approach generated accuracy gains when predicting each of our economic characteristics. Improvements relative to machine learning approaches that attempt to learn from the entirety of each patent's text were quite large. The relative success of our modified approach, when compared to other pure machine learning methods, is driven by the amount of extraneous information in patents' full texts, figure descriptions, and detailed claims. The presence of extraneous features reduced these algorithms' ability to pinpoint specific, economically relevant patent characteristics. Constraining the feature space to include only keywords, their synonyms, and neighboring contexts allows the machine learning algorithm to learn more efficiently.

Relative to a keyword search, our algorithm's greatest improvements in accuracy were gains of three percentage points for the quality-oriented traits we term "comfort" and "appearance." The improvement in accuracy comes entirely from gains in specificity: The modified approach learns to discriminate keywords whose context nullifies their meaning. Although a three percentage point gain in accuracy is modest, researchers will tend to realize larger gains for text analysis problems with greater degrees of polysemy.

B2.6 Lessons for Implementing Best Practice Text Analysis

In this section, we illustrate several key inputs to best practice text analysis. While text analysis tasks necessarily confront many setting-specific challenges, the dimensions of best practice we discuss should apply quite generally. They include an approach for assessing the optimal size of a training set, the importance of generating a training set that covers all contexts that a researcher targets, and an approach for assessing the implications of inaccurate predictions

"improving the prediction performance of the predictors, providing faster and more cost-effective predictors, and providing a better understanding of the underlying process that generated the data" (Guyon and Elisseeff 2003),

[27] Computation time was dramatically reduced using our approach when compared to other machine learning algorithms. This stems from the reduced feature space, allowing quicker model training.

for the estimates in which a study is ultimately interested.

Determining Optimal Sample Size

We conducted a systematic analysis of how the performance of various algorithms evolved as we expanded the size of our training data set. Text analysis tasks may differ substantially with respect to the complexity of each piece of text and with respect to the severity of setting-specific sources of polysemy and synonymy. Consequently, it is not possible to prescribe a "rule-of-thumb" size for a training set. One can nonetheless use the relationship between accuracy and sample size to make inferences regarding the returns to further expansions of the training set.

Using our preferred modified approach, the size of the training set required to reach tolerable balanced accuracy scores varied across traits. For the trait we term simplicity, for example, our balanced accuracy score exceeded 90 percent with training sets containing fewer than 200 observations. For the trait we term comfort, by contrast, the accuracy score approached 90 percent as training sets contained roughly 700 observations. For the trait we term materials, the accuracy score remained below 90 percent even on our full training set of 1,200 observations.

On what basis should the size of the training set be determined? Expanding a training set requires project resources. On the margin, the key question is whether increases in the size of the training set yield non-trivial returns. As a way to gauge the relevant returns, we recommend constructing "learning curves," like those displayed in Figure B19. We constructed these figures by evaluating our model's accuracy when trained and tested on samples of varying sizes. More specifically, we executed a bootstrap estimation of our model's balanced accuracy score when trained on different sample sizes from our manually coded data, with the remaining un-sampled data forming the test set. The solid green line in each panel traces the mean of the balanced accuracy score across 400 iterations of this procedure at ascending sample sizes. The shaded green area extends from the 10th to the 90th percentiles of the distribution of results. The bootstrap approach assures that our estimate for any given sample size is not skewed by particularly "favorable" or "unfavorable" draws, meaning draws on which the algorithm happens

to have a particularly easy or difficult time with its prediction task.

Panel A of Figure B19 shows that the balanced accuracy score for "comfort" is relatively low with small samples. Further, the score for comfort exhibits non-trivial improvement as the training set expands to include as many as 1,000 patents. The band extending from the 10th to the 90th percentiles of the distribution is quite large in comparison with the band presented in panel B, for the trait we term simplicity.

Panel B of Figure B19 shows that the balanced accuracy score for "simplicity" is high with small samples. Further, the score asymptotes quickly. It exhibits no further improvement once the training set includes 400 observations. Notably, the band extending from the 10th to the 90th percentiles of the distribution is relatively tight. This further supports the point that the performance of the algorithm is not particularly dependent on the patent documents used to train it.

Our analysis of alternative machine learning algorithms provides additional evidence that performance can depend crucially on sample size. On samples of the sizes we consider, we found that non-neural network machine learning algorithms perform better than deep learning algorithms and that our modified machine learning approach performs better than both deep learning and non-neural network machine learning models trained on the entire text of each patent.[28]

Assessing the Stability of Economic Estimates

What constitutes an acceptable accuracy threshold? Alternatively, how can one gauge the implications of incremental changes in model accuracy for the primary estimates of an analysis? We shed light on this question through a simulation of how our estimates evolve as we systematically *reduce* the accuracy of our preferred algorithm's estimates.

[28]These results are fairly consistent across the economic traits we analyze. All machine learning hyper-parameters are tuned using randomized grid-search methods (Bergstra and Bengio 2012). Deep learning models we considered were Bidirectional Encoder Representations from Transformers (Devlin et al. 2018), Convolutional Neural Networks (Kim 2014), Recurrent Neural Networks with long short-term memory (Hochreiter and Schmidhuber 1997), and Multi-Layer Perceptrons (Rosenblatt 1961).

The procedure we conduct is straightforward. Starting with the data generated by our preferred modified approach, we inject noise by altering the coding of a given fraction of the observations for an outcome variable of interest. We do this for fractions ranging from 1 percent to 50 percent. We select the observations we miscode at random, then estimate β_1 from equation (2.4.4). As in our analysis of "learning curves," we implement a bootstrap-style procedure. That is, for each degree of noise, we repeat the basic procedure 40 times to generate a range of new estimates. Figure B20 reports the resulting means and distributions.[29]

Panel A of Figure B20 presents estimates for the trait we term "comfort" during the World War I period. Our baseline estimate for comfort is -0.14, indicating that wartime prosthetic device patents were 14 percentage points less likely than pre-war prosthetic device patents (net of the equivalent change for the synthetic control group) to emphasize comfort. As we reduce the accuracy of our comfort variable's coding, this estimate quite rapidly converges towards zero. The magnitude of the estimate for comfort was halved before we had reduced accuracy by 10%.[30]

Panel B of Figure B20 presents the sensitivity of estimates of β_1 from equation (2.4.4) for "simplicity." Our baseline estimate for simplicity is 0.13, indicating that wartime prosthetic device patents were 13 percentage points more likely than pre-war prosthetic device patents (net of the equivalent change for the synthetic control group) to emphasize simplicity. Interestingly, the rate of convergence to zero differs non-trivially when comparing the estimates for comfort and simplicity. Estimates for simplicity converge more slowly, as the magnitude of the estimate is halved when we had reduced accuracy by roughly 20%.

Coding accuracy is clearly important for generating unbiased estimates in analyses of both comfort and simplicity. In both cases, 20% reductions in accuracy would render the

[29]Note that the estimate we produce using the data generated from our preferred model serves as the benchmark. Since our modified approach does not predict with perfect accuracy, the current observations already have a small amount of measurement error corresponding to the error associated with the model's performance in predicting "comfort."

[30]As the accuracy of the data approaches 50%, the estimate converges to zero. As the algorithm's accuracy dips below 50% the estimate will begin to converge to the opposite sign of the true estimate. To see why note that altering the coding of 100% of the observations would yield a variable that is the inverse of the original variable.

estimates from our analyses much smaller economically. In addition to being economically smaller, the attenuated estimates are less likely to be statistically distinguishable from zero. Differences in the rate of convergence towards zero suggest that the tolerability of error may be higher in the case of simplicity than in the case of comfort. It is not obvious why this is the case. A natural hypothesis, into which more research is needed, is that estimates' sensitivity to reductions in accuracy may depend in part on a trait's baseline prevalence within both the treatment and control groups.

Context Specificity

The performance of a trained model may be limited outside the context of its training data. We term this concept "context specificity." Limitations on a model's validity outside of its training set can result from variations in word meanings and usage across domains and across time. In our case, a model trained to recognize the traits in artificial limb patents may perform poorly when applied to patents from classes we use as controls. A model's performance might be impaired if the training set lacks sufficient data from all considered domains.

To illustrate this point, we conduct the following exercise. Our data can be described as consisting of four contexts, namely Civil War-era prosthetic devices, Civil War-era control categories, World War I-era prosthetic devices, and World War I-era control categories. We train our model on a single context, then asses its accuracy in all four contexts. Doing this for each of the contexts separately generates a total of sixteen balanced accuracy scores, four of which involve applying the model to the context on which it was trained. To ensure that differences in accuracy scores across contexts are not driven by differences in sample size, we constrain the size of the training set to be equal in all cases.

The results of conducting this exercise for our "comfort" and "simplicity" traits can be found in Table B25. In each panel, the main diagonal of the matrix of balanced accuracy scores corresponds to our model being applied to the context on which it is trained. This is done using cross-validation within the given domain and time period. The antidiagonal entries correspond

to our model being trained on a different patent class (prosthetic devices vs. the control classes) and historical episode (Civil War vs. World War I) than the corresponding left-out test data set. Differences in the average value of the balanced accuracy scores along the main diagonal relative to the antidiagonal provide information on the relevance of context-specificity.

Consistent with our priors, we find that context-specificity is more important for traits for which the problems of polysemy and synonymy are relatively severe. In the examples presented in Table B25, we find that the difference in accuracy scores when comparing the main diagonal to the antidiagonal is greater for "comfort" than it is for "simplicity." The differences in accuracy scores for comfort are non-trivial. On average, the score along the main diagonal is 92.5 percent, while the average score along the antidiagonal is 86.5. The difference of 7 percentage points is non-trivial when put in the context of our analysis from the previous section. For comfort, injecting a 7 percentage point reduction in accuracy led our estimate of β_1 from equation (2.4.4) to decline by nearly half.

More generally, we find that it is important to account for context specificity when predicting attributes whose meaning is domain- and time-dependent. In our setting, attributes that exhibited this time- and domain-dependence include "appearance", "materials", and "comfort." By contrast, accuracy scores were relatively insensitive to the training set's context for the traits we term "cost," "simplicity," and "adjustability."

Acknowledging Limitations

In some cases, even a well-chosen algorithm trained using a large data set may yield low accuracy scores. Even with our preferred algorithm, for example, we obtained an accuracy score of 87 percent when predicting the trait we term materials. What drives this result and how should it shape our presentation of the evidence?

"Materials" was a difficult trait to predict because keywords that describe the introduction of novel materials tend to have no previous mentions. When few observations contain a keyword, an algorithm's opportunities to learn how best to classify out-of-sample observations with that

keyword are limited. Keywords that were consistently used to describe new materials—like material, alloy, chemical, composition, or mixture—also tended to be used in the description of a device's construction whether or not the associated materials were new. Further, new material innovations were relatively rare. They occurred in only six percent of the observations in our sample, resulting in a small number of reliable positive observations.

As shown earlier, reductions in model accuracy tend to attenuate our estimates. Properly interpreting our estimates thus requires knowing the accuracy of the model used to generate the dependent variable. We recommend presenting two key pieces of information. First, analyses of this sort should present readers with an accuracy metric that is appropriate to the setting.[31] In Table B26, for example, we present the full set of balanced accuracy scores along with the underlying sensitivity and specificity scores. Second, "stability curves" of the sort we present in section B2.6 provide valuable information for inferring the biases associated with inaccurate predictions. We thus recommend coupling these key pieces of information within a discussion of the implications of prediction errors.

In some cases, predictive accuracy may be sufficiently low that the resulting biases will lead point estimates to be highly misleading. In such cases, we recommend that readers be directly warned to interpret the estimates "with caution." In some cases, it may be possible to pair this caution with the best estimate of the potential magnitude of the associated bias. If the only bias is a straightforward form of attenuation bias, then interpretable estimates can be recovered by applying a correction factor. If a correction factor cannot be estimated, the best approach may be to describe estimates as being useful for "illustrative purposes" only.

[31] While the balanced accuracy score is a sensible metric for our setting, alternative metrics might be more suitable elsewhere.

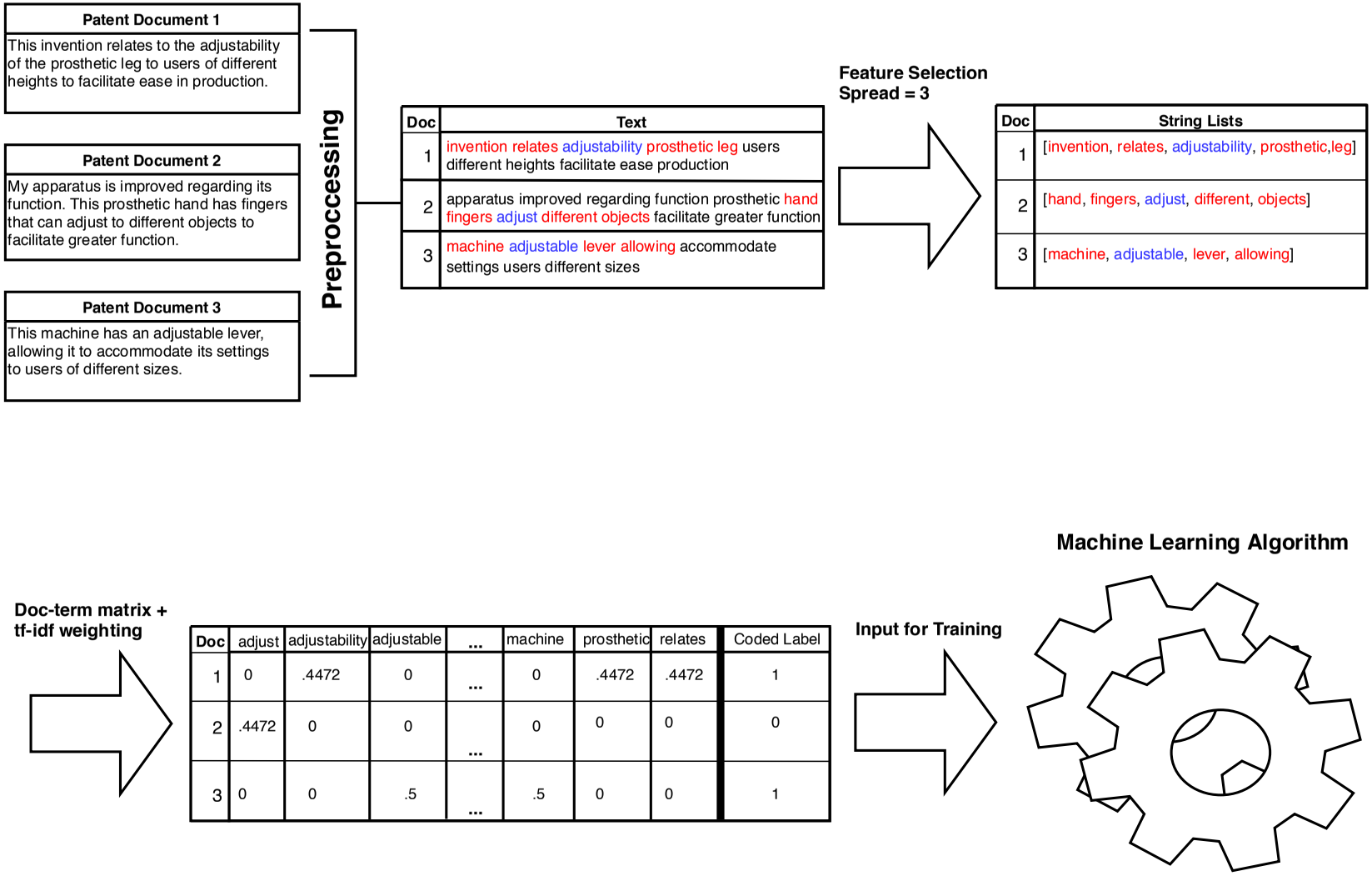
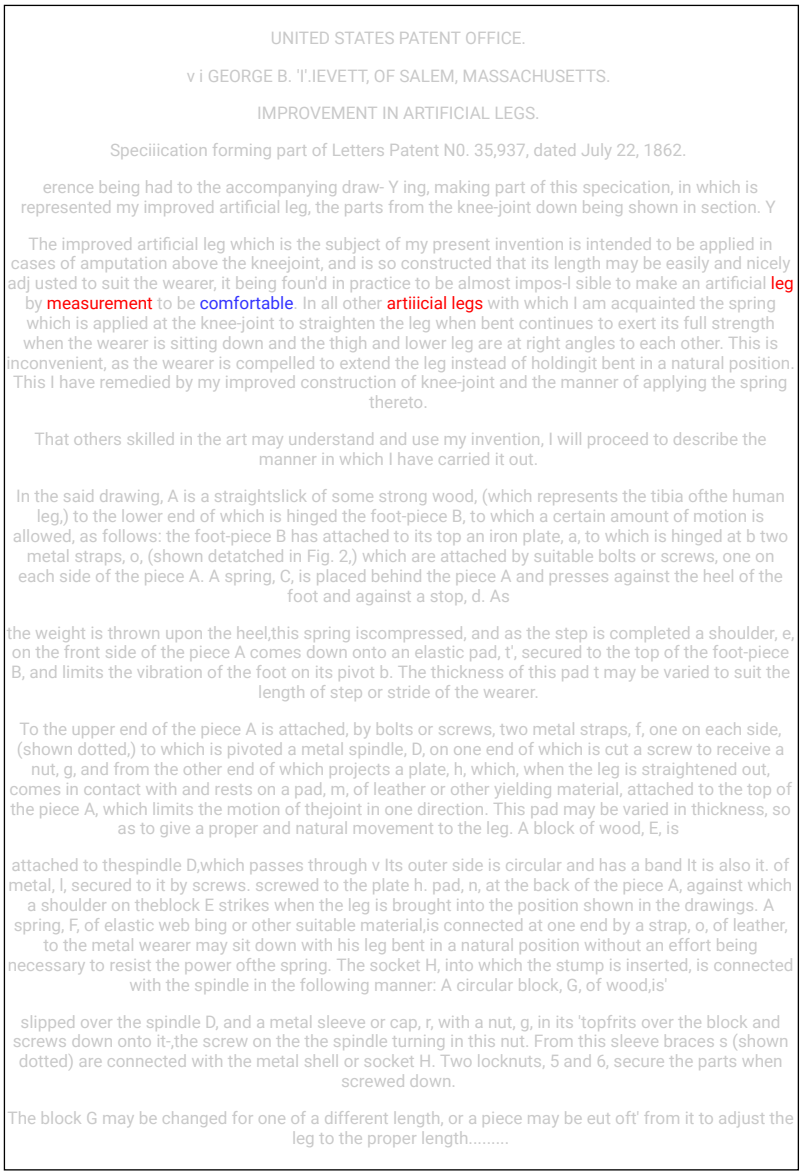


Figure B17. Flowchart of Modified Approach for Adjustability Characteristic. Note: The figure presents a flowchart of our modified approach. First, the text documents are preprocessed by correcting spelling errors, setting characters to lowercase, removing stop words, punctuation, word fragments, numbers, and extremely frequent or rare words. Then we select keywords and their surrounding context as features. After, we create a doc-term matrix with each entry representing the tf-idf weighting of relative importance. Lastly, this doc-term matrix is fed into the machine learning algorithm for training.



UNITED STATES PATENT OFFICE.

v i GEORGE B. 'I'.IEVETT, OF SALEM, MASSACHUSETTS.

IMPROVEMENT IN ARTIFICIAL LEGS.

Speciiication forming part of Letters Patent N0. 35,937, dated July 22, 1862.

erence being had to the accompanying draw- Y ing, making part of this specication, in which is represented my improved artificial leg, the parts from the knee-joint down being shown in section. Y

The improved artificial leg which is the subject of my present invention is intended to be applied in cases of amputation above the kneejoint, and is so constructed that its length may be easily and nicely adj usted to suit the wearer, it being foun'd in practice to be almost impos-l sible to make an artificial leg by measurement to be comfortable. In all other artiiicial legs with which I am acquainted the spring which is applied at the knee-joint to straighten the leg when bent continues to exert its full strength when the wearer is sitting down and the thigh and lower leg are at right angles to each other. This is inconvenient, as the wearer is compelled to extend the leg instead of holdingit bent in a natural position. This I have remedied by my improved construction of knee-joint and the manner of applying the spring thereto.

That others skilled in the art may understand and use my invention, I will proceed to describe the manner in which I have carried it out.

In the said drawing, A is a straightslick of some strong wood, (which represents the tibia ofthe human leg,) to the lower end of which is hinged the foot-piece B, to which a certain amount of motion is allowed, as follows: the foot-piece B has attached to its top an iron plate, a, to which is hinged at b two metal straps, o, (shown detatched in Fig. 2,) which are attached by suitable bolts or screws, one on each side of the piece A. A spring, C, is placed behind the piece A and presses against the heel of the foot and against a stop, d. As

the weight is thrown upon the heel,this spring iscompressed, and as the step is completed a shoulder, e, on the front side of the piece A comes down onto an elastic pad, t', secured to the top of the foot-piece B, and limits the vibration of the foot on its pivot b. The thickness of this pad t may be varied to suit the length of step or stride of the wearer.

To the upper end of the piece A is attached, by bolts or screws, two metal straps, f, one on each side, (shown dotted,) to which is pivoted a metal spindle, D, on one end of which is cut a screw to receive a nut, g, and from the other end of which projects a plate, h, which, when the leg is straightened out, comes in contact with and rests on a pad, m, of leather or other yielding material, attached to the top of the piece A, which limits the motion of thejoint in one direction. This pad may be varied in thickness, so as to give a proper and natural movement to the leg. A block of wood, E, is

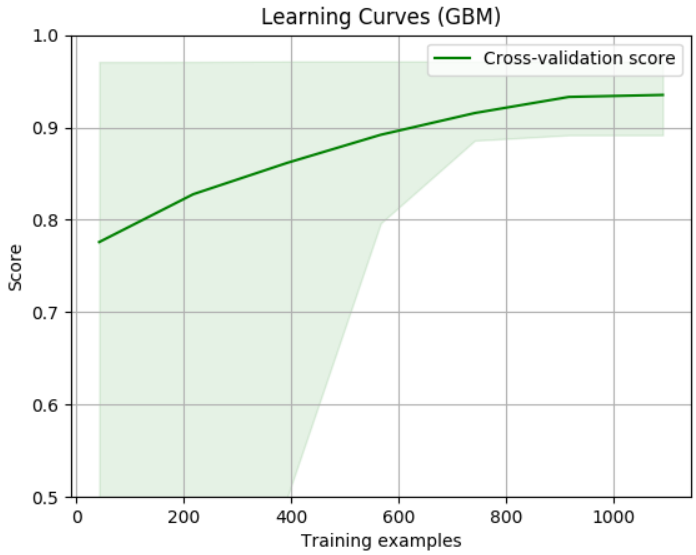
attached to thespindle D,which passes through v Its outer side is circular and has a band It is also it. of metal, l, secured to it by screws. screwed to the plate h. pad, n, at the back of the piece A, against which a shoulder on theblock E strikes when the leg is brought into the position shown in the drawings. A spring, F, of elastic web bing or other suitable material,is connected at one end by a strap, o, of leather, to the metal wearer may sit down with his leg bent in a natural position without an effort being necessary to resist the power ofthe spring. The socket H, into which the stump is inserted, is connected with the spindle in the following manner: A circular block, G, of wood,is'

slipped over the spindle D, and a metal sleeve or cap, r, with a nut, g, in its 'topfrits over the block and screws down onto it-,the screw on the the spindle turning in this nut. From this sleeve braces s (shown dotted) are connected with the metal shell or socket H. Two locknuts, 5 and 6, secure the parts when screwed down.

The block G may be changed for one of a different length, or a piece may be eut oft' from it to adjust the leg to the proper length.....

Figure B18. Patent Document Example for "Comfort" with Spread at 3. Note: The figure presents a patent document example. We focus the machine learning algorithm's attention to the keywords (blue) and the surrounding context (red). In this case spread is 3 and the trait of interest is "comfort". We correct spelling errors using a preprocessing procedure.

Panel A: Comfort



Panel B: Simplicity

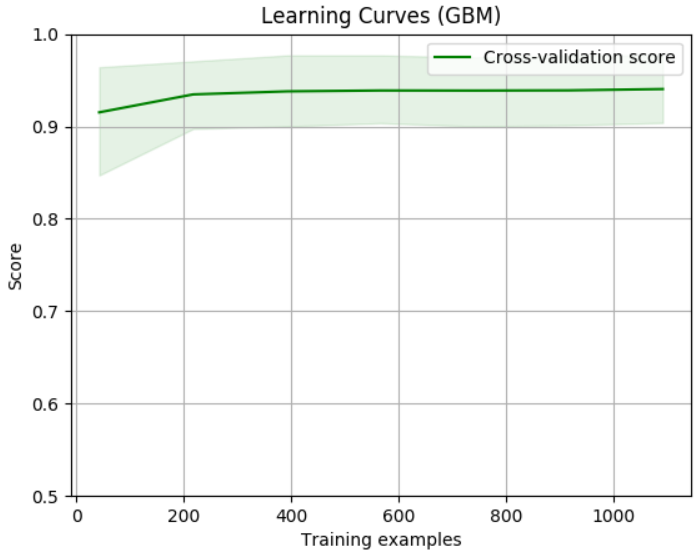
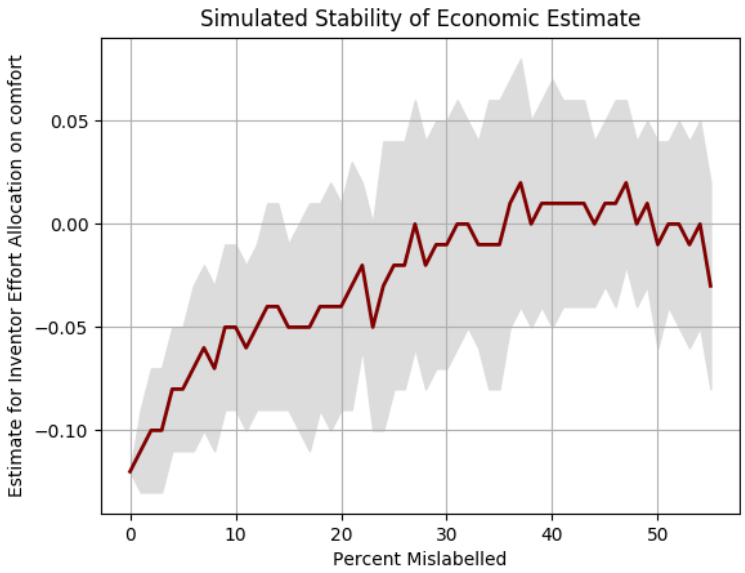


Figure B19. Learning Curve Balanced Accuracy Score. Note: The figure presents the "learning curves" for our preferred modified approach using a GBM algorithm when predicting the presence of our traits in patent documents. Panel A shows the learning curve for "comfort," and panel B shows the learning curve for "simplicity." The solid green line in each panel traces the mean of the balanced accuracy score across 400 iterations of a bootstrap cross-validation procedure at ascending sample sizes. Each bootstrap iteration randomly selects a training set of the "training examples" size to train the model, and the model's accuracy is then tested on the remaining un-sampled data. The shaded green area extends from the 10th to the 90th percentiles of the distribution of results. Balanced accuracy is reported in decimals (0.9 = 90% correctly predicted).

Panel A: Comfort



Panel B: Simplicity

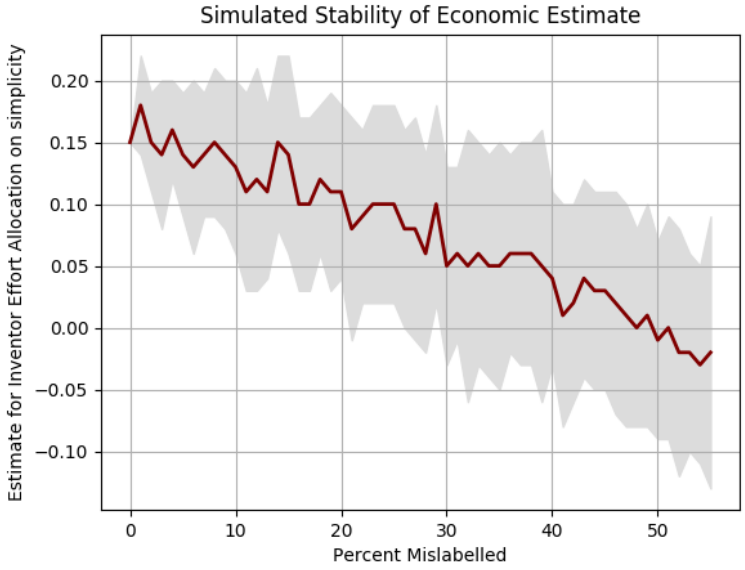


Figure B20. Estimate Stability To Reductions in the Accuracy Score. Note: The figure shows the simulated stability of our economic estimates as we reduce the accuracy of our preferred algorithm. Panel A shows the simulated stability for our "comfort" variable, and panel B shows the simulated stability of our "simplicity" variable. Using all the data generated by our preferred modified approach, we inject noise at random by altering the coding of a given percentage of the observations for our estimates of interest. We then re-estimate beta-one from equation (2.4.4) using a synthetic control procedure. We do this 40 times, sampling with replacement for each percent mislabeled. The red line in each panel traces the mean of the estimates of beta-one from equation (2.4.4) at each percent mislabeled. The shaded grey area shows one standard deviation above and below the mean.

Table B25. Balanced Accuracy Scores Across Training and Test Set Contexts. Note: The table shows the ability of our preferred modified approach applied to a GBM model to predict our traits within and outside the context of the model's training data. We present balanced accuracy scores across wars and broad patent technological classes. Panel A shows the balanced accuracy scores when predicting "comfort," and panel B shows the balanced accuracy scores when predicting "simplicity". Balanced accuracy is reported in percentage terms (78.4 = 78.4% correctly predicted). The main diagonal presents the balanced accuracy means that are obtained through repeated 10-fold cross-validation, using the same context for training and testing. Off-diagonal entries present the model's once-calculated balanced accuracy on the given left-out test set of a different context. The i, j entry corresponds to using the data from row header context i in GBM training to predict the left-out data from column header context j. CWP uses Civil War prosthesis patents, CWC uses Civil War control patents, WWP uses WWI prosthesis patents, and WWC uses the WWI control patents. To ensure that differences between balanced accuracy scores across contexts are not driven by differences in sample size, we constrain the size of the training set to be equal in all cases.

Panel A: Comfort

		Test Data			
		CWP	CWC	WWP	WWC
Training Data	CWP	93.9	84.4	91.8	78.4
	CWC	93.1	91.6	91.8	75.8
	WWP	93.6	84.4	92.7	78.4
	WWC	91.3	84.0	90.0	91.6

Panel B: Simplicity

		Test Data			
		CWP	CWC	WWP	WWC
Training Data	CWP	97.0	86.0	94.8	89.1
	CWC	96.7	94.8	93.8	93.0
	WWP	95.8	86.0	94.8	89.1
	WWC	98.4	92.7	95.4	93.5

Table B26. Performance of Algorithm Across Attributes Using All Patents. Note: The table shows the performance of our modified approach applied to a GBM algorithm across our traits of interest. We present the sensitivity (true-positive rate), specificity (true-negative rate), and the balanced accuracy (simple average of mean sensitivity and specificity). Sensitivity and specificity means are taken across repeated 10-fold cross-validation, and the corresponding standard errors are reported below each point estimate in parenthesis. All evaluation metrics and standard errors are reported in percentage terms (94.8 = 94.8% correctly predicted). All manually coded observations are used in the cross-validation procedure.

Characteristic	Sensitivity	Specificity	Balanced Accuracy
adjustability	94.8	91.0	92.9
	(3.2)	(3.3)	
comfort	91.8	96.3	94.0
	(5.6)	(2.3)	
simplicity	92.7	94.3	93.5
	(5.3)	(2.6)	
materials	81.6	92.4	87.0
	(15.7)	(2.6)	
appearance	91.8	96.1	93.9
	(7.1)	(1.7)	
cost	94.7	98.9	96.8
	(4.3)	(1.1)	

B3 Additional Discussion of the Synthetic Control Strategy for Analyzing Patent Traits

Table B27 presents data on the baseline means for our patent trait variables for prosthetic devices, for the full sample of other medical and mechanical control classes, and the synthetic control group for each trait. The synthetic control procedure successfully brings the baseline means for the control groups much closer to the means for prosthetic devices. Notably, although the mean for appearance is matched quite closely for the World War I sample, the mean for the Civil War control group remains moderately below the mean for prosthetic devices. This reflects both the difficulty of matching quality-oriented traits and the moderate size of our samples of Civil War-era patents relative to World War I-era patents. Consequently, results for our analysis of appearance during the Civil War period ought to be interpreted with caution.

Tables B28 and B29 present the weights our synthetic control procedure assigns to the classes that contribute to each synthetic control group. We make several observations regarding the synthetic control weights. First, the synthetic control groups for our production process traits strike us as being reasonable. At the same time, they are not particularly illuminating. This is reassuring since, as noted above, improvements in the production process can be described using language that is common across mechanical and medical technologies, making the choice of control group relatively inconsequential. Second, the classes that form a synthetic control for "comfort" are quite intuitive. These classes include surgical categories, dentistry, and land vehicles. Third, the classes that form our Civil War synthetic control for "appearance" are superficially counterintuitive, as they include the category "Ammunition and explosive-charge making." An inspection of the underlying patents, however, reveals that the relevant ammunition patents devote attention to the "finishing" process, which indeed denote improvements in product appearance. Nonetheless, we take this as illustrative of the challenges of selecting control groups for a technology's quality-oriented attributes.

Table B27. Baseline Summary Statistics for Prosthetic Devices, All Control Classes, and Re-Weighted Synthetic Control Classes. Note: This table presents baseline means for three samples, namely prosthetics, the "all controls" sample, and the "synthetic controls" sample. Panel A presents baseline means for the Civil War period, for which the baseline extends from 1855 to 1861. Panel B presents baseline means for the World War I period, for which the baseline extends from 1910 to 1915. The "all controls" sample consists of patents from all mechanical classes and all medical classes other than prosthetics. The "synthetic controls" sample was selected to match baseline prosthetics on their values across each year from 1855 to 1861 in panel A and across each year from 1910 to 1915 in panel B.

Panel A: Civil War	Prosthetics	All Controls	Synthetic Controls
production	0.188	0.227	0.189
usertraits	0.255	0.0694	0.245
cost	0.117	0.193	0.118
simplicity	0.102	0.185	0.11
adjustability	0.346	0.303	0.35
appliances	0	0.0445	
comfort	0.350	0.0685	0.346
appearance	0.415	0.0952	0.352
durability	0.730	0.622	0.729
materials	0.0327	0.0550	0.0328

Panel B: World War I	Prosthetics	All Controls	Synthetic Controls
production	0.318	0.355	0.318
usertraits	0.241	0.0778	0.241
cost	0.156	0.263	0.158
simplicity	0.363	0.391	0.362
adjustability	0.436	0.411	0.436
appliances	0.0744	0.0932	0.0744
comfort	0.426	0.0693	0.426
appearance	0.223	0.0708	0.222
durability	0.750	0.750	0.742
materials	0.0385	0.0585	0.0386

Table B28. Civil War Synthetic Control Classes by Trait. Note: The table presents sets of synthetic control "donor" classes for each trait from our Civil War sample. Class numbers are from the United States Patent Classification (USPC) system. A synthetic control weight for each donor class is provided for each trait.

Trait	Class Title	Class	Weight
Adjustability	Traversing Hoists	212	0.31
	Rotary Shafts, Gudgeons, Housings...	464	0.27
	Lubrication	184	0.19
	Optical: Systems And Elements	359	0.19
	Vehicle Fenders	293	0.04
Appearance	Land Vehicles: Bodies And Tops	296	0.56
	Ammunition And Explosive-Charge Making	86	0.36
	Severing By Tearing Or Breaking	225	0.08
Comfort	Advancing Material Of Indeterminate Length	226	0.5
	Ventilation	454	0.28
	Land Vehicles	280	0.09
	Surgery: Light, Thermal, And Electrical Application	607	0.07
	Dentistry	433	0.06
Cost	Elevator, Industrial Lift Truck, Or Stationary Lift...	187	0.49
	Ammunition And Explosive-Charge Making	86	0.26
	Abrading	451	0.22
	...	...	...
Durability	Metal Working	29	0.23
	Winding, Tensioning, Or Guiding	242	0.21
	Wireworking	140	0.2
	...	...	...
Materials	Railway Wheels And Axles	295	0.09
	Fluid Sprinkling, Spraying, And Diffusing	239	0.08
	Wood Turning	142	0.07
	Coopering	147	0.07
	...	...	...
Production	Fasteners (Expanded, Threaded, Driven, etc.)	411	0.36
	Surgery: Light, Thermal, And Electrical Application	607	0.29
	Endless Belt Power Transmission Systems...	474	0.11
	...	...	...
Simplicity	Optical: Systems And Elements	359	0.32
	Ammunition And Explosive-Charge Making	86	0.29
	Railway Rolling Stock	105	0.17
	...	...	...
User	Ventilation	454	0.85
	Cutters, For Shaping	407	0.15

B4 Supplemental Analysis, Figures, and Tables

This appendix presents additional evidence on the effects of wartime demand on counts of medical innovation. First, Table B30 presents estimates of equation (2.4.2). The estimates in table B30 differ from the estimates in table 2.4 exclusively by model choice. That is, they are estimates of the Poisson model described by equation (2.4.2) rather than the OLS model described by equation (1.4.1). All estimates are between 0.54 and 0.88, suggesting that wartime demand shocks led to large increases in flows of prosthetic device patents. As in table 2.4, the estimates in panels B and C reveal economically larger increases during the Civil War than during World War I.

Second, figure B21 presents estimates of the following event-study model:

$$E[N_{t,c}|X_t] = exp(\gamma_{c,w} + \gamma_{t,w} + \sum_{t \neq 0} \beta_t 1\{\text{Prosthetic}\}_c \times 1\{\text{Year of War}\}_t + \varepsilon_{c,t}). \tag{B4.1}$$

In contrast with our estimates of equations (1.4.1) and (2.4.2), for which we collapsed the data into multi-year time periods, we estimate equation (B4.1) using data that are collapsed at an annual frequency. In the summation, the omitted interaction between the prosthetic device indicator variable and the time dummy variables corresponds with the first full year of either the Civil War or World War I (i.e., the year for which $t = 0$ is the first full year of either war). Each β_t can thus be described as a difference-in-differences style estimate of the change in the prosthetic device patenting rate relative to patenting rates in the control categories from year t relative to the first full year of each war. In panel A, the control patent classes consist of all classes other than prosthetic devices that are either medical or mechanical classes. In panel B, the control patent classes are restricted to other medical classes. Standard errors are clustered at the patent class-by-war episode level. For reasons discussed in the main text, these standard errors are likely to be insufficiently conservative, which motivates our use of randomization methods for inference when we assess the statistical significance of our primary estimates of interest.

The estimates trace out the differential changes one can observe through careful inspection of the time series in figure 2.1. Crucially, the point estimates associated with years prior to each war (i.e., $t < 0$) exhibit no discernable pattern that might be suggestive of a worrisome pre-existing trend. The point estimate for year $t = -1$ is fairly close to 0, is moderately smaller than the estimates for year $t = -2$ through $t = -5$, is moderately larger than the estimates for $t = -8$ through $t = -6$ and is economically indistinguishable from the estimate for years $t = -9$ through $t = -12$. Prosthetic device patenting exhibits a strong increase relative to the control categories across years $t = 1$ through $t = 7$. There is a notable peak in years $t = 3$ and $t = 4$, which correspond with the 4th and 5th full calendar years following the onset of each war.

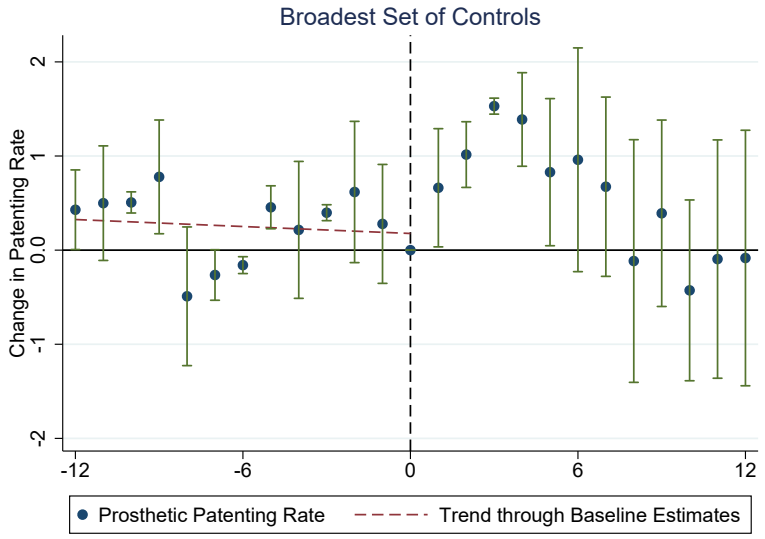
Third, note that the standard errors are presented in parenthesis below the estimates in panel A of table 2.4 and table B30 are conventional cluster-robust standard errors. Due to the small number of "treated patent class episodes" in our sample, however, conventional cluster-robust standard errors may result in insufficiently conservative inference (Bertrand et al. 2004, Cameron et al. 2008). In such settings, randomization inference has been found to generate p-values that confer appropriate degrees of statistical significance (Cameron et al. 2008, Imbens and Rosenbaum 2005). Figure B23 displays our prosthesis point estimates (dashed vertical lines) in the context of distributions generated from three distinct randomization inference procedures.[32] In each case, the "true point estimate" is larger in magnitude than nearly the entirety of the "placebo distribution." One of the 500 estimates exceeds the true estimate when using assignment algorithm A, two when using algorithm B, and zero when using algorithm C. The implication, in each case, is that our estimates are statistically distinguishable from zero at the p $<$.01 level.

[32]We use three distinct procedures for assigning placebo treatment status. In each case, we assign placebo treatment status to two patent class-by-episode observations. The sample from which these are drawn includes mechanical and medical patent classes other than prosthetic devices. For the first procedure (presented in panel A of figure B23), we assign placebo treatment status at random across both treatment episodes. For the second (presented in panel B of figure B23), we assign treatment at random to one patent class from each of the treatment episodes. For the third, we restrict the sample to patent classes that appear in both the Civil War and World War I sub-samples, then assign treatment at random to a single patent class. The dispersion of the distributions of placebo point estimates is only modestly affected by these alternative assignment mechanisms.

Table B29. World War I Synthetic Control Classes by Trait. Note: The table presents sets of synthetic control "donor" classes for each trait from our World War I sample. Class numbers are from the United States Patent Classification (USPC) system. A synthetic control weight for each donor class is provided for each trait.

Trait	Class Title	Class	Weight
Adjustability	Photocopying	355	0.28
	Surgery	600	0.28
	Compound Tools	7	0.22
	Abrasive Tool Making Process...	51	0.22
Appearance	Plastic And Nonmetallic Article Shaping Or Treating	264	0.43
	Roll Or Roller	492	0.22
	Solid Anti-Friction Devices...	508	0.17
	Surgery: Light, Thermal, And Electrical Application	607	0.1
	Needle And Pin Making	163	0.08
Appliances	Optics: Motion Pictures	352	0.48
	Wood Turning	142	0.16
	Optics: Image Projectors	353	0.13
	Alloys Or Metallic Compositions	420	0.11
Comfort	Surgery	128	0.67
	Ventilation	454	0.26
	Surgery: Light, Thermal, And Electrical Application	607	0.07
Cost	Selective Cutting (E.G., Punching)	234	0.55
	Sheet Feeding Or Delivering	271	0.35
	Surgery: Light, Thermal, And Electrical Application	607	0.08
	Roll Or Roller	492	0.02
Durability	Surgery: Light, Thermal, And Electrical Application	607	0.7
	Rotary Kinetic Fluid Motors Or Pumps	415	0.3
Materials	Cutters, For Shaping	407	0.56
	Railway Wheels And Axles	295	0.23
	Conveyors, Chutes, Skids, Guides, And Ways	193	0.17
	Solid Anti-Friction Devices...	508	0.04
Production	Selective Cutting (E.G., Punching)	234	0.29
	Motors: Spring, Weight, Or Animal Powered	185	0.24
	Roll Or Roller	492	0.16
	...	...	...
Simplicity	Sheet-Material Associating	270	0.6
	Needle And Pin Making	163	0.18
	Lubrication	184	0.12
	...	...	...
User	Surgery: Light, Thermal, And Electrical Application	607	0.34
	Ventilation	454	0.29
	Surgery: Splint, Brace, Or Bandage	602	0.27
	Compound Tools	7	0.09

Panel A



Panel B

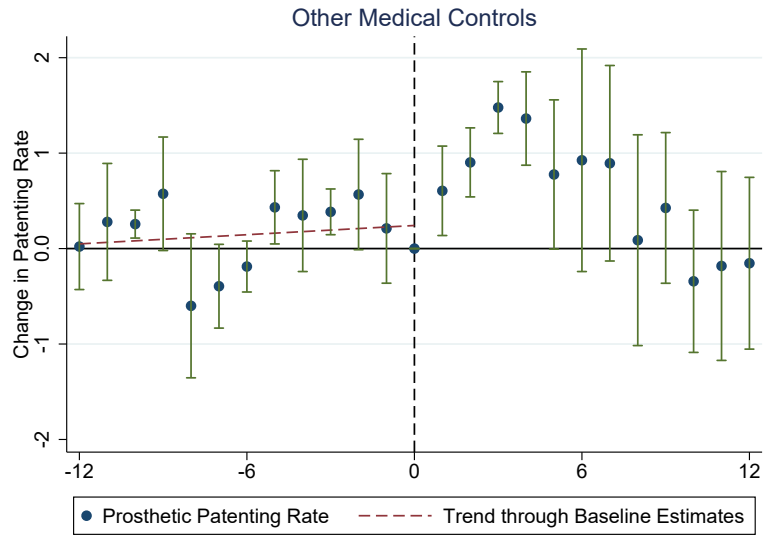
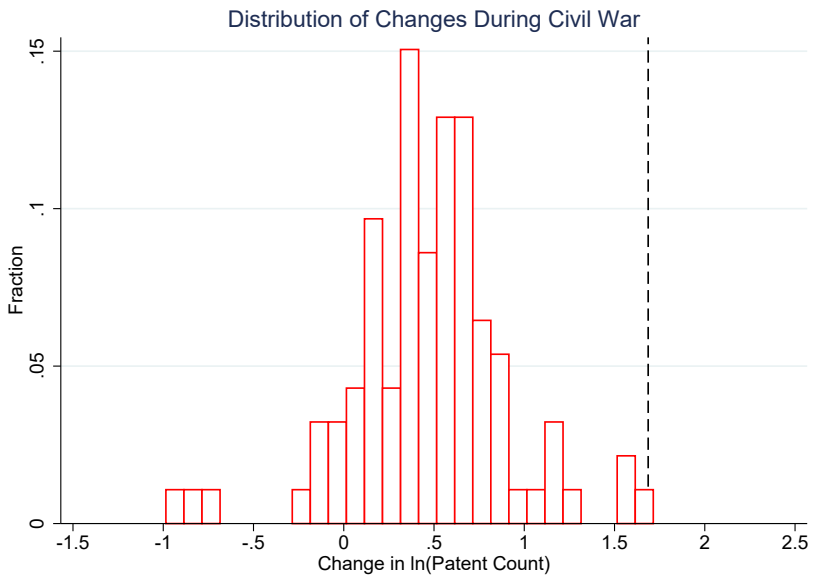


Figure B21. Event Study Estimates of Changes in Prosthetic Device Patenting Rates During the Civil War and World War I. Note: The figure presents estimates of the β_t coefficients from equation (B4.1). Data are analyzed at an annual frequency. The omitted year corresponds with the first full year of either the Civil War or World War I, such that each β_t can be described as a difference-in-differences style estimate of the change in the prosthetic device patenting rate relative to patenting rates in the control categories from year t relative to the first full year of each war. In panel A, the control patent classes consist of all classes other than prosthetic devices that are either medical or mechanical classes. In panel B, the control patent classes are restricted to other medical classes. Standard errors are clustered at the patent class-by-war episode level. For reasons discussed in the main text, these standard errors are likely to be insufficiently conservative, which motivates the use of randomization methods for inference when we assess the statistical significance of our primary estimates of interest.

Panel A



Panel B

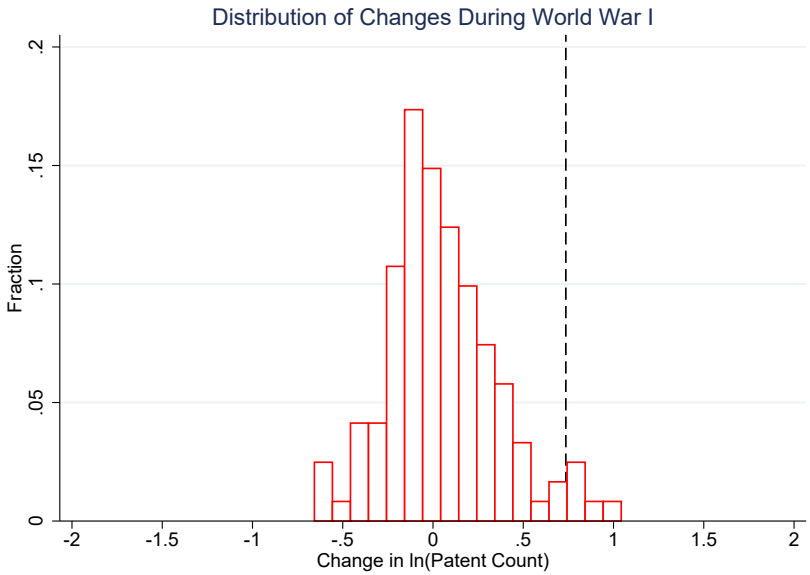


Figure B22. Patents in Prosthetic Devices and Mechanical Classes. Note: This figure presents distributions of changes in the log of patents per year. Each data point in each distribution corresponds with a change for an individual USPTO class. The changes in panel A are calculated from a "base" period extending from 1855 to 1861 to a "war" period extending from 1862 to 1866. The changes in panel B are calculated from a "base" period extending from 1910 to 1915 to a "war" period extending from 1916 to 1922. The vertical dashed line in each panel corresponds with the change that occurred in USPTO class 623 "Prosthesis."

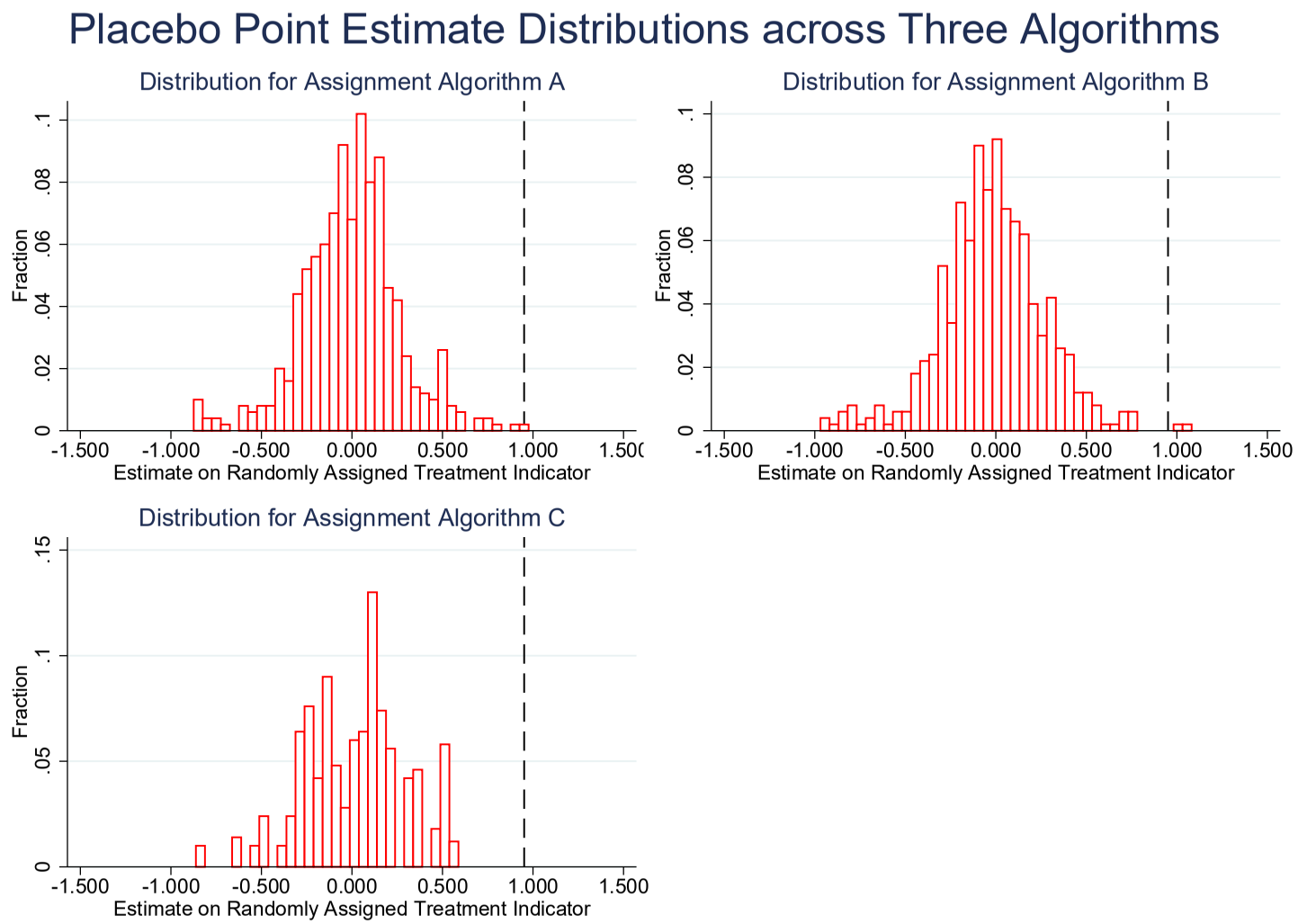
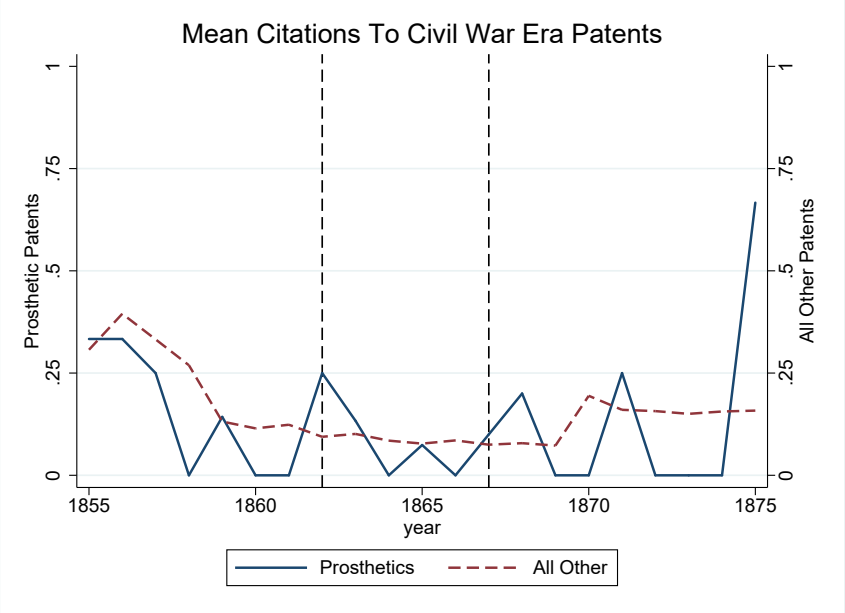


Figure B23. Placebo Point Estimate Distributions across Three Algorithms. Note: The figure presents distributions of placebo point estimates generated through the application of a randomization inference procedure (Imbens and Rosenbaum 2005). The distribution in each panel corresponds with a different algorithm for assigning placebo treatment status. In each case, we assign placebo treatment status to two patent class-by-episode observations. For observations associated with the Civil War, the pre-war period extends from 1855 to 1861, while the period over which the war influenced prosthetic device patenting is defined to extend from 1862 to 1866. For observations associated with World War I, the pre-war period extends from 1910 to 1915, while the period over which the war influenced prosthetic device patenting is defined to extend from 1916 to 1922. The sample from which these are drawn includes all mechanical and medical patent classes other than prosthetic devices. For Panel A, we assign placebo treatment status at random across this full set of episodes. For Panel B, we assign treatment at random to one patent class from each of the war episodes. For Panel C, we restrict the sample to patent classes that appear in both the Civil War and World War I sub-samples, then assign treatment at random to a single patent class. In each panel, the true estimate associated with assigning treatment status to "Prosthesis" is presented by the dashed vertical lines.

Panel A



Panel B

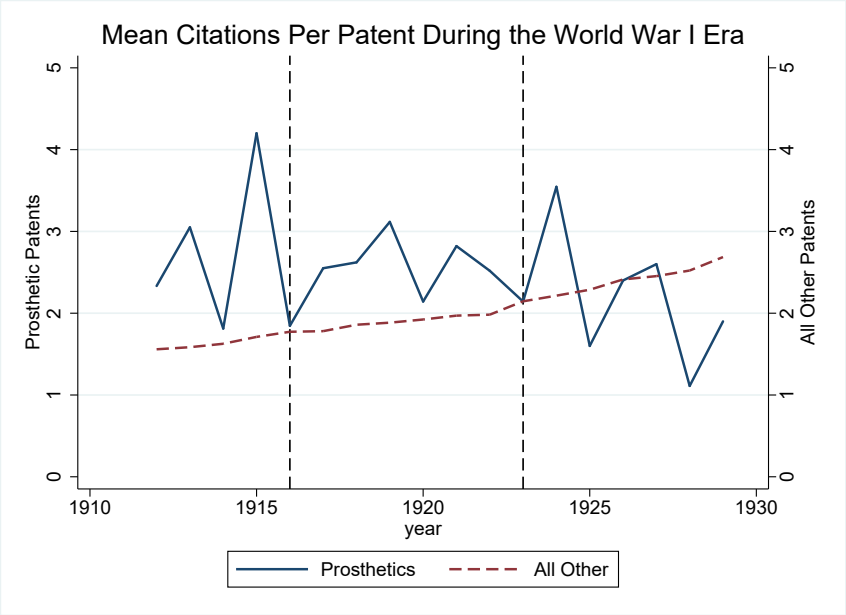


Figure B24. Mean Citations Per Patent. Note: This figure presents time series on mean citations per patent. The data come from the citation files associated with comprehensive patent data from Berkes (2018). Citation data from the Civil War period are sparse because, as discussed by Berkes (2018), citations in patent documents, and by extension in the database, became more systematic and comprehensive over time. Dashed vertical lines indicate the periods we associate with wartime prosthetic device patenting, namely 1862 to 1866 during the Civil War and 1916 to 1922 during World War I.

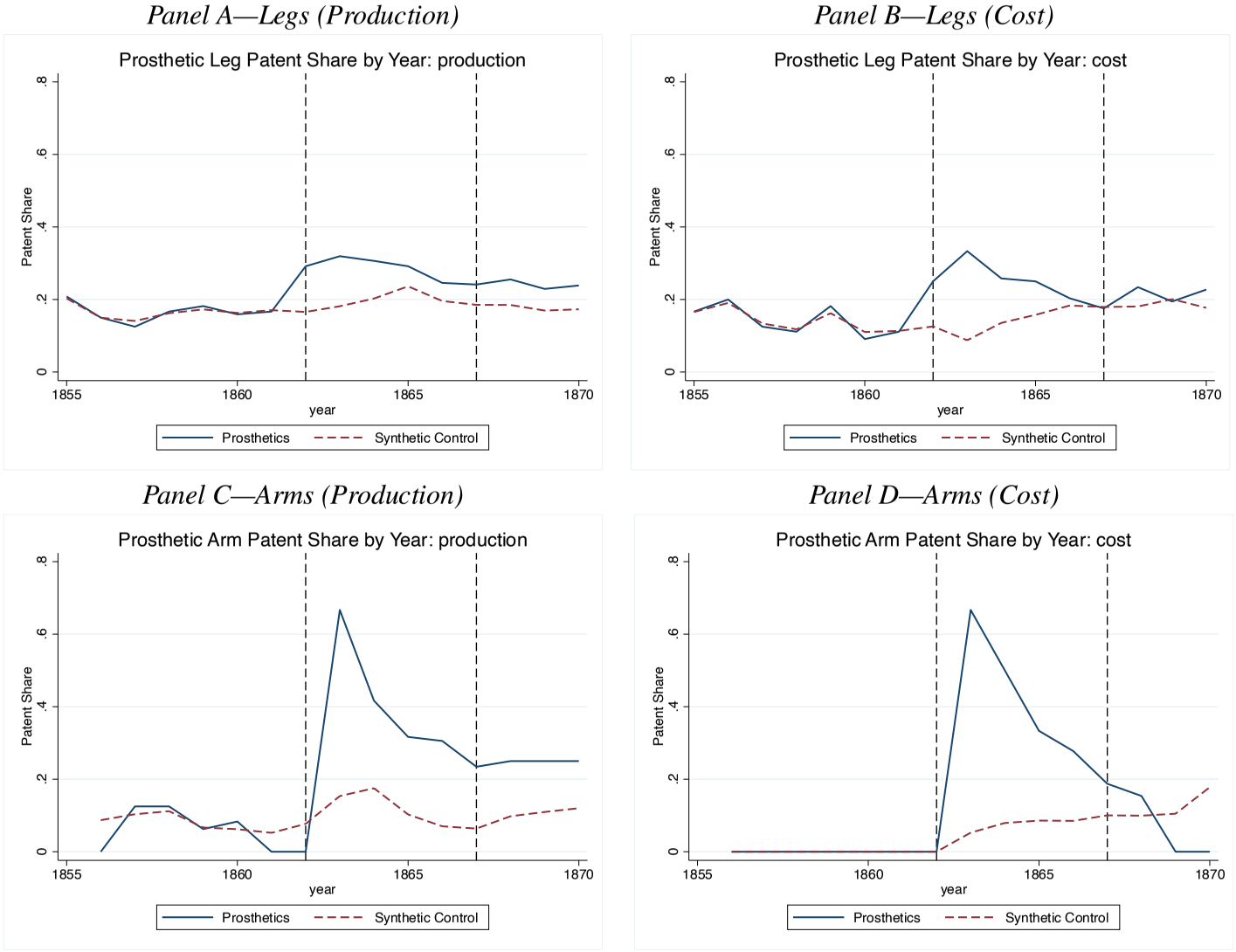
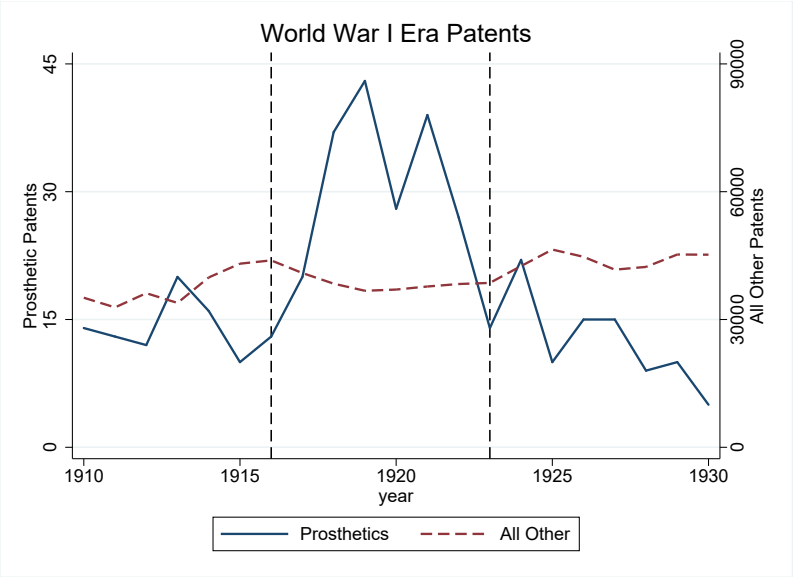


Figure B25. Civil War Changes in the Cost-Oriented Traits: Prosthetic Legs vs. Arms. Note: The figure presents data on the "treatment" and "synthetic control" series that describe the evolution of patents' emphases on averages across the "cost" trait and our aggregate "production" trait. The time series are calculated as 4-year moving averages. The series plot the share of patents in a given class ("Prosthesis" or the "Synthetic Control") and prosthetic limb type (prosthetic leg or prosthetic arm) that emphasize a given trait. In Panels A and B, which represent changes in cost-oriented traits for prosthetic legs, the "Pre War" baseline extends from 1855 to 1861, and the "Wartime" period extends from 1862 to 1866. In Panels C and D, which represent changes in cost-oriented traits for prosthetic arms, the "Pre War" baseline extends from 1856 to 1861, and the "Wartime" period extends from 1862 to 1866. We generate the synthetic control group using the "synth" package written by Abadie and Hainmueller (2010). "Donor weights" are chosen to match the treatment group on values extending from 1855 to 1861 for Panels A and B, and from 1856 to 1862 for Panels C and D as the procurement program for prosthetic arms was officially implemented after the prosthetic leg program.

Panel A



Panel B

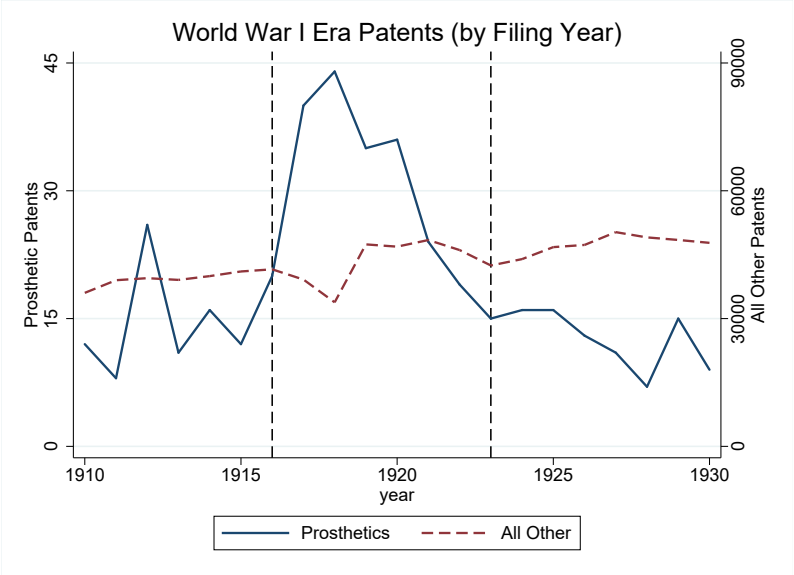


Figure B26. Patent Time Series. Note: This figure presents annual time series on patents, using USPTO categories as reported in Berkes (2018). In both panels, the solid blue line corresponds with patents from USPTO class 623 "Prosthesis." In the top panel, the patents are organized in accordance with the year in which the patent was issued, while in the bottom panel, the patents are organized in accordance with the year in which the patent was filed.

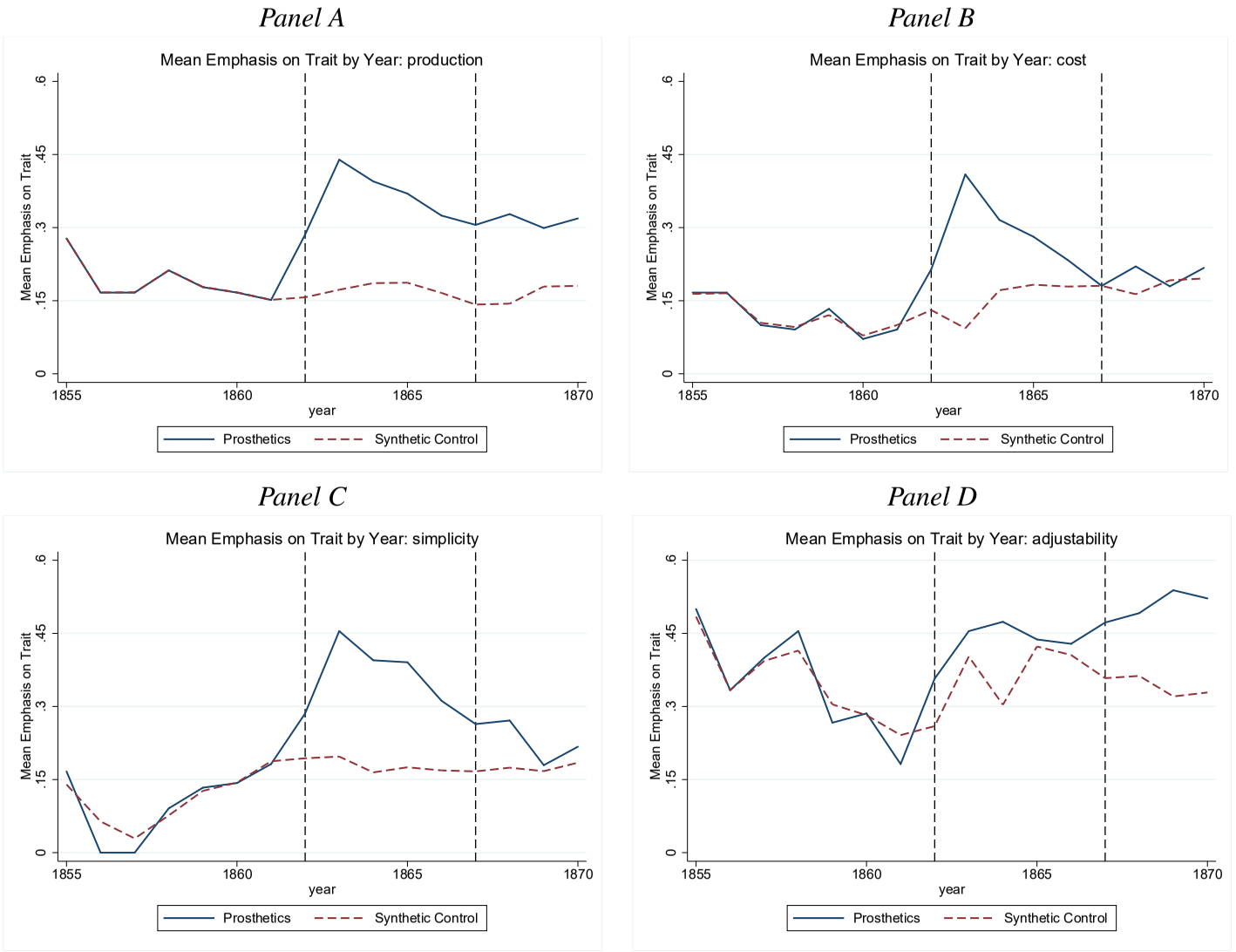


Figure B27. Production Traits: Civil War Synthetic Controls. Note: The figure presents the "treatment" and "synthetic control" series that describe the evolution of patents' emphases on the traits we term 'production," "cost," "simplicity," and "adjustability." Further information on the definitions of each trait can be found in table 2.3 as well as in the main text. All series in the figure are calculated as 4-year moving averages. The series plot the share of patents in a given class ("Prosthesis" or the "Synthetic Control") that emphasize a given trait. We generate the synthetic control group using the "synth" package written by Abadie and Hainmueller (2010). "Donor weights" are chosen to match the treatment group on values extending from 1855 to 1861.

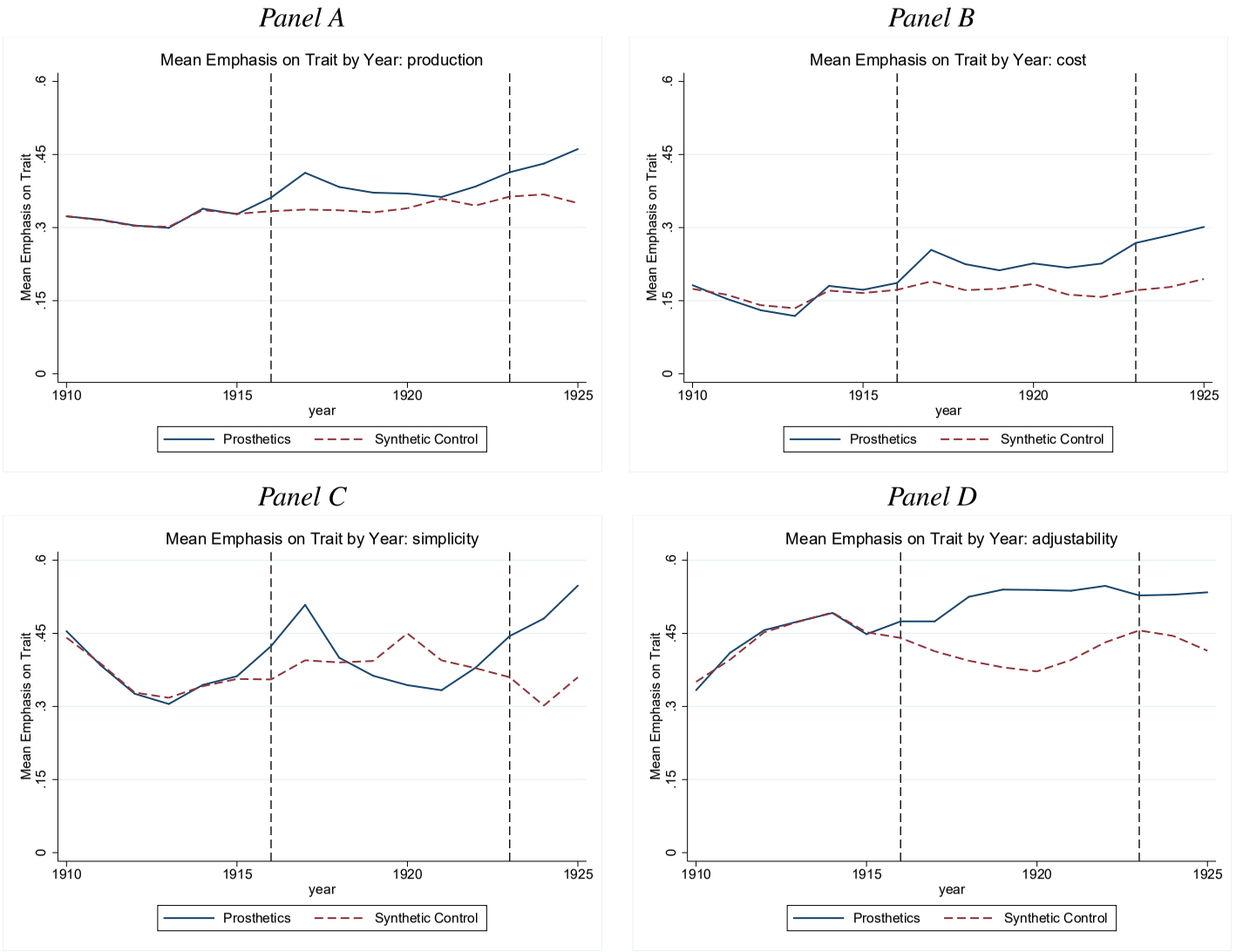


Figure B28. Production Traits: World War I Synthetic Controls. Note: The figure presents the "treatment" and "synthetic control" series that describe the evolution of patents' emphases on the traits we term "production," "cost," "simplicity," and "adjustability." Further information on the definitions of each trait can be found in table 2.3 as well as in the main text. All series in the figure are calculated as 4-year moving averages. The series plot the share of patents in a given class ("Prosthesis" or the "Synthetic Control") that emphasize a given trait. We generate the synthetic control group using the "synth" package written by Abadie and Hainmueller (2010). "Donor weights" are chosen to match the treatment group on values extending from 1910 to 1915.

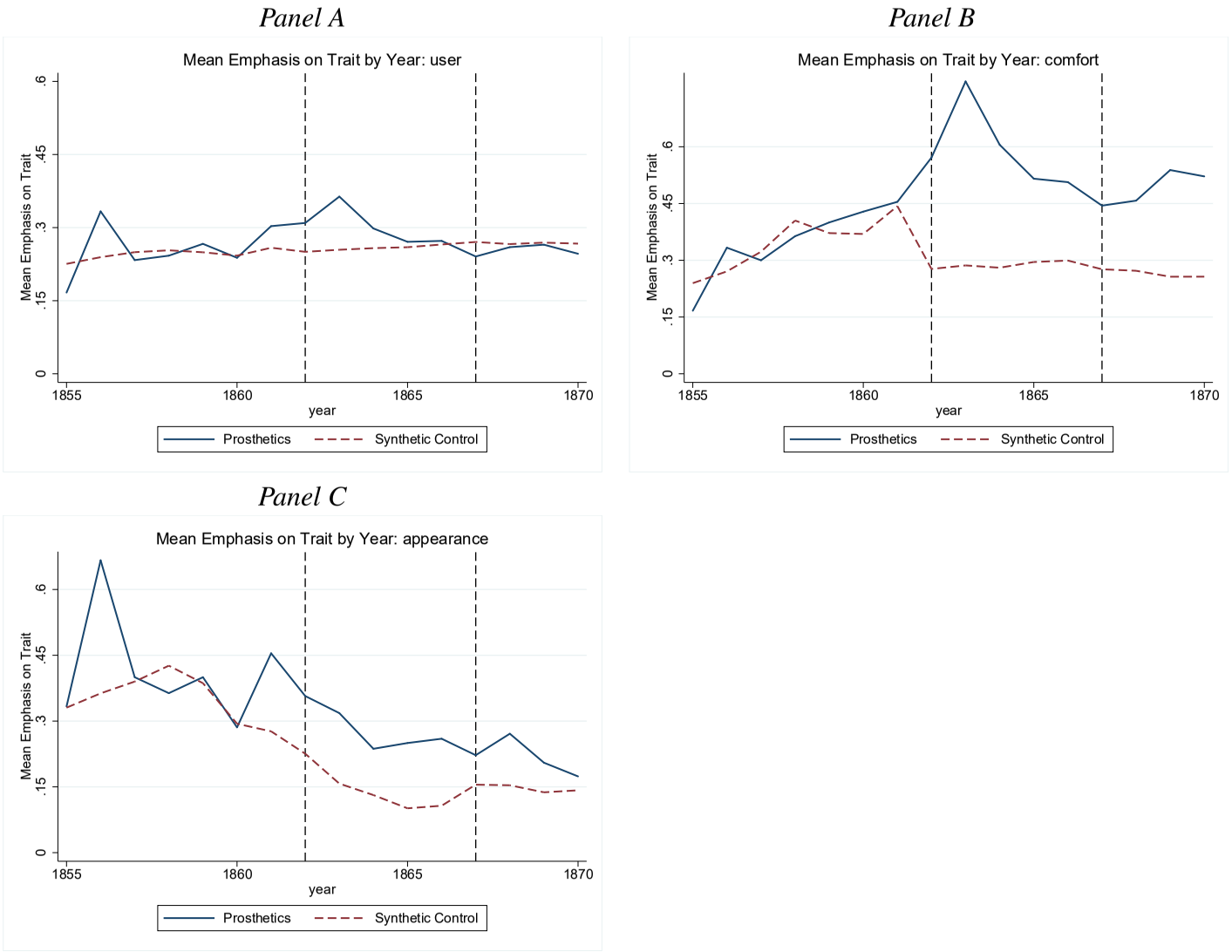


Figure B29. User Traits: Civil War Synthetic Controls. Note: The figure presents the "treatment" and "synthetic control" series that describe the evolution of patents' emphases on the traits we term "user traits," "comfort," and "appearance." Further information on the definitions of each trait can be found in table 2.3 as well as in the main text. All series in the figure are calculated as 4-year moving averages. The series plot the share of patents in a given class ("Prosthesis" or the "Synthetic Control") that emphasize a given trait. We generate the synthetic control group using the "synth" package written by Abadie and Hainmueller (2010). "Donor weights" are chosen to match the treatment group on values extending from 1855 to 1861.

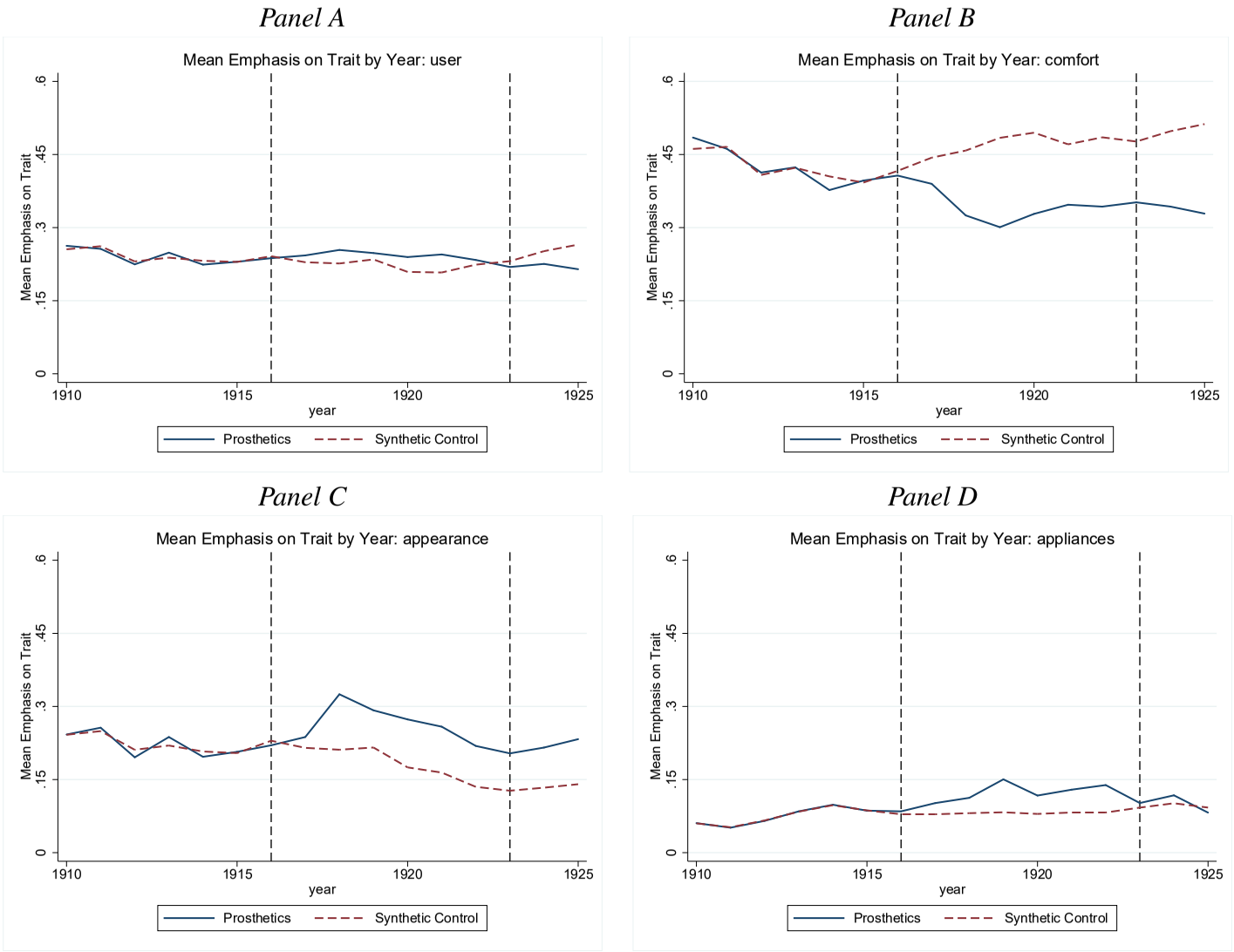


Figure B30. User Traits: World War I Synthetic Controls. Note: The figure presents the "treatment" and "synthetic control" series that describe the evolution of patents' emphases on the traits we term "user traits," "comfort," "appearance," and "appliances." Further information on the definitions of each trait can be found in table 2.3 as well as in the main text. All series in the figure are calculated as 4-year moving averages. The series plot the share of patents in a given class ("Prosthesis" or the "Synthetic Control") that emphasize a given trait. We generate the synthetic control group using the "synth" package written by Abadie and Hainmueller (2010). "Donor weights" are chosen to match the treatment group on values extending from 1910 to 1915.

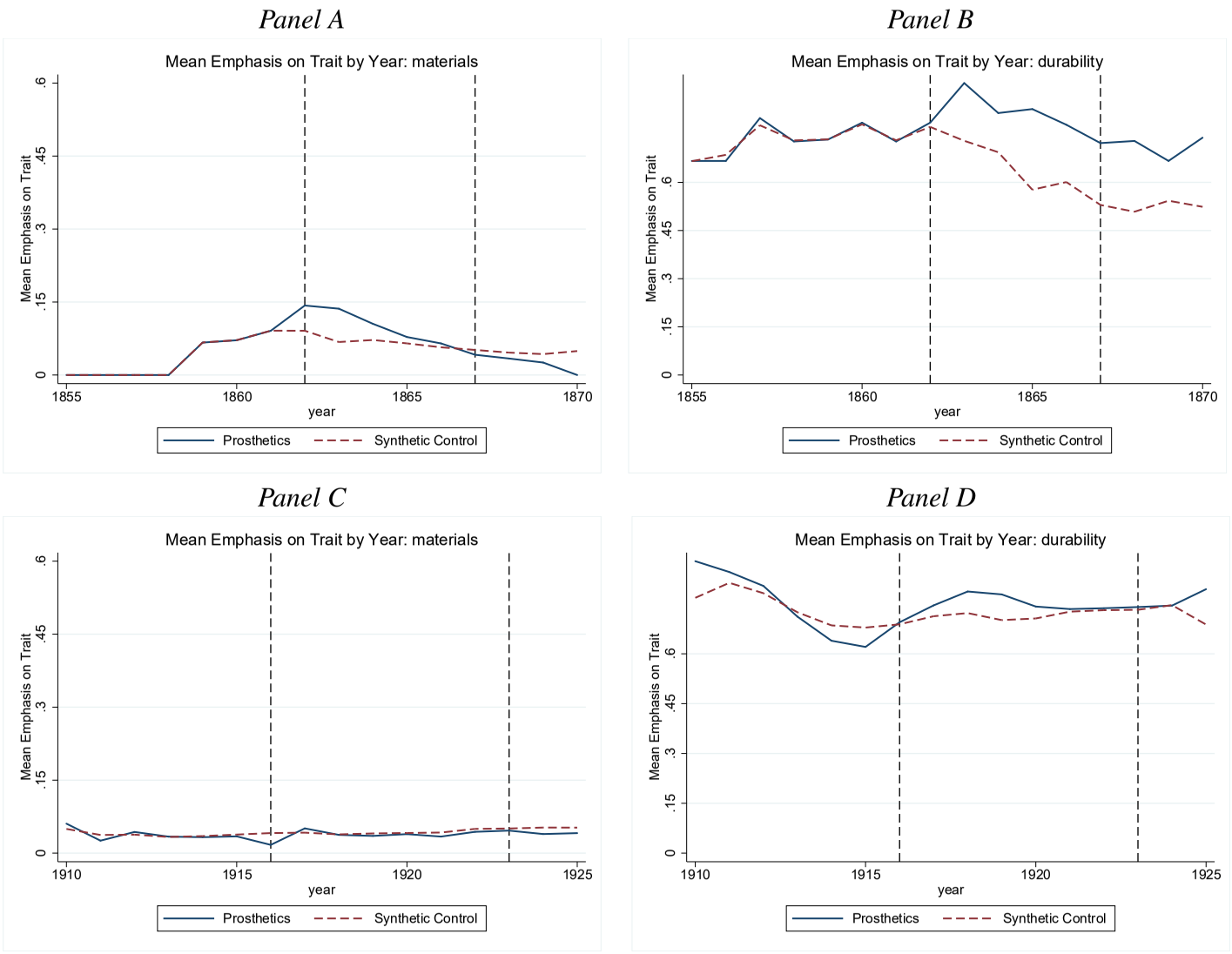


Figure B31. Materials and Durability: Civil War and World War I Synthetic Controls. Note: The figure presents the "treatment" and "synthetic control" series that describe the evolution of patents' emphases on the traits we term "materials" and "durability." Further information on the definitions of each trait can be found in table 2.3 as well as in the main text. All series in the figure are calculated as 4-year moving averages. The series plot the share of patents in a given class ("Prosthesis" or the "Synthetic Control") that emphasize a given trait. We generate the synthetic control group using the "synth" package written by Abadie and Hainmueller (2010). "Donor weights" for panels C and D are chosen to match the treatment group on values extending from 1855 to 1861. "Donor weights" for panels C and D are chosen to match the treatment group on values extending from 1910 to 1915.

Table B30. Relative Increases in Prosthetic Device Patenting During the Civil War and World War I. Note: The table presents estimates of equation (2.4.2). The control group used for each regression is described in the column heading. The sample for Panel A includes both the Civil War and World War I episodes, while the sample for Panel B consists solely of the Civil War episode and the sample for Panel C consists solely of the World War I episode. For observations associated with the Civil War, the pre-war period extends from 1855 to 1861, while the period over which the war influenced prosthetic device patenting is defined to extend from 1862 to 1866. For observations associated with World War I, the pre-war period extends from 1910 to 1915, while the period over which the war influenced prosthetic device patenting is defined to extend from 1916 to 1922. In Panel A, the standard errors reported in parentheses allow for clusters at the patent class-by-war episode level. In each panel, the p-values reported in rows labeled "Randomization Inf" are based on the position of the point estimate in the distribution of placebo point estimates that are constructed using a procedure along the lines recommended by Imbens and Rosenbaum (2005). Additional details are reported in the main text.

	(1) All Cntrls	(2) Matched	(3) Medical	(4) Misc. Mech.	(5) Metal	(6) Mater. Proc.	(7) Non War
Panel A: Full Sample							
Prosthetics x War	0.812	0.542	0.797	0.776	0.818	0.879	0.808
	(0.142)	(0.175)	(0.097)	(0.119)	(0.153)	(0.194)	(0.147)
N	432	88	34	128	56	92	362
Clusters	216	44	17	64	28	46	181
Estimator	Poisson	Poisson	Poisson	Poisson	Poisson	Poisson	Poisson
Class-by-Episode FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Period Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SEs in Parentheses	Clustered	Clustered	Clustered	Clustered	Clustered	Clustered	Clustered
Randomization Inf.	$P < .01$	$P < .01$	$P < .01$	$P < .01$	$P < .01$	$P < .01$	$P < .01$
Panel B: Civil War							
Prosthetics x War	1.233	0.529	1.134	1.110	1.251	1.436	1.243
Randomization Inf.	$P < .01$	$P < .01$	$P < .01$	$P < .01$	$P < .01$	$P < .01$	$P < .01$
N	188	88	14	56	24	42	156
Panel C: WWI							
Prosthetics x War	0.681	0.409	0.711	0.677	0.687	0.699	0.673
Randomization Inf.	$P < .05$	$P < .2$	$P < .01$	$P < .03$	$P < .01$	$P < .01$	$P < .04$
N	244	32	20	72	32	50	206

Table B31. Hand-Coded Training Set Tabulations. Note: The table presents sample means for the patents in our hand-coded training data set. For the complete hand-coded data set, the patents in the Civil War sample extend from 1840 to 1890, while the patents in the World War I sample extend from 1890 to 1940.

	(1)	(2)	(3)	(4)
	Civil War		World War I	
	Prosthetics	Controls	Prosthetics	Controls
production	0.243	0.226	0.360	0.356
user	0.346	0.0501	0.295	0.0475
cost	0.174	0.231	0.235	0.302
simplicity	0.226	0.148	0.394	0.380
adjustability	0.328	0.301	0.450	0.387
comfort	0.497	0.0551	0.371	0.0426
appearance	0.195	0.0451	0.219	0.0525
durability	0.687	0.363	0.384	0.269
materials	0.0462	0.0551	0.0530	0.0852
Observations	195	399	302	305

Table B32. Full Sample Tabulations. Note: The table presents sample means for all the "treatment" and "control" patents in the data set we generate using machine learning methods. For the complete data set, the patents in the Civil War sample extend from 1840 to 1890, while the patents in the World War I sample extend from 1890 to 1940.

	(1)	(2)	(3)	(4)
	Civil War		World War I	
	Prosthetics	Controls	Prosthetics	Controls
production	0.285	0.257	0.378	0.377
user	0.256	0.0562	0.257	0.0784
cost	0.186	0.200	0.245	0.299
simplicity	0.247	0.223	0.405	0.426
adjustability	0.423	0.350	0.484	0.407
appliances	0.0515	0.0568	0.0855	0.0797
appearance	0.222	0.0605	0.276	0.0741
comfort	0.495	0.0513	0.410	0.0813
durability	0.753	0.624	0.747	0.786
materials	0.0464	0.0319	0.0435	0.0611
Observations	194	151038	620	593706

Table B33. Correlations across Patent Attributes. Note: The table presents a simple correlation matrix across the economic traits we have defined and coded. The sample underlying the matrix is the sample of prosthetic device patents extending from 1840 to 1940.

	cost	simplicity	adjustability	appliances	appearance	comfort	durability	materials
cost	1.0000							
simplicity	0.3746	1.0000						
adjustability	0.0512	0.0392	1.0000					
appliances	0.0485	0.0654	0.0040	1.0000				
appearance	0.0568	-0.0151	0.0088	0.0568	1.0000			
comfort	0.0893	-0.0183	0.0684	-0.1123	0.1296	1.0000		
durability	0.1165	0.1192	-0.0045	-0.0014	0.0508	0.0980	1.0000	
materials	0.0523	0.0096	-0.0107	-0.0399	0.0887	0.1028	0.0009	1.0000

Table B34. Hand-Coded Training Set Tabulations and Changes. Note: The table presents sets of means and changes in means for our hand-coded training data set. The means in columns 1 through 4 are calculated separately for baseline prosthetics, wartime prosthetics, baseline controls, and wartime controls. As in our regressions, the Civil War baseline corresponds with 1855 to 1861, while the World War I baseline extends from 1910 to 1915. The Civil War "wartime" period corresponds with 1862 to 1866, while the World War I "wartime" period extends from 1916 to 1922. Column 5 presents the change from baseline to wartime for prosthetics, while column 6 presents the change from baseline to wartime for the controls. Column 7 presents the difference between these differences.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Prosthetics		Other Mechanical			Differences	
Panel A: Civil War	Pre-Boom	Boom	Pre-Boom	Boom	Prosth. Diff	Other Diff	Diff-in-Diff
production	0.0952	0.313	0.218	0.284	0.218	0.0665	0.151
user	0.381	0.364	0.0413	0.0902	-0.0168	0.0488	-0.0656
cost	0.0952	0.247	0.231	0.279	0.152	0.0473	0.104
simplicity	0.0476	0.321	0.132	0.230	0.273	0.0973	0.176
adjustability	0.143	0.370	0.289	0.344	0.228	0.0550	0.173
comfort	0.381	0.506	0.0413	0.0492	0.125	0.00786	0.117
appearance	0.381	0.222	0.0413	0.131	-0.159	0.0898	-0.249
durability	0.714	0.654	0.264	0.508	-0.0600	0.244	-0.304
materials	0.0476	0.0741	0.0826	0.0492	0.0265	-0.0335	0.0599
Panel B: World War I	Pre-Boom	Boom	Pre-Boom	Boom	Prosth. Diff	Other Diff	Diff-in-Diff
production	0.325	0.372	0.329	0.384	0.0465	0.0556	-0.00909
user	0.347	0.271	0.0436	0.0533	-0.0765	0.00971	-0.0862
cost	0.188	0.237	0.248	0.360	0.0485	0.112	-0.0632
simplicity	0.365	0.406	0.369	0.393	0.0411	0.0242	0.0169
adjustability	0.424	0.473	0.369	0.400	0.0499	0.0309	0.0190
comfort	0.506	0.319	0.0537	0.0333	-0.187	-0.0204	-0.167
appearance	0.188	0.222	0.0336	0.0733	0.0340	0.0398	-0.00579
durability	0.318	0.396	0.235	0.300	0.0785	0.0651	0.0134
materials	0.0353	0.0628	0.0671	0.107	0.0275	0.0396	-0.0120

Table B35. Full Sample Tabulations and Changes. Note: The table presents sets of means and changes in means for the full data set we generate using machine learning methods. The means in columns 1 through 4 are calculated separately for baseline prosthetics, wartime prosthetics, baseline controls, and wartime controls. As in our regressions, the Civil War baseline corresponds with 1855 to 1861, while the World War I baseline extends from 1910 to 1915. The Civil War "wartime" period corresponds with 1862 to 1866, while the World War I "wartime" period extends from 1916 to 1922. Column 5 presents the change from baseline to wartime for prosthetics, while column 6 presents the change from baseline to wartime for the controls. Column 7 presents the difference between these differences.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Prosthetics		Other Mechanical			Differences	
Panel A: Civil War	Pre-Boom	Boom	Pre-Boom	Boom	Prosth. Diff	Other Diff	Diff-in-Diff
production	0.159	0.346	0.252	0.241	0.187	-0.0114	0.198
user	0.270	0.276	0.0507	0.0482	0.00588	-0.00254	0.00842
cost	0.0952	0.247	0.197	0.187	0.152	-0.0105	0.162
simplicity	0.0952	0.333	0.213	0.218	0.238	0.00543	0.233
adjustability	0.286	0.457	0.347	0.318	0.171	-0.0291	0.200
appliances	0	0.0494	0.0403	0.0477	0.0494	0.00742	0.0420
appearance	0.429	0.247	0.0682	0.0547	-0.182	-0.0135	-0.168
comfort	0.381	0.531	0.0436	0.0420	0.150	-0.00151	0.151
durability	0.762	0.778	0.638	0.611	0.0159	-0.0274	0.0433
materials	0.0476	0.0741	0.0356	0.0376	0.0265	0.00205	0.0244
Panel B: World War I	Pre-Boom	Boom	Pre-Boom	Boom	Prosth. Diff	Other Diff	Diff-in-Diff
production	0.318	0.391	0.360	0.371	0.0737	0.0108	0.0629
user	0.247	0.240	0.0641	0.0721	-0.00712	0.00804	-0.0152
cost	0.153	0.232	0.270	0.294	0.0789	0.0238	0.0551
simplicity	0.353	0.396	0.412	0.429	0.0432	0.0167	0.0265
adjustability	0.447	0.546	0.397	0.389	0.0988	-0.00807	0.107
appliances	0.0706	0.135	0.0699	0.0784	0.0647	0.00857	0.0561
appearance	0.224	0.256	0.0566	0.0639	0.0325	0.00739	0.0251
comfort	0.447	0.329	0.0658	0.0739	-0.119	0.00815	-0.127
durability	0.694	0.758	0.765	0.772	0.0643	0.00784	0.0565
materials	0.0353	0.0435	0.0419	0.0467	0.00818	0.00483	0.00336

Table B36. Tabulations and Changes with Medical Control Classes Only. Note: The table presents sets of means and changes in means for our full data set, but with the control group restricted to medical patent classes only. The means in columns 1 through 4 are calculated separately for baseline prosthetics, wartime prosthetics, baseline controls, and wartime controls. As in our regressions, the Civil War baseline corresponds with 1855 to 1861, while the World War I baseline extends from 1910 to 1915. The Civil War "wartime" period corresponds with 1862 to 1866, while the World War I "wartime" period extends from 1916 to 1922. Column 5 presents the change from baseline to wartime for prosthetics, while column 6 presents the change from baseline to wartime for the controls. Column 7 presents the difference between these differences.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Prosthetics		Other Mechanical			Differences	
Panel A: Civil War	Pre-Boom	Boom	Pre-Boom	Boom	Prosth. Diff	Other Diff	Diff-in-Diff
production	0.159	0.346	0.207	0.237	0.187	0.0296	0.157
user	0.270	0.276	0.137	0.110	0.00588	-0.0268	0.0326
cost	0.0952	0.247	0.0962	0.142	0.152	0.0460	0.106
simplicity	0.0952	0.333	0.128	0.213	0.238	0.0850	0.153
adjustability	0.286	0.457	0.397	0.355	0.171	-0.0421	0.213
appliances	0	0.0494	0.0192	0.0305	0.0494	0.0112	0.0382
appearance	0.429	0.247	0.0833	0.0609	-0.182	-0.0224	-0.159
comfort	0.381	0.531	0.308	0.239	0.150	-0.0691	0.219
durability	0.762	0.778	0.526	0.533	0.0159	0.00735	0.00852
materials	0.0476	0.0741	0.0577	0.0508	0.0265	-0.00693	0.0334
Panel B: World War I	Pre-Boom	Boom	Pre-Boom	Boom	Prosth. Diff	Other Diff	Diff-in-Diff
production	0.318	0.391	0.354	0.364	0.0737	0.00988	0.0638
user	0.247	0.240	0.161	0.170	-0.00712	0.00991	-0.0170
cost	0.153	0.232	0.251	0.271	0.0789	0.0199	0.0591
simplicity	0.353	0.396	0.388	0.410	0.0432	0.0222	0.0210
adjustability	0.447	0.546	0.424	0.412	0.0988	-0.0124	0.111
appliances	0.0706	0.135	0.141	0.154	0.0647	0.0127	0.0520
appearance	0.224	0.256	0.113	0.118	0.0325	0.00432	0.0282
comfort	0.447	0.329	0.227	0.240	-0.119	0.0128	-0.131
durability	0.694	0.758	0.651	0.675	0.0643	0.0245	0.0399
materials	0.0353	0.0435	0.0616	0.0881	0.00818	0.0265	-0.0183

Table B37. Crude Matching Sample Tabulations and Changes. Note: The table presents sets of means and changes in means for data sets in which the control group is constrained using a simple matching procedure. Specifically, the control group is selected to include all control-group patent classes for which the baseline mean is within 6 percentage points of the mean for prosthetic devices for a given economic trait. The one exception is "comfort" during the World War I episode, for which the control-group patent classes consist of those for which the baseline mean is within 20 percentage points of the mean for prosthetic devices. This reflects the fact that there were no close matches for prosthetic devices with respect to "comfort" during the World War I period. The means in columns 1 through 4 are calculated separately for baseline prosthetics, wartime prosthetics, baseline controls, and wartime controls. As in our regressions, the Civil War baseline corresponds with 1855 to 1861, while the World War I baseline extends from 1910 to 1915. The Civil War "wartime" period corresponds with 1862 to 1866, while the World War I "wartime" period extends from 1916 to 1922. Column 5 presents the change from baseline to wartime for prosthetics, while column 6 presents the change from baseline to wartime for the controls. Column 7 presents the difference between these differences.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Prosthetics		Other Mechanical		Differences		
Panel A: Civil War	Pre-Boom	Boom	Pre-Boom	Boom	Prosth. Diff	Other Diff	Diff-in-Diff
production	.159	.346	.179	.221	.187	.0423	.145
user	.27	.276	.239	.176	.00588	-.0629	.0688
cost	.0952	.247	.115	.139	.152	.0241	.128
simplicity	.0952	.333	.119	.231	.238	.112	.126
adjustability	.286	.457	.281	.286	.171	.00436	.167
appearance	.429	.247	.423	.0936	-.182	-.329	.147
comfort	.381	.531	.357	.379	.15	.0223	.128
durability	.762	.778	.746	.715	.0159	-.0314	.0473
materials	.0476	.0741	.0374	.0373	.0265	-.000149	.0266
Panel B: World War I	Pre-Boom	Boom	Pre-Boom	Boom	Prosth. Diff	Other Diff	Diff-in-Diff
production	.318	.391	.335	.352	.0737	.0171	.0565
user	.247	.24	.252	.251	-.00712	-.000115	-.00701
cost	.153	.232	.177	.219	.0789	.0419	.0371
simplicity	.353	.396	.37	.399	.0432	.0295	.0137
adjustability	.447	.546	.444	.437	.0988	-.00689	.106
appliances	.0706	.135	.0489	.0553	.0647	.00632	.0584
appearance	.224	.256	.214	.183	.0325	-.0316	.0641
comfort	.447	.329	.42	.417	-.119	-.00372	-.115
durability	.694	.758	.692	.708	.0643	.0154	.0489
materials	.0353	.0435	.0309	.0383	.00818	.00736	.000821

Appendix: Chapter 3

C1 Additional Tables and Figures

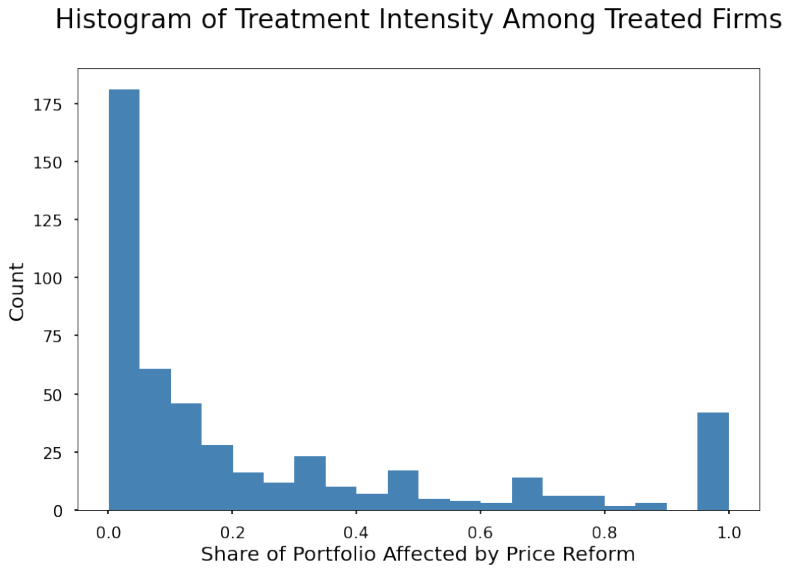
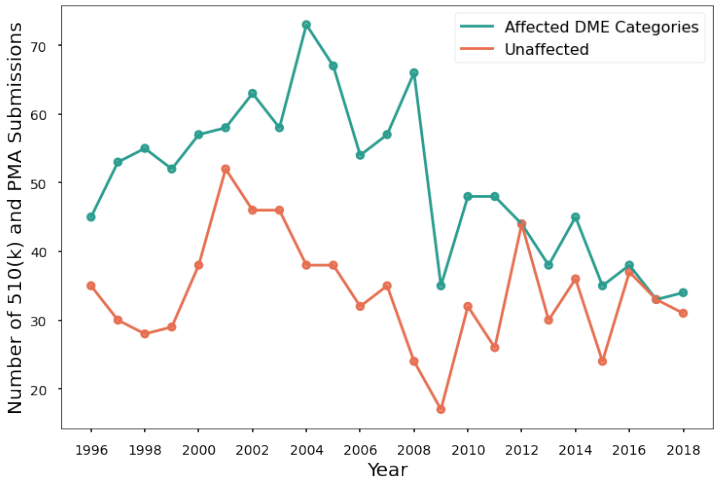
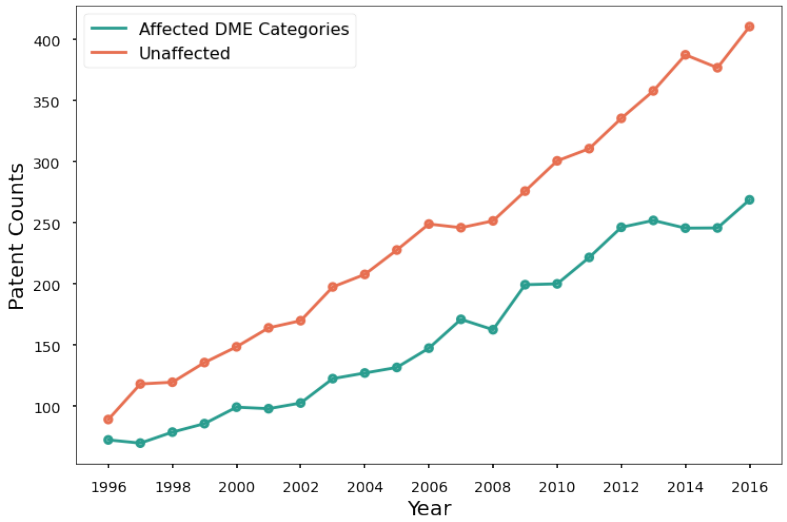


Figure C1. Percent of Manufacturers' Patent Portfolios Affected by Price Reform. Note: The figure presents a histogram depicting the distribution of firms based on different values of patent portfolio exposure to price reform. The figure includes the 486 firms in our firm-level analysis. No firms have zero exposure since all firms must have had at least one affected DME patent pre-reform. A share value of one corresponds to a 100% exposure to price reform, indicating that all of the patents held by the firm were in affected DME categories.

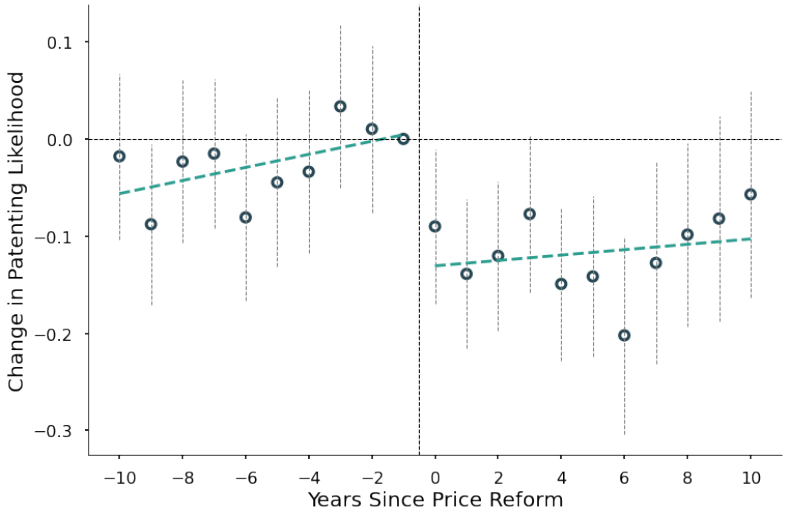


(a) Pre-Market Approval and 510(k) Counts per Year

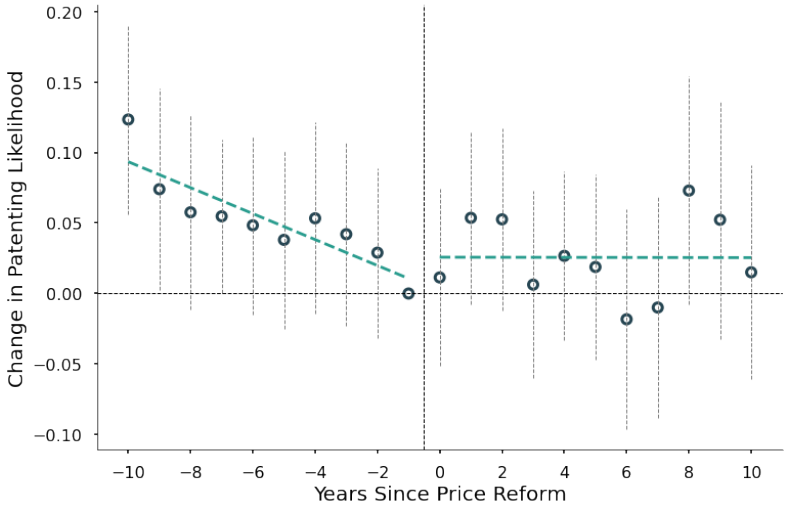


(b) Patent Counts per Year

Figure C2. Raw Trends in Innovation. Note: The figure plots the number of pre-market approvals and 510(k)s submitted per year in panel (a) and the number of patents filed per year in panel (b), separately for DME subject to the price reform and those that are not.



(a) Affected DME



(b) Unaffected DME

Figure C3. Event Study: Patents, Firm Level. Note: The figure presents the coefficients obtained from estimating equation (3.3.3) for our patenting likelihood outcome, which represents the extensive margin or the probability that a firm filed any patent in a given year. It illustrates the temporal evolution of outcomes from firms more exposed to price reform relative to those less exposed, with a reference period at $t = -1$. Panel (a) presents our event-study estimates for changes in firm patenting likelihood within affected DME categories, and panel (b) provides these estimates within unaffected categories. 95% confidence intervals are provided.

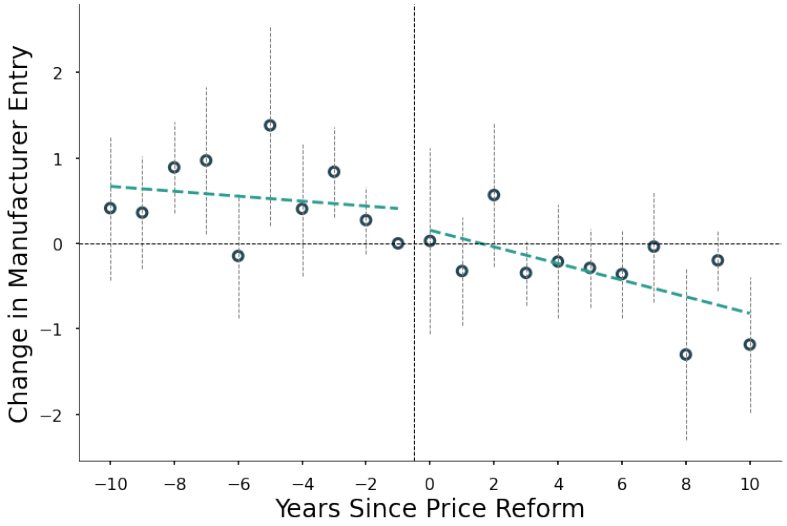
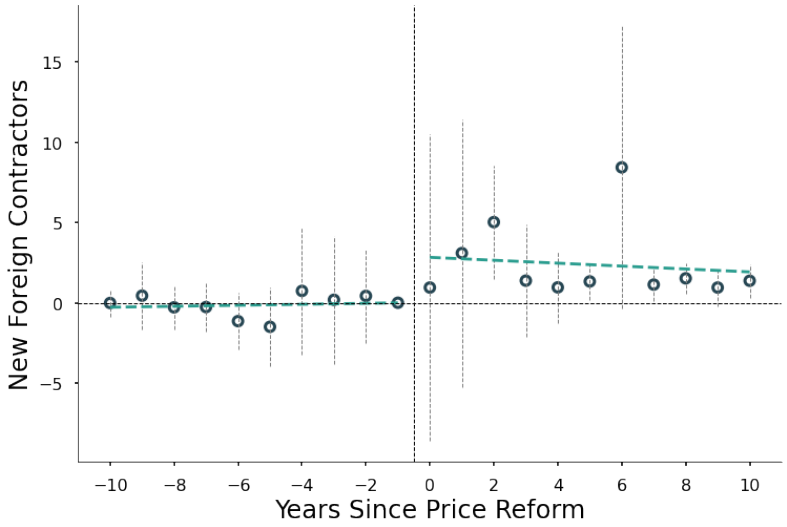
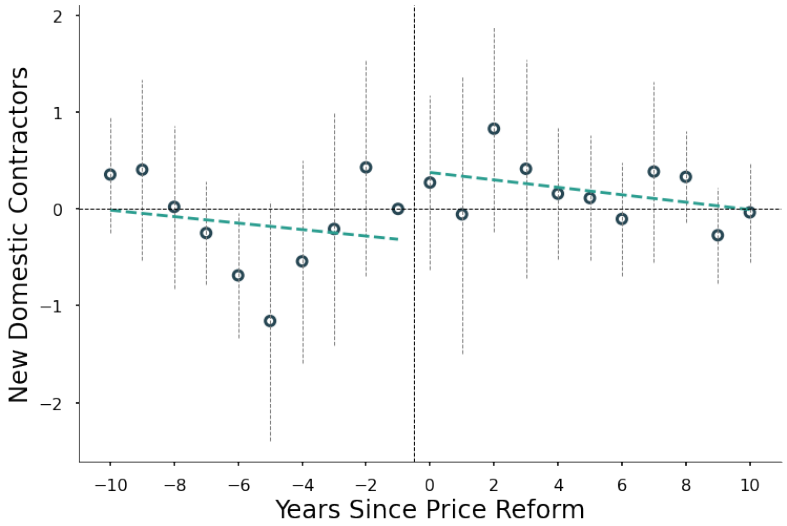


Figure C4. Event Study: Entry, DME Level. Note: The figure presents the coefficients obtained from estimating equation (3.3.1) for our firm entry results derived from FDA submissions. It illustrates the temporal evolution of entry in DME categories affected by the event, relative to those unaffected, with a reference period at $t = -1$. 95% confidence intervals are provided.



(a) Outsourcing to Foreign Contractors



(b) Outsourcing to US Contractors

Figure C5. Event Study: Outsourcing, DME Level. Note: The figure presents the coefficients obtained from estimating equation (3.3.1) for the outsourcing outcomes, separately for foreign and US contractors. It illustrates the temporal evolution of contracting rates (per year) in DME categories affected by the event, relative to those unaffected, with a reference period at $t = -1$. Panel (a) presents the event-study estimates for changes in contracting with foreign manufacturers, while panel (b) presents estimates for changes in contracting with domestic manufacturers. 95% confidence intervals are provided.

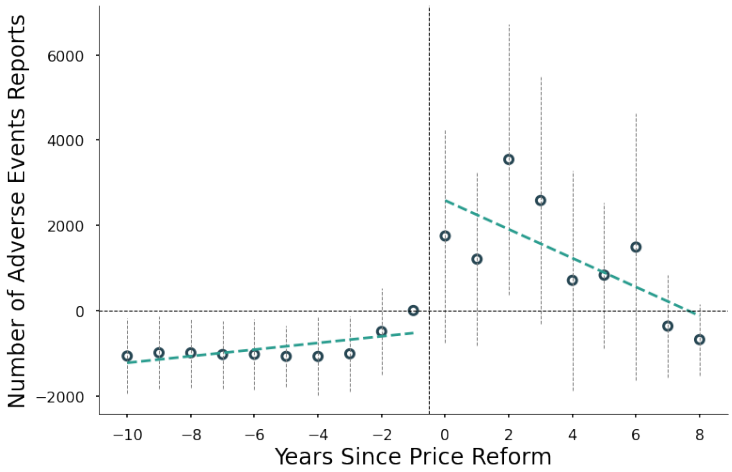


Figure C6. Event Study: Change in Product Quality – Adverse Events. Note: The figure presents the coefficients obtained from estimating equation (3.3.1) for our adverse event outcomes derived from FDA reports. It illustrates the temporal evolution of adverse event reports in DME categories affected by the event, relative to those unaffected, with a reference period at $t = -1$. 95% confidence intervals are provided.

Table C38. Impact of Price Reform on Patents Filed in US by Country of Origin. Note: The table presents results from estimating equation (3.3.2) for patents filed in the US by firm type. Column (1) reports the pre-event (before price reform) mean across treated groups. Column (2) presents the estimates, with standard errors reported in parentheses below the estimates. Column (3) shows the percent change in the outcome relative to the pre-event mean. Described are changes in patenting rates (per year) within affected DME categories relative to unaffected ones, differentiated by firm origin (i.e., US or foreign). Statistical significance is denoted by +, *, **, and *** correspond to significance levels of 0.10, 0.05, 0.01, and 0.001 levels, respectively.

	Change with Price Reform		
	Pre-Event Mean	Estimate	% Change
	(1)	(2)	(3)
Number of Patents Filed by US Firms	48.08	-15.62	-32%
		(19.89)	
Number of Patents Filed by Foreign Firms	28.23	-5.02	-18%
		(8.10)	

Table C39. Impact of Price Reform on Adverse Events by Firm Type. Note: The table presents results from estimating equation (3.3.2) for our adverse event outcomes by firm type. Column (1) reports the pre-event (before price reform) mean across treated groups. Column (2) presents the estimates, with standard errors reported in parentheses below the estimates. Column (3) shows the percent change in the outcome relative to the pre-event mean. Described are changes in adverse event rates within affected DME categories relative to unaffected ones from foreign and domestic manufacturers, with overall changes and differentiation between contractors and non-contractors. Statistical significance is denoted by +, *, **, and *** correspond to significance levels of 0.10, 0.05, 0.01, and 0.001 levels, respectively.

	Change with Price Reform		
	Pre-Period Mean (1)	Estimate (2)	% Change (3)
Adverse Events from Foreign Manufacturers			
All Events	27.5	129.3* (57.76)	470%
Contractors	5.69	178.6* (84.0)	3,136%
Non-Contractors	22.9	70.2 (50.1)	306%
Adverse Events from Domestic Manufacturers			
All Events	515.3	2034.6 (1329.9)	395%
Contractors	3.88	8.9 (71.1)	228%
Non-Contractors	510.3	1906.3 (1313.5)	374%