From Scientific Research to Healthcare Markets

Empirical Essays on the Economics of
Pharmaceutical Innovation

Dennis Byrski

Dissertation

2021

Max Planck Institute for Innovation and Competition
LMU Munich, Department of Economics

From Scientific Research to Healthcare Markets

Empirical Essays on the Economics of
Pharmaceutical Innovation

Inaugural-Dissertation
zur Erlangung des akademischen Grades

Doctor oeconomiae publicae
(Dr. oec. publ.)

an der Volkswirtschaftlichen Fakultät der
Ludwig–Maximilians–Universität München

vorgelegt von
Dennis Byrski

2021

Für meine Familie in Dankbarkeit.

Acknowledgements

This dissertation would not have been possible without the encouragement and support of many people. First and foremost, I would like to thank my advisor, Prof. Dietmar Harhoff, for his excellent guidance and for providing great academic freedom over the past years. He truly inspired me with his entrepreneurial spirit and intellectual integrity. I am also very grateful to my co-advisor, Prof. Monika Schnitzer, for her valuable advice throughout my doctoral studies. In particular, she encouraged me to think deeper about the policy relevance of my research.

I am indebted to my co-authors Stefano Baruffaldi, Fabian Gaessler, and Matthew J. Higgins who have taught and mentored me. Their in-depth knowledge and enthusiasm for Innovation Economics have been a continuous source of motivation. Their support was invaluable for the completion of this dissertation. In addition, I would like to thank my fellow and friend Felix Poege for all our stimulating conversations. He made me learn more and push harder.

I would also like to express my gratitude to the *Max Planck Institute for Innovation and Competition* for their excellent working environment. In particular, the administration, the IT team, and my student research assistants have been a great support. Furthermore, I thank the Jackson Laboratory for providing me with access to its archive. This made one of my research projects possible in the first place. I am particularly grateful to the Laboratory of Innovation Science at Harvard University for their hospitality while writing parts of this dissertation.

My special gratitude goes to my wonderful colleagues and friends at the MPI and MGSE. Our discussions and their feedback greatly improved my research. More importantly, our scientific and political debates over lunch, dinner, or after-work drinks were one of the reasons why I enjoyed going to the institute day after day. I am particularly thankful for sharing these memories with Laura Bechthold, Marc Fabel, Svenja Frieß, Corinna Hartung, Jonas Heite, Markus Nagler, Timm Opitz, Zhaoxin Pu, Myriam Rion, Michael Rose, Cristina Rujan, Stefan Sorg, Magdalena Streicher, Alexander Suyer, Gisela Stingl, and Rainer Widmann.

I am grateful to my close friends who made life so much more enjoyable and exciting. I was able to share joy and occasional despair, while I could always rely on their backing. Most of all, I thank my family. Without the unconditional love and trust of my sister, grandmother, and parents, none of this would have been possible. I dedicate this thesis to them.

Dennis Byrski
Munich, October 2021

Contents

CONTENTS

*"Progress in combating disease depends upon an
expanding body of new scientific knowledge."*

Vannevar Bush (1945, Chapter 2)

Preface

Improving health is a fundamental goal of modern societies. Healthiness and longevity, while being valuable in their own right, also have a positive impact on economic growth through human capital accumulation, investment in human capital, and direct productivity effects (*e.g.*, Bloom et al., 2019; Prinz et al., 2018; Weil, 2014). Breakthroughs in medicine have contributed significantly to progress in public health. In particular, new pharmaceutical treatments, such as anti-cholesterol drugs, antibiotics, or new cancer drugs, have accounted for a substantial increase in life expectancy and well-being (Cutler et al., 2007; Jayachandran et al., 2010; Lichtenberg, 2019). Yet, the development of new drugs is a costly endeavor. Recent estimates suggest average development costs to oscillate around \$2.5 billion (DiMasi et al., 2016). Hence, it is a first-order economic and policy concern how to effectively and efficiently promote biomedical science and pharmaceutical innovation, and how to direct them to areas of highest societal benefit.

The development of new pharmaceutical treatments depends critically on the interplay between markets and public interventions. Competitive markets may provide insufficient incentives for investments in innovation (Nordhaus, 1969). The nature of ideas as a public good causes a gap between private and social returns from innovation and induces market failures (Arrow, 1962; Nelson, 1959). Governments can address them either by lowering the private cost of innovation, so-called "push" policies, or by increasing the private return to innovation, so-called "pull" policies (for a discussion, see *e.g.*, Kyle, 2020; Lakdawalla, 2018). These incentive mechanisms aim at increasing research and development (R&D) activities in the pharmaceutical industry towards the socially desirable level.

"Push" approaches touch upon the determinants of (biomedical) scientific productivity and public science policy as input for pharmaceutical innovations (Cockburn and Henderson, 2000). These policies target the direct provision and incentivization of new scientific discoveries. A variety of empirical studies suggests a high degree of complementarity between public research and private drug development (Azoulay et al., 2019; Blume-Kohout, 2012; Toole, 2012; Ward and Dranove, 1995). More specifically, an increasing body of work identifies which inputs to knowledge production are important and how policy can influence them. This includes, for example, human capital (Azoulay et al., 2010), funding (Jacob and Lefgren, 2011; Myers, 2020), as well as access to and diffusion of research tools (Furman and Stern, 2011; Murray et al., 2016). If the scientific domain is effective in producing high-quality research,

it can translate into valuable and novel commercial applications like new drugs (Ahmadpoor and Jones, 2017; Poege et al., 2019; Watzinger and Schnitzer, 2019).

The need for "pull" approaches arises from the very long nature and high upfront investment costs of the drug development process (Adams and Brantner, 2006). Patents, protecting a drug against unwanted imitation, are an effective and the most widely applied policy tool (Cohen et al., 2000; Mansfield, 1986). However, there is emerging economic literature discussing the inefficiencies of the patent system related to patent characteristics (Budish et al., 2015), follow-on innovation (Gaessler et al., 2019; Galasso and Schankerman, 2015; Hall and Harhoff, 2012; Sampat and Williams, 2019; Williams, 2013), and patent proliferation (Sternitzke, 2010). These recent contributions have created growing concerns about whether societies should primarily rely on patents to incentivize drug development. Thus, the toolkit of government interventions includes further supporting institutions, *e.g.*, policy-induced expansions to demand (Blume-Kohout and Sood, 2013; Finkelstein, 2004), innovation prizes such as transferable market exclusivities or advanced market commitments (Batista et al., 2019; Kremer and Williams, 2010), data exclusivity (Gaessler and Wagner, 2020), and competition policy (Cunningham et al., 2021; Higgins and Rodriguez, 2006). It is, however, unclear *ex-ante* which of these policies are effective and efficient. This remains an empirical question to be answered.

This dissertation sheds light on three stimuli to biomedical science and pharmaceutical innovation, each in a self-contained chapter. The first chapter investigates the functioning of research tool markets. These are important input factors into science, and, hence, a fundamental determinant of "push" approaches. Specifically, it observes why short-term distortions to supply have enduring effects on tool adoption and, thus, the direction of scientific research. The second chapter examines whether pharmaceutical companies adjust their follow-on innovation activities when patentability standards increase. To this end, it analyses changes in the innovation incentives caused by a drug's marketing authorization. The third chapter explores the relationship between downstream shifts in demand and upstream research. It studies whether a policy-induced increase in market size, considered as "pull" approach, affects scientific publishing by universities and corporations.

Chapter 1, in joint work with Stefano Baruffaldi and Fabian Gaessler, examines the functioning of research tool markets. Exploiting an unforeseen negative supply shock to laboratory mice in 1989, the study shows that new tools may fail to diffuse widely due to path dependency created by switching costs.

New research tools are central to scientific and technological progress (Mokyr, 2002). Institutions, often supported by the public, and markets increasingly play a role in their production and diffusion (Furman and Stern, 2011; Walsh et al., 2007). They embody positive feedback loops that are conducive to the continuous accumulation of knowledge (Mokyr, 2002). Hence,

gaining a better understanding of the mechanisms that govern the adoption of research tools is of particular importance due to their role in the process of knowledge production (Stephan, 2012). However, upfront investments and uncertainties create switching costs that scientists face when adopting new research tools (Klemperer, 1987). This has important implications for the functioning of such markets due to path dependency: new tools may fail to diffuse widely, whereas old tools remain dominant, leading to suboptimal market equilibria (David, 1985; Dosi, 1982). Under these circumstances, short-term changes in the costs of adoption can lead to long-term changes in demand.

To explore whether the existence of these frictions undermines the functioning of research tool markets, we study the consequences of a negative supply shock on the use of research tools and the production of scientific knowledge. For this purpose, we leverage a natural experiment in the market for research tools and exploit the Morrell Park fire at the world's largest mice breeding facility, the Jackson Laboratory (JAX), in 1989. It killed approximately 400,000 mice and caused a substantial but temporary supply shortage in certain mice strains – essential tools for biomedical research (Murray et al., 2016) – for the scientific community.

We identify from JAX *archival* records mice strains that were in short supply in the aftermath of the fire. To quantify their use by the research community, we link each strain to scientific publications. This allows us to trace the use of JAX strains as well as of identical mice strains from other suppliers. To study the consequences of the shock at the individual level, we identify scientists exposed to the supply shortage and construct their full publication history. At both mice and scientist level of the analysis, we deploy Difference-in-Differences estimations, comparing groups with different levels of exposure to the supply shortage.

We find that the fire-induced supply shortage, albeit temporary, had long-lasting consequences on the use of mice strains. The use of affected JAX mice strains declined relative to both spared JAX strains and strains provided by other suppliers. In contrast, the adoption of spared JAX strains appears to gradually increase in the period after the JAX reconstruction. These effects are explained by those mice where pre-fire switching costs were presumably higher. We find corroborating evidence for the proposed mechanism in our analysis at the scientist level. Scientists affected by the supply shortage are more likely to use spared mice strains relative to comparable scientists that were not affected by the fire. This adoption persists even years after the fire. The affected scientists' productivity is temporarily compromised, as captured by a decrease in annual research output. This suggests some initial switching costs when adopting the substitute strain. Since strains are imperfect substitutes, we find that adoption of different mice, as a consequence of the shock, leads to durable changes in the scientists' direction of research.

The study contributes to the literature on the inputs to knowledge production. Scholars have initially focused on human capital and funding (*e.g.*, Azoulay et al., 2010, 2018; Oettl, 2012; Jacob and Lefgren, 2011). More recently, a literature stream on research tools has emerged showing that access to and diffusion of research tools enables cumulative knowledge

production (Furman and Stern, 2011), leads to exploration in research (Furman and Teodoridis, 2020; Murray et al., 2016), and influences the composition of teams (Teodoridis, 2017; Zyontz, 2019). Other studies have looked at the effect of negative shocks on physical capital per se (Baruffaldi and Gaessler, 2021; Waldinger, 2016) and find medium- to long-term effects on the rate and direction of research. To our knowledge, this study is the first to investigate the nature of markets for research tools with evidence based on an exogenous supply shock. Our results provide first evidence on tool-inherent frictions in the market for research tools.

Evidence across all these studies has important policy implications. It suggests that the institutional arrangements that determine the allocation of investments, the level of standardization, and the distribution of risks associated with the development and access to research tools have important consequences for downstream scientific activities. The findings of our study provide a further empirical justification for the support of institutional, industry, and community level efforts to steer research tool markets towards desirable trajectories.

Chapter 2 investigates how drug approval affects follow-on innovation activities. Leveraging variation in the length of time from patent filing until a drug's approval, it finds innovation activities conducted to prevent generic entry to decrease after the marketing authorization.

Investments in follow-on innovation are characteristic of the pharmaceutical industry (Kyle, 2020). Some of these innovations, defined as *"improvement innovations"*, provide a meaningful therapeutic benefit (Arcidiacono et al., 2013; Bokhari and Fournier, 2013). Others are argued to be trivial modifications to the original drug conducted to delay generic entry by expanding patent protection, defined as *"enforcement innovations"* (Amin and Kesselheim, 2012; Frakes and Wasserman, 2020; Gurgula, 2020; Sternitzke, 2010, 2013). Given that a longer market exclusivity allows the originator company to earn supracompetitive profits (Budish et al., 2015), there is an ongoing economic and political discussion about the net consumer surplus derived from incremental innovations in the pharmaceutical industry (Yin, 2017).

The event of drug *approval* is of capital importance in a drug's life cycle. The transition from pre-approval to post-approval may have two countervailing effects on follow-on innovation activities: on the one hand, entering the market should stimulate originators to protect the approved drug from competition and to improve its features (European Commission, 2009; Sternitzke, 2010, 2013). On the other hand, the marketing authorization impedes the enforceability of follow-on patents that are filed after approval by increasing the patentability standards. In turn, this decreases the incentives of originators to invest in enforcement innovations. The main reason is that (trial-related) information disclosed in the course of the drug approval can render an invention obvious and serve as novelty threatening prior art (Breckenridge and Jacob, 2019; Kallenbach and Vallazza, 2018; Mello et al., 2013). Thus, this study investigates how drug approval affects the two types of follow-on innovation activities.

To this end, I collect novel data on pharmaceutical patents that combines detailed information on the focal inventions with follow-on innovations and data on marketing authorizations.

I exploit the fact that follow-on innovations can be observed through patent citations (Galasso and Schankerman, 2015) and distinguish between those likely related to enforcement, *e.g.*, "secondary" patents and patents in the same field of application, and those likely related to improvements, *e.g.*, "product" patents and patents in different fields of applications. The empirical strategy leverages differences in the timing of drug approval relative to patent filing (*time to approval*). This relies on the assumption that the timing of marketing authorization is to some degree independent of drug quality and commercial value. I provide evidence that inventions with early and late marketing authorization do not differ in ex-ante patent or drug characteristics and exhibit parallel pre-approval forward citation trends.

I find an immediate and long-lasting decrease in enforcement innovation activities after the marketing authorization. This decrease in innovation is not only associated with the focal firm, but also with third parties, who are likely in a vertical relationship with the originator. Improvement innovation activities instead remain unaffected by the drug approval. The immediate timing of the decrease in enforcement innovation is consistent with the interpretation that the missing innovations are of marginal value. Moreover, I do not find similar effects on enforcement innovation for prior milestones in the drug development process that are not related to increases in patentability standards. These findings can be interpreted as corroborating evidence for the enforceability mechanism. Marketing authorization-related increases in patentability standards delineate incremental innovation activities: pharmaceutical companies are self-adjusting their innovation behavior so that investments in improvement become relatively more important than investments in enforcement with little therapeutic benefit.

The findings contribute to the literature on *patent fencing* and *evergreening* in the pharmaceutical industry (*e.g.*, Abud et al., 2015; Hemphill and Sampat, 2012; Sternitzke, 2010). I provide evidence that secondary pharmaceutical patenting becomes less likely with drug approval. This is embedded in the broader literature on the incentives for innovation in the pharmaceutical industry. While the existence of intellectual property rights as an effective "pull" policy to encourage innovation is well understood (Gaessler and Wagner, 2020; Kyle and McGahan, 2012), I show that enforcement type innovation activities are responsive to subtle changes in the patentability standards, *e.g.*, due to drug approval.

This has important policy implications. Some scholars advocate "fixing" the existing patenting system concerning patent proliferation and secondary patents. For example, Frakes and Wasserman (2020) propose to empower patent examiners to properly investigate the prior art of secondary pharmaceutical patents and, thus, to increase patentability standards. This study supports their idea: if patent examiners had access to all trial information submitted for drug approval, this would increase their ability to identify secondary patent filings of low quality. Eventually, self-enforcement may lead pharmaceutical companies to decrease enforcement innovation activities and resulting secondary patent filings themselves.

Chapter 3, in joint work with Fabian Gaessler and Matthew J. Higgins, sheds light on demand-driven incentives in basic science. Exploiting quasi-experimental variation in market size introduced by Medicare Part D, we show that while drug development appears to respond to downstream shifts in market demand, upstream scientific research fails to do so.

Dating back at least to Schumpeter (1939), scientists have been exploring the factors that drive innovation. As aforementioned, some scholars suggest that both supply-side ("push") and demand-side ("pull") factors might shape the pattern of investments in innovation. Prior literature has established a link between market size and innovation, especially in the pharmaceutical industry (Acemoglu and Linn, 2004; Blume-Kohout and Sood, 2013; Finkelstein, 2004). These studies, however, almost universally refer to the traditional rubric of "development" activities, such as clinical trials or new drug approvals, as opposed to "research", such as biomedical science. Efforts to extend this linkage back to upstream research had limited success. Acemoglu and Linn (2004) and Finkelstein (2004), for example, do not find a relationship between demographic-driven or policy-driven expansions in market size and patenting. Bhattacharya and Packalen (2011) establish a link to research, but cannot disentangle disease prevalence from profit incentives. Thus, the broader link, if it exists, between market pull incentives and scientific research remains elusive.

In this study, we rely on a major policy intervention in the context of U.S. drug prescriptions that affect a wide range of therapeutic areas and diseases: the 2003-introduction of Medicare Part D, which substantially increased the demand for drugs more relevant for the elderly in the U.S. (Blume-Kohout and Sood, 2013; Dranove et al., 2020). We build novel data that combines all U.S. biomedical and life sciences publications, patent-paper links, and drug development efforts mapped to disease categories. To this end, we employ a verified crosswalk between publicly used disease codes and a controlled vocabulary to index publications (Bhattacharya and Packalen, 2011). For each scientific publication, we add bibliographic information, which allows us to accurately categorize research activities across universities and corporations as well as to differentiate the type of research. Moreover, we map publications to patents to approximate whether scientific research was recognized in commercially relevant applications (Marx and Fuegi, 2020). We replicate prior findings on drug development, sales, prices, and revenues to verify our data construction. To examine the effects of quasi-experimental variation in market size introduced by Medicare Part D in 2003 on upstream research, we employ a Difference-in-Differences estimation. The exposure to Medicare Part D is calculated based on the pre-treatment share of Medicare patients among the total population for each disease group (*e.g.*, Duggan and Scott Morton, 2010). We account for demographic changes, public research funding, and new research opportunities. The pre-2004 trends in scientific publications are remarkably similar.

Over a decade following the implementation of Medicare Part D, we find no evidence of an overall relationship between market size and biomedical science. Effect sizes are substantially smaller than any effect on drug development activities found in the prior literature and

in our own replication. However, we illustrate substantial effect differences by the type of affiliation. Any demand response is concentrated only among corporate research, and decreases in magnitude by distance to the market (*e.g.*, universities). More precisely, our results show that Medicare Part D caused an increase primarily in corporate affiliated publications linked to both clinical trials and pharmaceutical products, which are residuals from drug development activities. Consistently, we do not find any causal relationship between any type of research and market size for universities. Moreover, we observe that only in the years directly following the Part-D enactment, greater exposure to Medicare Part D caused an initial increase in corporate-affiliated patent-weighted research. Our back-of-the-envelope calculation suggests that an expansion of market size by \$43 million would only lead to one additional scientific publication. These findings demonstrate a disconnect between the incentives for investments in innovation in the private sector and the incentives for the production of scientific knowledge.

The study has important implications for firms and policymakers. The pharmaceutical industry is highly dependent upon the external market for technologies (Higgins and Rodriguez, 2006), with much of that research emanating from universities (Cockburn and Henderson, 2000). While drug development (*i.e.*, drugs in clinical trials) appears to respond to downstream shifts in market demand, our results show that upstream research, especially at universities, fails to do so. Firms face the prospect that the flow of research may not meet the kind of development needs, they require. This disconnect suggests that more active policy intervention may be needed. Policymakers may want to complement a market expansion with early-stage research incentives, such as public research funding, which is known to be an effective tool in fostering scientific research (Azoulay et al., 2019).

In summary, this dissertation offers new insights into the determinants of biomedical science and pharmaceutical innovation. Evidence from these micro-economic analyses may contribute to designing effective and efficient public policies that help stimulate R&D activities, foster the development of new pharmaceutical treatments, and eventually improve public health.

<div style="text-align: right; font-size: 4em;">**1**</div>

Fire and Mice

The Effect of Supply Shocks on Research Tool Adoption

1.1 Introduction

New research tools have often been central to scientific and technological progress (Mokyr, 2002). While scientists create research tools for their own needs (Franzoni, 2009; Rosenberg, 1992), institutions have become increasingly important in their provision and diffusion. This external supply is advantageous because it lowers costs due to economies of scale, reduces uncertainty thanks to higher standardization, and fosters collaboration and knowledge sharing (Furman and Stern, 2011; Walsh et al., 2007). A functioning market for research tools, therefore, embodies positive feedback loops between science and technology that are conducive to the continuous accumulation of knowledge (Mokyr, 2002) and, ultimately, economic growth (Romer, 1990). Hence, gaining a better understanding of the mechanisms that govern the supply and demand of research tools is of particular importance (Stephan, 2012).

We posit that switching costs for scientists influence the adoption of research tools and bear important implications for the functioning of such markets (Klemperer, 1987). First, the adoption of research tools requires investments in specific tacit knowledge and complementary assets (Zyontz, 2019). Second, their full potential and range of applications is initially unknown and changes dynamically over time. These upfront investments and uncertainties create switching costs that scientists face when adopting new research tools. This generates path dependency: past adoption determines the continued use of tools, irrespective of their present characteristics relative to available alternatives. New tools may fail to diffuse widely whereas old tools remain dominant, leading to suboptimal market equilibria (David, 1985; Dosi, 1982; Greve and Seidel, 2015; Huckman, 2003). Under these circumstances, short-term

---

This chapter is based on joint work with Stefano Baruffaldi and Fabian Gaessler.

changes in the costs of adoption can lead to long-term changes in demand. For example, a temporary supply shock to some research tools may lead to the permanent adoption of substitutes.

In this study, we investigate the consequences of a negative supply shock on the use of research tools and the production of scientific knowledge. For this purpose, we leverage a natural experiment in the market for research tools. More specifically, we investigate a fire at the world's largest mice breeding facility, the Jackson Laboratory (JAX), in 1989. This fire caused a substantial but temporary supply shortage in certain mice strains – essential tools for biomedical research – for the scientific community. We first study how this shock affected the use of different mice strains in the long run. We then analyze whether changes in usage are due to the adoption of different mice strains by scientists exposed to the supply shortage. Finally, we test whether the adoption of different mice strains, which represent imperfect substitutes, comes with switching costs, reflected by changes in the scientists' research productivity and research direction.

The empirical setting is ideal to study the dynamics of research tool markets. First, laboratory mice are an iconic example of research tools (Stephan, 2012; Murray et al., 2016). They are complex biological models whose development and use is accompanied by uncertainty. While self-development is an option, their breeding requires care and effort (The Jackson Laboratory, 2009a). Second, the fire was an unforeseen event that led to a substantial but temporary supply shortage.[1] This allows us to overcome endogeneity issues, which are a common feature in the literature on research tool adoption. Third, the supply shortage applied to some distinct mice strains, while the supply of other JAX mice strains remained unaffected. Moreover, some strains remained available from alternative commercial suppliers. This creates substantial heterogeneity. Fourth, despite these relative changes in supply, JAX, as a non-profit organization, did not adjust prices.

From JAX archival records, we identify mice strains that were in short supply in the aftermath of the fire. To quantify their use by the research community, we link each strain to scientific publications based on information from the Mouse Genome Database and by textual search in the Scopus database. This allows us to trace the use of JAX strains as well as of identical mice strains from other suppliers. We further characterize each mice strain by the extent to which its genetic code was understood at the time of the fire, as a direct proxy of the level of uncertainty associated with their use. To study the consequences of the shock at the individual level, we identify scientists exposed to the supply shortage and construct their full publication history. Both at mice and scientist level of analysis, we deploy Difference-in-Differences (DiD) estimations, comparing groups with different levels of exposure to the supply shortage. At scientist level, we use Coarsened Exact Matching (CEM) to account for observable differences between researchers, such as the career life cycle.

---

[1]Thanks to large public investments, JAX returned to full capacity in two years' time.

1. FIRE AND MICE

We find that the fire-induced supply shortage had lasting consequences on the use of mice strains. The use of affected JAX mice strains declined relative to both spared JAX strains and strains provided by other suppliers. This decline shows no sign of convergence to pre-fire levels even a decade after the fire (8 years after supply was fully restored). In contrast, the use of unaffected JAX strains appears to have gradually increased in the period after the JAX reconstruction. Heterogeneity analyses show that these effects are stronger when the mice supply shortage induced larger switching costs for scientists. In particular, the long-term decline is primarily driven by affected mice strains with low pre-fire JAX sales, not available from other commercial mice providers, and with a higher level of uncertainty associated with their use. For these mice, scientists are more likely forced to incur larger switching costs to cope with the temporary supply shortage, and, at the same time, face higher costs in case they intend to switch back to the previously used strains.

We find corroborating evidence for the proposed mechanism in our analysis at the scientist level. Scientists affected by the supply shortage are more likely to use spared mice strains relative to comparable scientists that were not affected by the fire. This adoption persists even years after the fire. The affected scientists' productivity is temporarily compromised, as captured by a decrease in annual research output. This suggests some initial switching costs when adopting the substitute strain. Moreover, we find that the adoption of imperfect substitutes, as a consequence of the shock, leads to durable changes in the scientists' direction of research. In particular, their research output seems less related to their previous work, as the scientists make fewer self-references, and more applied, as the scientists publish more often in clinical research journals and receive more citations from patents. Consistent with our findings, further analyses show that scientists with a relatively higher degree of dependence on mice in prior research, and, thus, higher switching costs, adopt new strains at lower levels.

To our knowledge, our paper is the first to investigate the nature of markets for research tools with evidence based on an exogenous supply shock. The existence of significant switching costs implies that short-term distortions regarding the accessibility of research tools can have long-lasting effects on their adoption. Our results provide first evidence on tool-inherent frictions in the market for research tools. Thus, our paper contributes to the literature on the inputs to knowledge production. Scholars have initially focused on human capital and funding (*e.g.*, Azoulay et al., 2010, 2018; Oettl, 2012; Jacob and Lefgren, 2011). More recently, a literature stream on research tools has emerged. Related studies have shown that access to and diffusion of research tools enables cumulative knowledge production (Furman and Stern, 2011), leads to exploration in research (Furman and Teodoridis, 2020), influences the composition of teams (Teodoridis, 2017; Zyontz, 2019), and is highly sensitive to scientists' search strategies (Kolympiris et al., 2021). Closest to our setting, Murray et al. (2016) study the effect of IP rights on the use of specific new mice technologies, finding both increased use and exploration of areas of application when such exclusion rights are removed. The findings of

11

our study provide a further empirical justification for the support of institutional, industry, and community level efforts to steer research tool markets towards desirable trajectories.

Other studies have looked at the effect of negative shocks – of varying intensity – to physical capital per se and research tools in particular on knowledge production, with mixed evidence. Waldinger (2016) finds that damage to universities during World War II had more transitory consequences than the departure of scientists. Hill (2019) finds transient productivity consequences of adverse weather conditions during telescope viewing sessions in astronomy. Baruffaldi and Gaessler (2021) look at adverse events causing damage in research laboratories and find that the loss of generic capital has transitory consequences, while the loss of specific, self-developed capital can have long-term effects on the rate and direction of research of scientists. Our findings are consistent to the extent that scientists suffered at most a delay in accessing (imperfectly) substitutable mice strains, for which they did not sustain development costs. In general, evidence across these studies suggests that the institutional arrangements that determine the allocation of investments, the level of standardization, and the distribution of risks associated with the development and access to research tools have important implications for downstream scientific activities and deserve further attention in future research.

The remainder of the paper is organized as follows. In the next section, we describe the empirical setting related to the 1989 Morrell Park fire at JAX. Subsequently, we introduce the data, our empirical strategy, and selected descriptive statistics. We then present our main findings at the mice and scientist level and conclude with a brief discussion and outlook.

1.2 Empirical Setting

1.2.1 Laboratory Mice as Research Tools

Since 1902, when the French biologist Lucien Cuénot tried to test Mendel's Laws of Inheritance for the first time on an animal (Cuénot, 1902), researchers have used laboratory mice to study the inheritance of genes and their relationship to diseases. Laboratory mice are the most widely used mammal in biomedical research as they share many genes with humans (Paigen, 1995). Two features of laboratory mice are noteworthy. First, mice differ with regard to their strain-specific genetic profile, which in turn determines their suitability as research tool for a particular disease (The Jackson Laboratory, 2009b). Table A-5 provides examples. Moreover, mice strains are likely to change their traits as a response to breeding conditions and differ across narrow sub-strains and vendors (Nevalainen, 2014; Bryant, 2011). This creates imperfect substitutability between strains within one research project (The Jackson Laboratory, 2009b). Second, the use of laboratory mice in research is costly. A single laboratory may need more than 1,000 mice per year, with unit prices ranging between \$20 and \$60 (Stephan, 2012).

1.2.2 The Jackson Laboratory

Clarence Cook Little, known for having developed the first inbred mouse strain, founded the Jackson Laboratory (JAX) in 1929 as a non-profit institution committed to research on and the provision of laboratory mice.[2] Buying mice from an external supplier, such as JAX, had several benefits for scientists. First, it was cheaper and less time-consuming than self-breeding. Second, given the high breeding standards at JAX, scientists could minimize the genetic variance between mice from the same strain, which is necessary for controlled and replicable experiments (Malakoff, 2000; The Jackson Laboratory, 2009b).

By 1989, JAX had become the leading research institution for laboratory mice, serving two million inbred and mutant mice annually to scientists in over 11,000 laboratories (JAX Archive-1).[3] With 1,700 different strains, it then accounted for 21% of all genetically defined mice used in the United States. JAX mice were used for research on a variety of diseases (*e.g.*, AIDS, cancer, diabetes, and neurological dysfunctions) and by a large number of institutions. About 51% of the customers were universities/medical schools, and 27% came from industry. JAX supplied mice globally (9% of JAX customers were abroad), but had a clear focus on the United States and Canada (see Table A-2 in the Appendix).

1.2.3 The 1989 JAX Fire and its Effect on Mice Supply

On May 10, 1989, a fire destroyed parts of the JAX production facilities in Bar Harbor, Maine, killed 400,000 mice, and reduced production capacity by more than 50%.[4] Since foundation stocks of the inbred strains were kept at a different location, no mice strain was permanently lost (JAX Archive-4).

The fire affected mice production at JAX in two ways. First, the physical capacity to rear mice was largely compromised and, second, the existing stocks of several mice strains intended for shipment and breeding were severely reduced. The drop in JAX's mice production capacity (JAX Archive-10) had direct consequences on mice supply. Sales dropped from 475,016 mice in the pre-fire quarter of 1989 to 220,988 mice in the first quarter of 1990 (Figure 1.2b). Notably, the impact of the fire was not uniform across all mice strains. The extent to which the production of a particular mice strain was reduced largely depended on the exact location of the respective mice cages in the main building. Figure 1.1 illustrates the microgeographic discontinuity of damage in the production facilities. As a result, the supply of a few strains remained largely unscathed with their unit sales in 1989 almost as high as in the previous year (see Figure 1.2a). For example, one of the best-selling strains of JAX at that time, the C57BL/6J strain – called B6, suffered a sales drop of only about 20%. In contrast, many other

---

[2]Inbred strains need to be maintained by continued sibling mating based on foundation stocks to avoid genetic drift (The Jackson Laboratory, 2009b). This is a knowledge intensive and highly specialized endeavor.

[3]Documents directly collected from the JAX archive (The Jackson Laboratory Archives, 2012a,b) are cited as "JAX Archive-XY". Table A-1 in the Appendix provides an overview of the collected documents.

[4]The blaze resulted from the ignition of "flammable vapors from adhesive[s]" being used in the room of origin (JAX Archive-3).

inbred and mutant strains (*e.g.*, the BALB/cJ strain) were subject to serious supply shortages, as their sales fell by up to 95% relative to the previous year (JAX Archive-5).[5] JAX estimated that the fire significantly interfered with mice-dependent biomedical research to the extent of about one billion dollars in annual research money (JAX Archive-1).[6]

Figure 1.1: Fire damage in the JAX production facilities
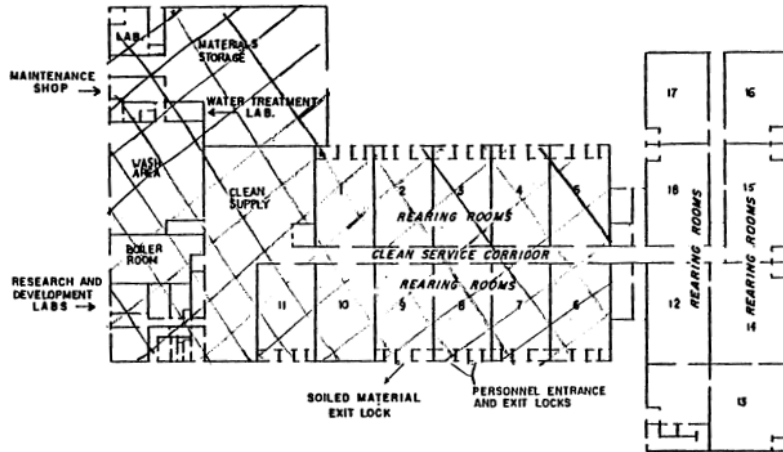
Notes: This figure illustrates the floor plan of the JAX main facilities (JAX-3). The striped area (25,895 square feet) was destroyed by the fire. The rooms in the blank area were largely spared from the fire.

Figure 1.2: Mice production capacity and sales at JAX

(a) Sales decline at strain level

(b) Production capacity over time
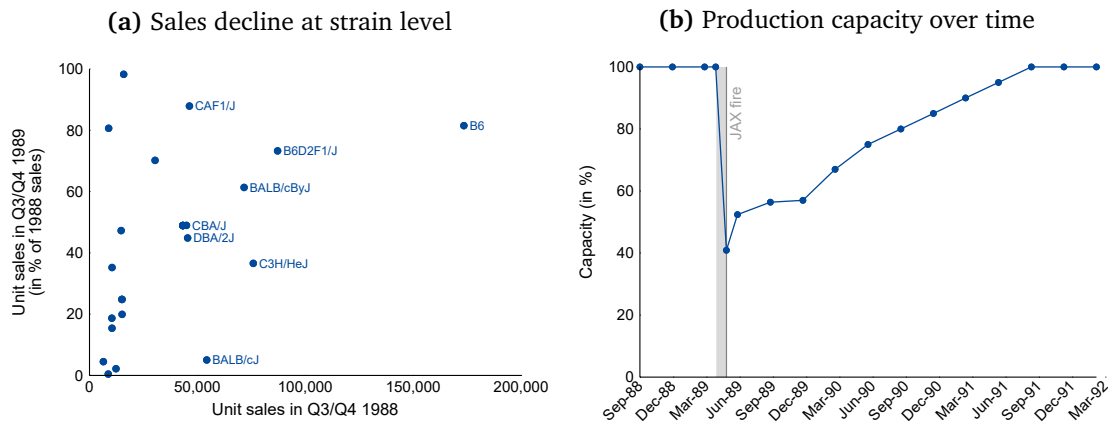
Notes: Figure 1.2a plots the 25 best selling JAX mice strains by their unit sales in the second half of 1988 and in the second half of 1989 (as a share of the 1988 sales). For reasons of clarity, only mice with 50,000 or more unit sales are labeled. Figure 1.2b illustrates the breeding capacity at JAX before and after the fire. Numbers from June 1990 onward are forecast numbers provided by JAX (JAX Archive-10, JAX Archive-11, JAX Archive-12).

---

[5]Hybrid stocks and highly specialized research mice remained rather unaffected since both were reared in the annex or main research building (JAX Archive-5).

[6]Expenses for JAX's own research, however, remained constant in the post-fire years (JAX Archive-6). In total, the fire resulted in an estimated loss of over $40 million, of which the accumulated operating deficits over the two-year recovery period were about $9 million (JAX Archive-2).

1. FIRE AND MICE

Quickly after the fire, JAX rolled out a recovery plan that involved both the construction of emergency production space and the restoration of the damaged production facilities. For this purpose, JAX initially used capital from insurance funds and private donations. In parallel, JAX undertook substantial political efforts to obtain further funds (JAX Archive-8).[7] In August 1990, JAX received an extramural NIH grant for the construction of new mouse breeding, production, and support facilities (JAX Archive-7).[8] This additional funding helped JAX return to pre-fire production levels in mid-1991, without substantially raising the unit prices of their mice.[9] Figure 1.2b illustrates the steady increase in supply due to the re-breeding and restoration efforts.

Despite its dominant position, JAX was not the sole supplier in the market for laboratory mice. About ten U.S.-based commercial suppliers provided an alternative source for about 15 to 20 mice strains (JAX Archive-2).[10] The offered strains were some of the most widely used strains (JAX Archive-13). Facing capacity constraints themselves, the commercial suppliers responded to the supply shortage at JAX with price increases of about 70-80% (JAX Archive-14).[11] We did not find any evidence in the historical records suggesting that JAX was favoring specific customers in need.

Qualitative accounts suggest that the temporary supply shortage of particular mice strains had a substantial impact on the demand side. The results of a representative survey of JAX customers conducted in late 1989 suggest that more than half of the respondents had been negatively affected by the supply shortage. These researchers partly switched to either self-breeding or alternative mice, with a negative effect on their research (JAX Archive-17). These findings are consistent with the consequences reported by NIH researchers relying on JAX mice. They experienced delays in research projects due to the necessary search for alternative mice strains (JAX Archive-16). Moreover, case studies conducted by JAX suggest that scientists indeed reorganized their research (JAX Archive-18):

- "If The Jackson Laboratory is unable to supply the strain and number of mice Dr. ▮▮▮▮ requires, she will have to switch projects and perhaps use different animals."

- "They would have to begin a massive breeding program on their own. However, because of the extremely high costs involved, they would eventually phase out projects requiring large numbers of mice [...]."

- "Instead of pursuing many lines of research simultaneously, she is focusing on one or two areas where the missing or hard-to-obtain strains are not required."

---

[7] For instance, JAX orchestrated more than 1,500 support letters from scientists all over the U.S. addressed to their local political representatives to lobby for extramural NIH funding.

[8] In 1991, JAX successfully applied for further funds to replace the temporary facilities with permanent, state-of-the-art premises (JAX Archive-9).

[9] Price increases were about 10% on average (JAX Archive-15, based on own calculations).

[10] Commercial suppliers of laboratory mice in the U.S. at that time included Charles River, Harian Sprague Dawley, Simonsen, Dominio, Bantin & Kingman, Taconic, Hilltop, Sasco.

[11] Even in combination, the commercial suppliers had a production capacity merely half as large as the one of JAX.

1.3 Data and Summary Statistics

In the following, we first describe the dataset, in which we link archival data from JAX with bibliometric information from publication databases. Subsequently, we outline our empirical strategy exploiting the 1989 Morrell Park fire as a plausibly exogenous shock to the availability of particular mice strains for the scientific community.

1.3.1 Mice Strain Level

Identifying Affected and Spared Mice Strains

We first determine which mice strains were affected by the fire. For this purpose, we make use of collected fire recovery files, price lists, and personal memorabilia from JAX. These documents contain information on mice strains that were in actual short supply (JAX Archive-19), in great demand (JAX Archive-20), only available with substantial delay (JAX Archive-21), or experienced a substantial loss in sales (JAX Archive-22). Other documents list several mice strains that remained unaffected by the fire and, thus, were in good supply (JAX Archive-23, JAX Archive-22).[12] In sum, we have sufficient information to classify 84 strains. Among these, we consider 39 strains as affected and 45 strains as unaffected.[13]

Mice Usage Data

We use scientific publications to measure the use of different mice strains in research. Although scientists typically mention the used mice in the publication, this does not happen in a systematic and structured manner. Scientists may not all use the same name for a given strain (but synonyms) and may mention the strain in different sections of the article. We rely on two approaches to overcome the challenge of associating papers with distinct mice strains. First, we draw on curated data from the *Mouse Genome Informatics* database. Second, we search for mice strains in the title, abstracts, and keywords in *Scopus*.[14]

We make use of the Mouse Genome database, which is freely accessible and administered by JAX (Blake et al., 2003). This database provides aggregate publication information on laboratory mice and their profiles to facilitate the study of human health and diseases. As the standard nomenclature of mice strains changes over time, we use the JAX order number, which is a unique and stable identifier of research mice in the MGI database.[15] We collect all

---

[12]Example documents can be found in Appendix A-1 and A-2.

[13]Table A-3 in the Appendix lists the classified strains and provides further details.

[14]A different approach is used by Murray et al. (2016), who use citations to seminal mice papers to proxy the use of a strain in publications. This approach is unfeasible for our set of strains: most inbred strains do not have one single seminal publication, and some strains were introduced in periods for which bibliographic data coverage is limited (*e.g.*, B6 was developed in 1920).

[15]To identify the JAX order number, we link the mice strains mentioned in the historical fire documents by name to the 1989-price list.

references in MGI for each mice strain over time. We find at least one pre-fire publication for 35 affected and 32 spared strains as referenced in the MGI data.

The MGI data does not claim full coverage for all mice strains and all mice-related publications. The reasons for this are twofold. First, the focus of MGI lies on mice strains developed and offered by JAX. Second, it predominantly includes publications about new mice models and newly identified genome-phenotype links of existing mice models. That is, the MGI data likely omits publications on more applied research.

We extend the collection of publications related to different strains by means of a text-based search in the Scopus database.[16] We identify publications making use of different strains, by searching for each strain name in the title, abstracts, or keywords of papers. Importantly, the procedure also allows telling apart strains most certainly provided by JAX from the same strains but provided by other different suppliers or bred locally. To do so, we leverage the nomenclature convention, which prescribes that a "J" is added at the end of the strain name if supplied by JAX. Conversely, strains from different suppliers have the same name but miss the "J" at the end (Standardized Genetic Nomenclature for Mice, 2016).

Some precautions are necessary to correctly identify publications related to specific strains. First, we extend each query with at least one generic keyword (*e.g.*, mouse, mice, rat, rats, strain) to minimize false positives. Second, we exclude in each query other strains with an identical substring of the focal strain.[17] Sublines of a particular strain often have the first part of the name in common. While these strains share a common lineage, their properties as well as their supply may differ, and therefore need to be distinguished. We find at least one pre-fire publication for 21 affected JAX strains, 12 spared JAX strains, and 28 strains available from other (non-commercial and commercial) suppliers.

Gene Data

Scientists rely on comprehensive information on the genetic profiles of mice when selecting the most suitable strain for a given research project. While the full sequencing of the mouse genome was only achieved in 2002 (Mouse Genome Sequencing Consortium and others, 2002), about 1,300 genetic loci had been indexed until 1989 (Lyon and Searle, 1989).[18] We digitize the information from Lyon and Searle (1989) and calculate the number of gene loci known for each strain. Figure A-4 in the Appendix illustrates the distribution of the number of loci across all strains and in our strain sample. We use this information to proxy how well scientists understood the genetic profile of a given strain at the time of the supply shock.

---

[16]To this end, we query via the RESTful API the on-line Scopus search-engine, which provides advanced search functionalities (*e.g.*, wildcards) and robustness to different spellings or double spacing between letters (Rose and Kitchin, 2019).

[17]Some strain names, particularly short ones, could be confounded with abbreviations in completely separate fields (e.g. engineering, chemistry, etc.). Example queries are detailed in Appendix A.1.

[18]An example of this matrix can be found in the Appendix A-3.

Table 1.1: Pre-fire characteristics of affected and spared mice strains (1979-1988)

MGI	Mean	Median	Std. Dev.	Min	Max
Affected strains (N = 35)					
First publication	1965.03	1967.00	11.92	1952	1985
Publications	39.37	21.00	50.92	1	217
JIF-weighted publications	254.11	184.35	299.29	0	1263
Citation-weighted publications	2623.83	1493.00	2982.21	39	12812
Spared strains (N = 32)					
First publication	1969.56	1972.50	11.29	1949	1985
Publications	24.00	6.50	55.44	1	315
JIF-weighted publications	142.49	50.79	317.43	3	1795
Citation-weighted publications	1940.34	1192.50	3887.20	29	22210
Scopus	Mean	Median	Std. Dev.	Min	Max
Affected strains (N = 39)					
First publication	1962.13	1961.00	12.00	1946	1987
Publications	628.21	76.00	1637.09	1	9497
JIF-weighted publications	1226.85	189.66	3002.63	0	16073
Citation-weighted publications	13882.79	2416.00	32981.74	6	187187
Spared strains (N = 22)					
First publication	1965.27	1965.00	11.61	1935	1980
Publications	516.23	40.00	1371.48	0	5873
JIF-weighted publications	1056.65	78.61	2959.50	0	13211
Citation-weighted publications	15825.68	1096.50	44403.35	0	196039

Notes: This table presents summary statistics of publications linked to affected and spared mice strains. The unit of observation is at the mice strain level. Publication counts are aggregated in the 1979-1988 period. Note that two spared strains have no Scopus publications in this ten-year period, but some in earlier years. 18 affected strains and 10 spared strains were available from other suppliers and are treated as separate observations in the Scopus data.

Summary Statistics

Table 1.1 reports summary statistics on bibliographic characteristics of affected and spared mice strains by data source. In the MGI database, affected strains are linked to an average of 39 publications in the ten years before the fire, whereas spared mice strains are linked to an average of 24 publications. This difference between affected and spared strains is robust to weighting publications by impact (JIF/citations). Moreover, affected strains are slightly older.

In the Scopus database, we find considerably more mice publications in the 10 years before the fire: about 630 for affected and 520 for spared strains. The reason for these higher counts is twofold. First, in the Scopus data, there are also non-JAX strains among affected and spared strains. These mice are genetically identical to the JAX mice but were provided by different suppliers or self-bred by the scientists. These non-JAX strains constitute a sizeable share of all mice-related publications. Second, in contrast to the MGI data, which focuses on studies whose contributions directly relate to the respective strain, the Scopus data also encompasses more applied studies, in which mice of a particular strain were employed as research material.

As the JAX-curated MGI data oversamples in-house mice, we can only draw on other – spared – JAX strains as a comparison group. For the Scopus data, we cannot only distinguish between affected and spared strains but also whether they were supplied by JAX or acquired from a different supplier (non-JAX). In a further analysis, we consider the group of affected non-JAX strains as additional controls.

1.3.2 Scientist Level

Identifying Affected and Control Scientists

We identify all scientists who published scientific articles using affected or spared mice strains prior to the JAX fire in 1989 based on the MGI and Scopus publication data. The initial data set includes 20,078 publications with 16,154 scientists until 1988. For each scientist, we add full bibliometric information from the Scopus database.[19]

We focus on all scientists with an affiliation in the U.S., Canada, or Western Europe and who had at least one mice-related publication between 1984 and 1988. This reduces our sample to 9,361 scientists.[20]

We determine the scientist level exposure to the supply shortage through the share of publications that use an affected mice strain, relative to publications that use an affected or spared mice strain, in the 5 years before the fire. To have sufficient variation when calculating the treatment exposure, we drop 5,251 scientists with only *one* publication related to either affected or spared mice strains.[21]

We define a scientist as being affected if her share of affected mice publications is larger than 50%. The affected scientist may have publications with multiple strains, partly affected and partly spared. To account for this in our exposure measure, we additionally require the affected scientist's share of publications that use a spared mice strain, to be below 50%:

$$
\textit{affected scientist}_i = \begin{cases} 0, \text{ if } \textit{share affected}_{i,84\text{-}88} \leq 0.5 \ \wedge \ \textit{share spared}_{i,84\text{-}88} \geq 0.5, \\ 1, \text{ if } \textit{share affected}_{i,84\text{-}88} \geq 0.5 \ \wedge \ \textit{share spared}_{i,84\text{-}88} \leq 0.5. \end{cases} \tag{1.1}
$$

This leaves us with 1,868 likely affected scientists and 465 likely control scientists.

Bibliometric information includes the age of a scientist (proxied by the time elapsed from the first publication), the full publication record including journal impact factor (JIF)

---

[19]All mice publications in MGI are connected to Scopus Publication IDs via the common Pubmed ID, if available. This allows us to extract disambiguated Scopus author identifiers for mice publications initially found in both MGI and Scopus. These author identifiers are then searched in Scopus again to extract the full bibliometric information of each scientist. To pull, cache and extract data we draw on the *pybliometrics* library (Rose and Kitchin, 2019).

[20]We drop scientists with presumably unclean Scopus author profiles, *e.g.*, with more than 50 years pre-fire publication record, and those who stopped publishing in 1988.

[21]This requirement on the exposure measure puts an high emphasis on more productive researchers. However, prior literature has highlighted their instrumental role for progress in science (*e.g.*, Iaria et al., 2018).

weights,[22] the global rank and country of a scientist's affiliation in 1988,[23] as well as the share of last authored publications, and citation counts. We add Medical Subject Headings (MeSH) from the National Library of Medicine[24] to each publication and identify mice publications through mice-related MeSH terms, such as mouse, mice, and murine.

Figure A-9 in the Appendix shows that the pre-treatment characteristics of all affected and all control scientists differ significantly with regard to scientific productivity. We, therefore, perform Coarsened Exact Matching (CEM) to identify more comparable dyads of scientists with common pre-fire characteristics, such as age, the number of JIF-weighted mice publications, and the number of pre-fire citations to pre-fire (mice) publications.[25] If there is more than one dyad combination possible within a CEM stratum, we choose the dyad with the smallest normalized distance in scientists' characteristics based on the above variables and additionally the affiliation ranking and the share of last authored publications. This leaves us with 283 affected-control scientist dyads (sample A). Additionally, we select a second sample of scientists in North America, the core market of JAX. This sample (sample USA) consists of 160 dyads, which should be more exposed to the shock given the geographic dimensions of the supply shortage. Overall, our sample emphasizes scientists who had a high propensity to stay in research.[26]

Summary Statistics

We compare pre-fire characteristics of affected and control scientists in the full sample (A) and the subsample (USA) summarized in Figure 1.3.[27] By construction, the affected and control group show no significant mean differences in those characteristics used in the matching procedure (age, JIF-weighted (mice) publications, citations, affiliation rank, and the share of last authored publications). Other independent characteristics, such as the set of (new) distinct coauthors, self-references, new keywords, new MeSH terms, publications in clinical journals,

---

[22]The journal impact factor relies on the Scimago journal and country rank data coming from https://www.scimagojr.com/journalrank.php [last accessed on March 8, 2021]. We use the annual Scimago Journal Rank indicator, which expresses the average number of weighted citations received in the selected year by the documents published in the journal in the three previous years. For most journals, the rank indicator information is available since the early 1990s. For earlier publications, we extrapolate the ranking.

[23]We use the global rank information of the Scimago Institutions Ranking, which can be found here: https://www.scimagoir.com/rankings.php [last accessed on March 8, 2021].

[24]MeSH is a controlled vocabulary thesaurus used for indexing articles in PubMed. More information on MeSH terms are available at: https://www.nlm.nih.gov/mesh/meshhome.html [last accessed on March 8, 2021].

[25]The selection of matching variables and cut-off points (deciles/3-year age bins) is similar to Azoulay et al. (2010).

[26]Since the CEM and selection of dyads is based on criteria chosen by the authors, we built further alternative samples with the following dimensions: a smaller sample B, which is stricter concerning age and scientific productivity, and the full sample C before the matching.

[27]The corresponding summary statistics can be found in Tables A-9 & A-10 in the Appendix.

patent-weighted number of publications are similar, too. We conduct various robustness checks with regard to the dependent variable construction.[28]

The scientists differ in those categories which should resemble the treatment exposure (bottom left part of Figure 1.3). The magnitudes of these differences indicate that most scientists are active in either of these two mice categories. Additionally, the affected and control scientists also differ with regards to strain types identified via MeSH terms from the full bibliometric information. Affected scientists have a substantially higher share of *affected mice-related publications* based on MeSH terms,[29] and control scientists have a much higher share of spared strain-related publications such as *B6* based on MeSH terms.[30] This provides confidence in our definition of treatment exposure.

Figure 1.3: Mean comparison of affected and control scientists
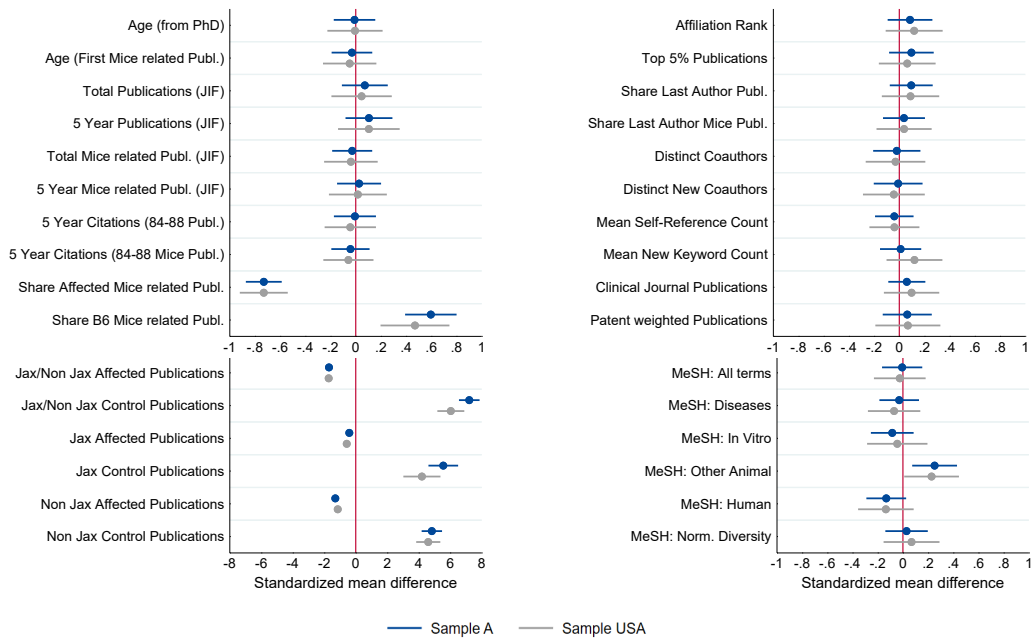
Notes: The figures presents comparisons of standardized mean differences of key pre-fire characteristics of the scientists in sample A and in sample USA after matching. The unit of observation is at the scientist level.

The pre-fire characteristics of the scientists are very similar not only in their means but also in their distributions. Figure A-8 in the Appendix compares the distributions of various pre-fire covariates. In the vast majority of cases, the distributions are nearly identical between the two groups.

---

[28]We calculate alternative dependent variables. For this, we either log-transform the count dependent variables or use simple counts instead of shares and averages. Moreover, we weight publications by the number of forward citations instead of journal-impact factors or by the inverse number of coauthors. Additionally, we restrict the dependent variable conservatively to only scientific journal articles with fewer than 20 authors.

[29]Affected mice identified by MeSH terms are: HRS, A, AKR, BALB/c, C3H, CBA, DBA, NOD, NZB, Obese.

[30]The MeSH terms include only four different types of unaffected strains (129, MRL-lpr, nude, and B6). Given the importance of B6 in the mice strain level analysis, we opted to focus on this unaffected strain specifically and not on all control strains.

1.4 Empirical Strategy

For the main part of our empirical analysis, we rely on Difference-in-Differences models, exploiting the plausibly exogenous supply shock caused by the 1989 Morrell Park fire at JAX. Unless mentioned otherwise, our baseline regressions have *annual JIF-weighted scientific publication counts* as the dependent variable. At the mice strain level, we investigate research output and the usage of particular strains in research in the ten years before and after the fire. The window of observation is restricted to the 10 years after the fire to minimize the overlap with subsequent changes in the market of laboratory mice.[31] For this purpose, we compare the number of publications linked to affected mice strains to the number of publications linked to spared mice strains before and after the fire. The model at mice strain level can be written as follows:

$$\mathbf{E}[y_{it}|X_{it}] = \exp[\alpha + \beta_1 \text{ post-fire} \times \text{affected strain}_{it} + \delta_t + \gamma_i], \tag{1.1}$$

where y represents the dependent variable, and the time-variant treatment variable *post-fire × affected strain* indicates whether the current observation belongs to affected strains in the post-fire period. Since laboratory mice became an increasingly popular research tool over time, we control for time trends using calendar year effects δ_t. Given the differences in importance and applicability (*e.g.*, number and strength of gene-phenotype links), we include strain fixed effects γ_i. We cluster standard errors at the mice strain level.

We explore the heterogeneity between different types of strain characteristics, such as the availability of commercial suppliers, pre-fire sales, or the number of known gene loci by using triple interactions. Furthermore, we restrict the sample to strains, for which there is no ambiguity concerning the exposure to the fire and disentangle industry trends towards the development of transgenic mice from our shock.

At the scientist level, we compare research activities conducted by potentially affected scientists to those of potentially unaffected scientists before and after the fire. First, we analyze the composition of mice strains in their research. Second, we investigate switching costs and adjustments in the type of research. Formally, we estimate the following Difference-in-Differences equation:

$$\mathbf{E}[y_{it}|X_{it}] = \exp[\alpha + \beta_1 \text{ post-fire} \times \text{affected scientist}_{it} + f(age_{it}) + \delta_t + \gamma_i]. \tag{1.2}$$

Again, the time-variant treatment variable *post-fire × affected scientist* indicates whether the current observation represents an affected scientist in the post-fire period. As the effect of the supply shock might be correlated with the scientists' career progress, we include age controls up to the 3rd degree polynomial to capture the potential non-linearity in research output

---

[31]Murray et al. (2016) analyze the substantial changes in the mice IP landscape related to Cre-lox and Onco mice due to NIH agreements in 1998 and 1999. This is around 10 years after the fire at JAX.

throughout a scientist's career. Moreover, δ_t corresponds to calendar year fixed effects and γ_i corresponds to scientist fixed effects. We cluster standard errors at the scientist level.

We explore potential effect differentials by scientist characteristics through a triple interaction. All scientist characteristics are binary indicators that split the sample around the median of the respective dimension.

At both levels of analysis, we estimate Poisson pseudo-maximum likelihood (PPML) regressions with high-dimensional fixed effects for count dependent variables.[32] We apply PPML regressions also in the context of annual averages, *e.g.*, mean annual number of new MeSH terms, to account for the long tail of the distribution. The coefficients represent semi-elasticities and allow a direct interpretation. For continuous dependent variables, which are in our setting log-transformed count dependent variables as well as shares, we run linear regressions models with high-dimensional fixed effects.[33] We conduct various robustness checks concerning the empirical specification.[34,35]

1.5 Results at Mice Strain Level

1.5.1 Descriptive Analysis

The top part of Figure 1.4 illustrates the trends in mice usage based on counts of publications linked to affected and spared JAX mice strains in the decade before and after the fire. The links between publications and mice strains are based on the MGI data. That is, we focus on JAX mice strains and their use in upstream research activities, such as gene sequencing or the development of new sub-strains. In Figure 1.4a, we compare the log-transformed annual number of publications linked to affected strains and spared strains as aggregate counts. In Figure 1.4b, we present JIF-weighted counts. Both groups follow a common path during the 10-years prior to the fire, *i.e.*, differences in means are statistically insignificant. However, shortly after the year of the fire, these trends diverge with more publications linked to the group of spared strains. Notably, the difference in mice usage seems to be long-lasting as we do not observe a convergence in trends despite the temporary nature of the fire-induced supply shortage.

One spared mice strain – B6 – is linked to a particularly large number of publications. Singling out this mice strain, we observe that the relative increase in the use of spared mice is largely, but not entirely, driven by the B6 strain. This strain has many favorable characteristics

---

[32]We use the *ppmlhdfe* Stata package as described in Correia et al. (2020).

[33]We use the *reghdfe* Stata package based on Correia (2016).

[34]First, we use fewer and no age controls. Second, we cluster the standard errors at the scientist-year level or the strain level (most often used strain by scientist). Third, we use a different operationalization of the exposure variable. We also use continuous instead of binary definitions of the treatment variable *affected scientist$_i$*.

[35]In our event study specifications, we follow Jaravel et al. (2018) and interact the affected strain/scientist variable with a full set of leads and lags from -10 to $+10$ years around the JAX fire $\sum_{t=\underline{j}}^{\bar{j}} \beta_t (year_t \times affected_i)$ with two-year bins. We normalize the coefficient $\beta_{t=-2}$ to zero and, hence, express the dynamic treatment effects relative to this pre-treatment year.

Figure 1.4: Mice usage – publication counts

Aggregated counts (MGI data)

(a) Publications

(b) Publications (JIF-weighted)

Average counts (Scopus data)

(c) Publications

(d) Publications (JIF-weighted)

Notes: The top left-hand figure presents the log-transformed annual sums of publications linked to affected and spared mice strains. The top right-hand figure presents the log-transformed annual sums of publications linked to affected and spared mice strains weighted by their respective journal impact factor. Publications linked to affected and spared mice are included twice. The bottom left-hand figure presents the log-transformed average publication counts at mice strain level. The bottom right-hand figure presents the log-transformed average publication counts weighted by their respective journal impact factor. Publications can be linked to multiple mice strains. In both bottom graphs, the unit of observation is the unique mice strain.

and is widely applicable.[36] In our analysis at the scientist level, we examine whether scientists have used B6 as the main substitute for affected strains.

In the bottom part of Figure 1.4, we depict average trends in mice usage at the strain level. To this end, we use the more comprehensive Scopus data, which also covers more applied research, and links ten times more publications to mice strains than the MGI data. Moreover, we can distinguish between publications linked to JAX and non-JAX mice, where the latter group refers to strains that either were self-bred or acquired from other (commercial) suppliers. Both simple and JIF-weighted publication counts show a notable decrease in the use

---

[36]As a matter of fact, JAX pointed in the aftermath of the fire towards this strain as a potential alternative for research activities (see Figure A-2 in the Appendix).

of affected strains provided by JAX whereas mice usage in the two comparison groups, spared mice strains from JAX and affected strains from non-JAX sources, continues its pre-fire trend.[37]

1.5.2 Multivariate Analysis

The results of our empirical analysis at the mice strain level indicate that the use of affected JAX mice decreased relative to that of spared JAX mice and affected non-JAX mice. This decline was long-lasting, unrelated to industry trends, and affected mice usage in upstream as well as downstream research. The long-lasting effect cannot be explained by fewer newcomer scientists starting research with the affected mice strains. Instead, incumbent scientists appear to have increasingly turned their back on strains that were temporarily unavailable.

In line with the descriptive results, we find a highly negative effect of the supply shortage on subsequent mice usage (Table 1.2). Compared to spared strains, affected strains are linked to about 70% fewer publications in the 10 years after the fire, presented in Column 1. In Column 3, we measure mice usage based on our second publication data source (Scopus) and find a statistically significant decrease of about 20%. Although less precisely estimated, we find a negative effect of similar magnitude when adding non-JAX mice to the unaffected group in Column 6. The results for simple publication counts are corroborated with those of JIF-weighted publication counts.

Table 1.2: Impact of mice supply shortage on mice usage (strain level)

Strains Sample:	(1)	(2)	(3)	(4)	(5)	(6)
	MGI JAX strains		Scopus JAX strains		Scopus all strains	
+ 10 years	Publ	Publ (JIF)	Publ	Publ (JIF)	Publ	Publ (JIF)
Post × affected	−0.721**	−0.980***	−0.233**	−0.454***	−0.306	−0.502*
	(0.366)	(0.377)	(0.113)	(0.125)	(0.207)	(0.283)
Strain FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1344	1344	693	693	1239	1239
Strains	64	64	33	33	59	59
Log-likelihood	−2403	−12508	−1375	−3583	−3238	−9570

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The outcome variable is the number of (JIF-weighted) publications. The unit of observation is the individual mice strain by year. Standard errors are clustered at the mice strain level and shown in parentheses. Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

One important indicator of the validity of Difference-in-Differences estimations are parallel trends prior to the exogenous supply shock following the fire. Figure 1.5a depicts event study estimates with publication counts as the dependent variables. We observe a significant reduction in the use of affected mice strains. Moreover, the lack of confounding pre-trends

---

[37]This result is robust to using a subset of treated mice strains that follows a more conservative classification (see Figure A-6 in the Appendix).

Figure 1.5: Impact of mice supply shortage on mice usage (strain level) – heterogeneity
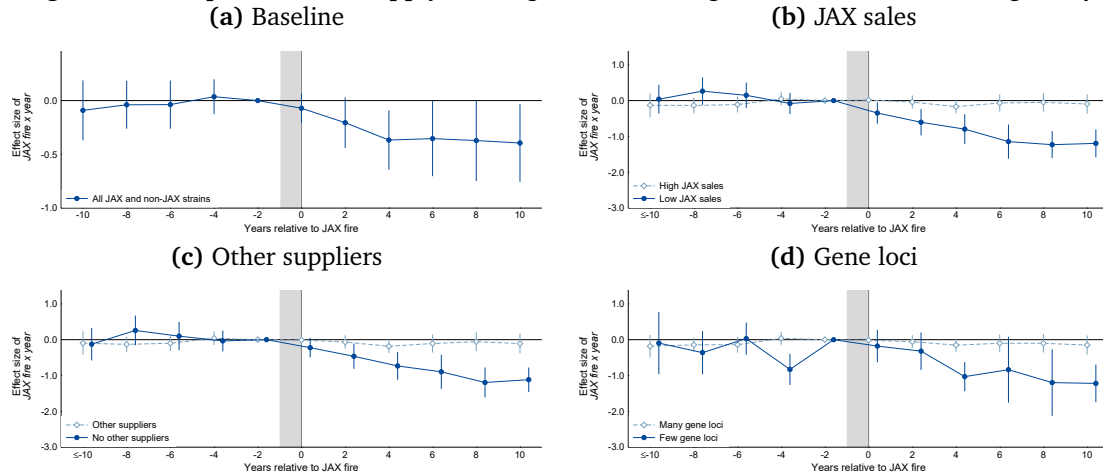
(a) Baseline **(b)** JAX sales

(c) Other suppliers **(d)** Gene loci

Notes: The figures show the event study estimates and the 90 percent confidence bands of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects following Equation 1.1 in (a) and with a triple interaction term in (b) to (d). The unit of observation is the individual strain by year. Standard errors are clustered at the strain level.

suggests that the parallel trends assumption is likely fulfilled in our setting.[38] The reduction in usage seems to be long-lasting, if not permanent.

In Figures 1.5b to Figures 1.5d, we further explore potential heterogeneity in the effect on research output by strain characteristics that proxy the likely substitutability of the affected strain through similar strains and/or different sources.[39] We first distinguish between JAX strains that belonged to the top-selling mice before the fire and niche mice with relatively low sales figures. We find that mice strains with low sales figures drive the average decline in mice usage. We further distinguish between mice strains that were available from *commercial mice suppliers* and those exclusively offered by JAX. Notably, the decrease in mice usage is driven by those affected strains which could not be purchased from other sources. Finally, we distinguish between strains with high and low numbers of gene loci publicly known at the time of the JAX fire. Mice with a large number of known gene loci are well understood in the scientific community and it should be easier to judge whether another strain has a similar genetic profile. The substitution of strains with *many known genes* should, hence, be easier compared to strains with few known gene associations. In fact, the decrease in mice usage is confined to strains with a below-average number of gene loci. These three differential effects suggest that the temporary supply shortage had long-term consequences primarily for those strains that were hard to adequately substitute. This provides one possible explanation for the long-term effect on mice usage: scientists who had to adopt imperfect substitutes continued using them given the necessary adjustments of their research line.

---

[38]Appendix Figure A-7 illustrates the event study results for weighted publication counts and for the MGI data.

[39]Appendix Table A-7 provides the corroborating results of triple Difference-in-Differences models.

We can rule out that industry trends led to the observed change in mice usage. Murray et al. (2016) elaborate that there was an ongoing development of new transgenic mice technologies like Onco, Cre-lox, and Knock-out in the 1990s.[40] To exclude any concerns that we depict trends that would have happened in the absence of the JAX fire, we exclude all publications from our sample, which are classified by the MeSH term "transgenic". The results in Table A-6 in the Appendix confirm that the decrease in publications related to affected strains is unrelated to the development of transgenic mice. Consistently, the coefficients have a smaller magnitude.

Moreover, we can exclude the alternative explanation that the long-term decline in mice usage is purely a cohort effect. In the aftermath of the fire, young researchers entering the biomedical field possibly started new research projects based on available (*i.e.*, spared) mice strains and kept on working with these permanently. If this was the case, the reduced research output would be due to the lack of newcomers conducting research with affected mice strains, whereas incumbent researchers continued working with their acquainted strains irrespective of the temporary supply shortage. We split publications by whether the author team age lies above the overall average team age in the respective publication year. As the results in Table A-8 illustrate, the decrease in total publications related to a specific strain is not solely due to a decrease in publications by young research teams. If anything, the decrease in publications by old research teams appears to be larger in magnitude.

1.6 Results at Scientist Level

1.6.1 Descriptive Analysis

Figure 1.6 illustrates the trends in the average research output of affected and control scientists over time for the full sample as well as for the North American sample. Pre-trends correlated with the treatment exposure may lead to concerns that estimates pick up pre-existing activity. However, the plotted relationships support the parallel trends assumption at the scientist level. The annual share in the usage of specific strains[41] as well as the JIF-weighted publication counts move largely parallel in the 5-year period before the fire. This holds also true when looking at longer pre-periods or unweighted simple publication counts. More formally, we estimate a placebo specification of Equation 1.2 using pre-period data.[42] The estimated coefficients are small in magnitude and statistically insignificant. This supports a causal interpretation of our estimates in the following multivariate analysis.

---

[40]Traditionally, the basis for these experiments has been the 129 strain. However, B6 became an increasingly popular model (Seong et al., 2004). Figure A-5 in the Appendix illustrates the different dynamics between those two strains. The sudden increase in the usage of B6 happens directly after the fire and substantially earlier then the transgenic mice induced increase of 129 strains.

[41]We illustrate the spared strain B6 in Figures 1.6a and 1.6b and the affected mice strains in Appendix Figure A-11.

[42]This regression includes an interaction of affected scientists with an indicator for the 1986-8 period. Appendix Table A-11 contains the results from this specification.

Figure 1.6: Scientist publications linked to affected and spared mice strains

(a) Share B6 publ. (JIF) - sample A

(b) Share B6 publ. (JIF) - sample USA

(c) Publications (JIF) - sample A

(d) Publications (JIF) - sample USA
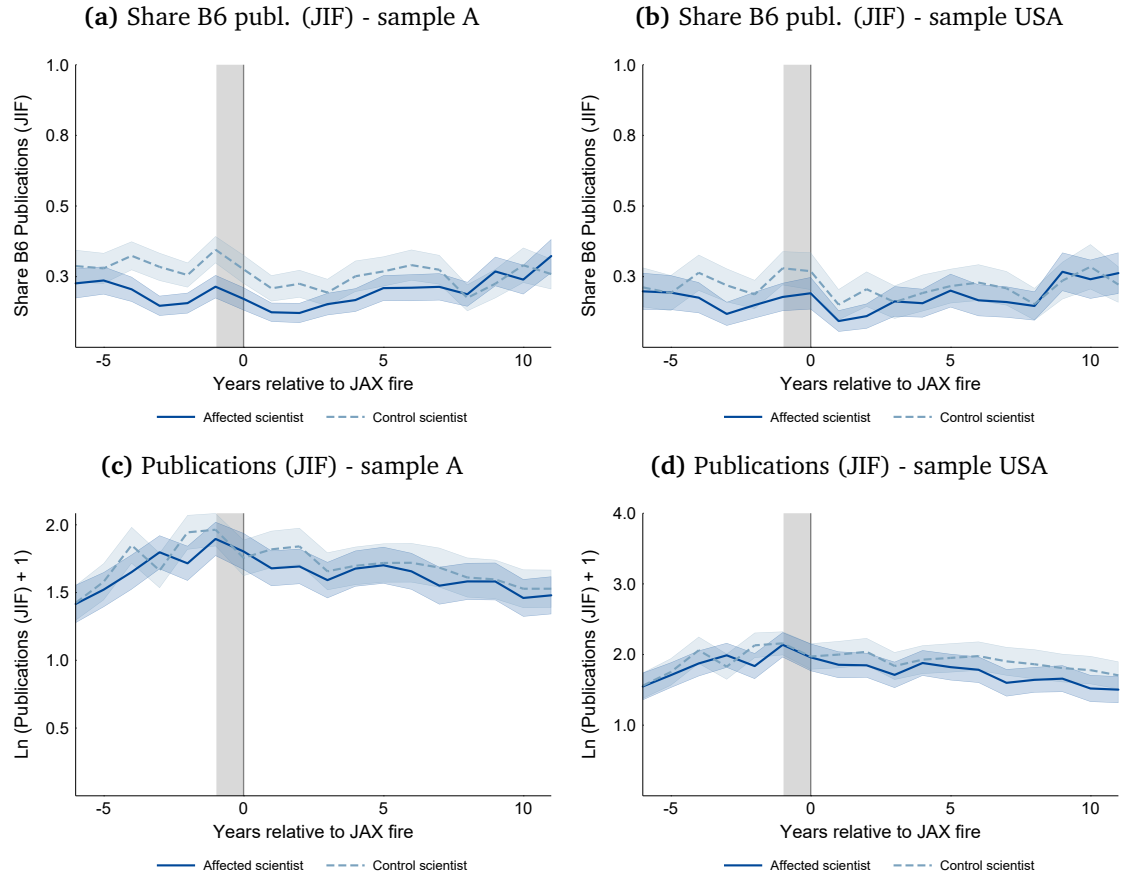
Notes: The top Figures (a) and (b) illustrate the mean JIF-weighted share of B6 publications over time for affected and control scientists. The bottom Figures (c) and (d) show the mean JIF-weighted publication counts. The unit of observation is the scientist level.

Figures 1.6 and A-11 also provide a first impression concerning the effect of the supply shock on the adoption of research material and switching costs. First, affected scientists increase (decrease) the share of JIF-weighted publications related to *spared strains (affected mice strains)* compared to control scientists. This suggests that affected scientists are more likely to adopt different strains. Second, the trends in overall publication counts seem to diverge in the years following the fire-induced supply shock, especially within the North American sample. Scientists that were unable to purchase the laboratory mice necessary for their research published a lower number of articles than comparable scientists that had not previously relied on affected JAX strains.[43] Taken together, these descriptive analyses suggest the occurrence of switching costs caused by the adoption of new tools.

---

[43]At the extensive margin in Appendix Figure A-10, both affected researchers and control researchers exit science (defined by the last observable publication of this person) with a similarly low likelihood after the fire. 75% of the scientists have their last publication 15 or more years after the fire.

1.6.2 Multivariate Analysis

Adoption of Mice

In this section, we show that different levels of use of mice strain after the fire result from scientists who previously worked with affected strains and then adopted spared mice strains. One of the most intriguing aspects is that the scientists did not switch back to their former strains after JAX was rebuilt. This suggests the existence of switching costs that make returning to the prior equilibrium less attractive than continuing in the new one.

Figure 1.7 plots the event study estimates of the Poisson pseudo-maximum likelihood model with scientist and year fixed effects (similar to Equation 1.2). The outcome variable is the JIF-weighted share of B6 (affected) mice publications. In line with our descriptive analysis, the composition of research mice follows similar patterns for both groups of scientists prior to the fire. After the fire, however, the use of affected mice strains decreases while the use of unaffected strains increases. This divergence grows throughout the full post-fire period.

Figure 1.7: Adoption of mice – event study
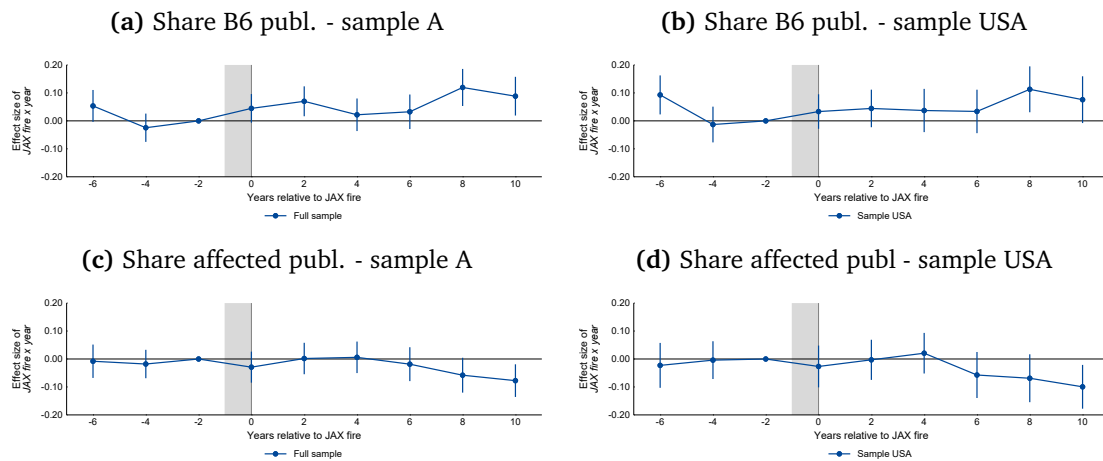
(a) Share B6 publ. - sample A

(b) Share B6 publ. - sample USA

(c) Share affected publ. - sample A

(d) Share affected publ - sample USA

Notes: The figures show the event study estimates and the 90 percent confidence bands of linear regressions with high-dimensional fixed effects following Equation 1.2. The outcome variable is the share of JIF-weighted B6/affected mice publications relative to all mice publications. In both samples, the unit of observation is the individual scientist by year. Standard errors are clustered at the scientist level.

Table 1.3 presents the estimates of Difference-in-Differences linear regression on the share of (JIF-weighted) publications based on either spared B6 mice or affected mice strains. We find statistically significant effects on the composition of mice strains used in the scientists' research.[44] Affected scientists are more likely to adopt B6 mice for their research, and less likely to use affected mice strains in the post-fire period. This holds true for the full sample of

---

[44]In the Appendix, we illustrate that these results are quantitatively similar when using alternative specifications (Table A-13), alternative dependent variables such as forward-citation weighted publications (Table A-16), scientific articles with fewer than 20 coauthors (Table A-23), publications weighted by the inverse number of coauthors (Table A-24), or alternative samples such as sample B (Table A-25) or the pre-matching sample C (Table A-26).

283 scientist dyads and the North American sample with 160 dyads. However, estimates are more precise in the full sample. In terms of magnitude, affected scientists publish about 10 percentage points more JIF-weighted B6 publications in year 10 after the fire. Given a "base" pre-fire share of about 20%, this resembles an increase of around 50% or *one* JIF-weighted B6 publication every two years. The effect is not driven by a change in the number of mice-related publications,[45] but rather by a higher share of these publications that rely on B6. In contrast, affected scientists publish about 10 percentage points fewer publications with affected mice strains in year 10 after the fire (see event studies in Figure 1.7).[46]

The change in the composition of research tools emerges and is already significant within a shorter period of ±5 years around the fire, albeit the estimates are smaller in magnitude and less precise (Appendix Table A-12). We argue that the smaller effect size in the ±5 is consistent with the idea the adoption of new research tools consolidates over time and becomes evident in publications with some delay.

Table 1.3: Adoption of mice – DiD

Share/Linear +10 years	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Sample A				Sample USA			
	B6	Affected	B6 (JIF)	Affected (JIF)	B6	Affected	B6 (JIF)	Affected (JIF)
Post × affected	0.046***	−0.037**	0.040***	−0.026*	0.025	−0.032*	0.020	−0.039**
	(0.015)	(0.015)	(0.015)	(0.015)	(0.018)	(0.019)	(0.018)	(0.019)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9622	9622	9622	9622	5440	5440	5440	5440
Scientists	566	566	566	566	320	320	320	320
Log-likelihood	−344	−478	−676	−679	159	−335	11	−498

Notes: Columns (1) to (8) show the estimates of linear regressions with high-dimensional fixed effects. The outcome variable is the share of (JIF-weighted) B6/affected mice publications relative to all mice publications. Columns (1) to (4) show the estimates for sample A and Columns (5) to (8) show the estimated for sample USA. The unit of observation is the individual scientist by year. Standard errors are clustered at the scientist level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Research Output

The results of our empirical analysis suggest that scientists who adopt new strains following the temporary supply shock faced switching costs measured by decreases in research productivity. These switching costs increase if adopting a different strain is difficult such as in North America.

---

[45]The Appendix includes Table A-12 for count and for log-transformed dependent variables.

[46]Table A-22 in the Appendix shows that affected scientists publish fewer publications with affected (JAX or Non-JAX) strains and more with spared (JAX or Non-JAX) strains. Moreover, affected scientists publish fewer publications with inbred mice, but not more publications with transgenic mice. This supports the notion that our results are not driven by industry trends towards new mice technologies (transgenic mice like Knockout mice) but within the existing range of technology.

Figure 1.8 depicts the event study results of the PPML model (Equation 1.2). The outcome variable is the JIF-weighted number of (mice) publications. In line with our descriptive analysis, the number of scientific publications follows similar patterns for both groups of scientists prior to the JAX fire. After the fire, there is a drop in the number of publications, especially for North American scientists. This decrease is smaller when looking at mice publications only. After the initial drop in 1991/1992, affected scientists return slowly to pre-fire levels of research output.[47]

Figure 1.8: Research output – event study
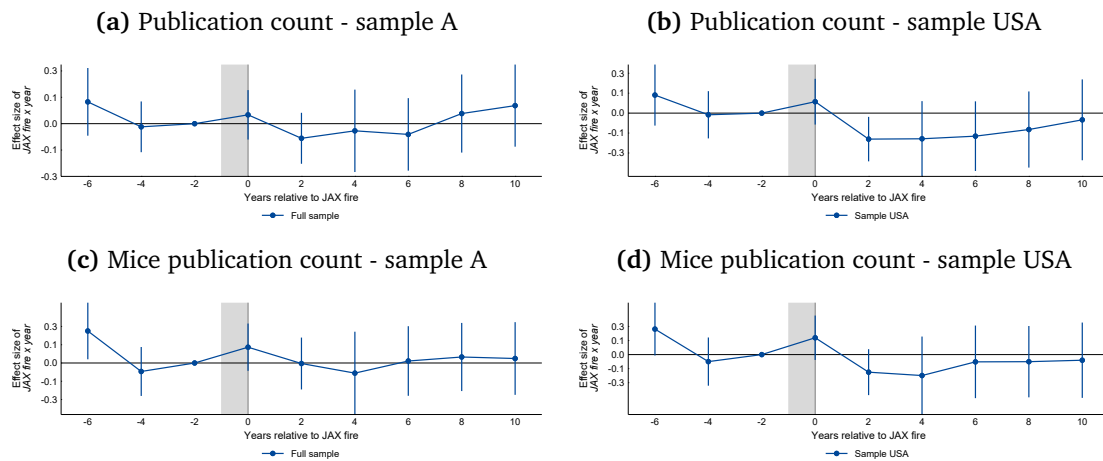
(a) Publication count - sample A

(b) Publication count - sample USA

(c) Mice publication count - sample A

(d) Mice publication count - sample USA

Notes: The figures show the event study estimates and the 90 percent confidence bands of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects following Equation 1.2. The outcome variable is the number of JIF-weighted (mice) publications. In both samples, the unit of observation is the individual scientist by year. Standard errors are clustered at the scientist level.

Table 1.4 presents the estimates of Difference-in-Differences estimations with (mice) publication counts, in either simple counts or JIF-weighted. We find small negative effects of the supply shock on all (JIF-weighted) publication activities. These effects are more pronounced and significant when looking at affected scientists from North America only – who presumably were most exposed to the shock. This is in line with our assumption that North American scientists faced higher switching costs adopting spared strains. For instance, other U.S. commercial suppliers increased their prices substantially in the aftermath of the fire, limiting the range of affordable strains as potential substitutes.

The point estimate ranges between -0.187 for the North American sample and -0.055 for the overall sample.[48] This refers to a decrease in the magnitude of 18.7% (5.5%) relative to the pre-fire mean of 4.33 (4.05) publications per year. Relative to the "base" mean publication counts, this translates into 8.1 (2.2) fewer publications within 10 years.[49] The dynamics of

---

[47]We illustrate the event study results of many other outcome variables in Figure A-12 in the Appendix.

[48]This results are broadly similar when looking at linear specifications in Table A-14 and alternative specifications in Table A-15 in the Appendix.

[49]In robustness checks we account for outliers by winsorizing all count variables at the 95th percentile. All results are robust using winsorized values.

the effect (±5 years shown in Appendix Table A-14) suggest that switching costs occur only temporarily at the moment of initial tool adoption. The results for all publications can be confirmed broadly when looking at *mice* publications directly. We estimate – except for the simple mice publication counts in the full sample – negative albeit insignificant effects.

To exclude the alternative explanation that changes in productivity might be driven by the need to work with new people, we investigate whether the adoption of new tools is associated with changes in the organization of research. This follows the discussion by Teodoridis (2017) and Zyontz (2019), who show that scientists respond to the availability of new tools with a re-organization of their scientific knowledge production. However, we show in Table A-17 in the Appendix that, for instance, the average number of coauthors per paper as well as the position of the scientists within a project, *e.g.*, last author as a proxy for senior scientist/principal investigator, remains unaffected.

Table 1.4: Research output – DiD

Count/PPML +10 years	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Sample A				Sample USA			
	Publ	Mice	Publ (JIF)	Mice (JIF)	Publ	Mice	Publ (JIF)	Mice (JIF)
Post × affected	−0.055	0.030	−0.039	−0.027	−0.187**	−0.025	−0.151	−0.122
	(0.069)	(0.079)	(0.089)	(0.111)	(0.079)	(0.103)	(0.103)	(0.139)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9622	9622	9622	9622	5440	5440	5440	5440
Scientists	566	566	566	566	320	320	320	320
Log-likelihood	−20655	−14323	−51237	−40923	−11783	−8291	−32981	−26303

Notes: Columns (1) to (8) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The outcome variable is the number of (JIF-weighted) publications/mice publications. Columns (1) to (4) show the estimates for sample A and Columns (5) to (8) show the estimated for sample USA. The unit of observation is the individual scientist by year. Standard errors are clustered at the scientist level and shown in parentheses. Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Type of Research

A potential reason for our finding is the imperfect substitutability of different research mice, requiring investments in complementary knowledge. If this was the case, one would expect long-term adjustments in research trajectories related to the adopted strains, adjustment costs, as well as a lower propensity to re-use old tools once they are available again.

Imperfect substitutability between old and newly adopted mice strains may require adjustments to the scientist's research pipeline. This is because different mice strains have different genetic compositions leading to different traits, which determines their suitability to study certain phenomena (see Table A-5 for examples) and their substitutability.[50] Substitute mice

---

[50]As a matter of fact, the case studies cited in Section 1.2.3 describe, for example, that the unavailability of certain strain may induce the researchers to switch topics as response to work with different mice.

strains may overlap with respect to some relevant genetic features but differ with regard to others. We show in this section that the adoption of new mice strains, as a response to the shock, has led to long-term changes in the type of conducted research, in particular research related to the newly adopted strains. Given that these adjustments became necessary when adopting imperfect substitutes to the affected strains, scientists faced considerable costs of switching back to the old strains once they had become available again.

Table 1.5 presents the effects of the supply shortage of laboratory mice on the scientists' likelihood to work on new research topics. We create three distinct measures to capture these changes in research direction. First, we use the mean annual number of self-references in post-fire publications as an indicator to what extent the scientist relied on her previous body of work (*e.g.*, Baruffaldi and Gaessler, 2021). Second, we use the mean annual number of keywords used for the first time by the scientist. Third, we leverage the detailed information on the content of biomedical publications provided by the Medical Subject Headings. We calculate the mean annual number of MeSH terms that are new to the scientists. This may include also information on research methods. Additionally, we restrict the new MeSH terms to disease-related vocabularies to proxy the novelty of the topic of research more specifically.

We find partial evidence that scientists, who adopt new mice strains, changed the direction of their research towards unrelated projects. Comparing affected scientists to their controls, our estimates show a negative effect on self-references to the respective scientist's own publications in the magnitude of -0.223 for the full sample and -0.201 for the North American sample. This decline is disproportionately larger than that for all references (see Appendix Table A-18 for comparison). The differential effect suggests that affected scientists are more likely to conduct exploratory research. However, the evidence remains mixed, as we do not see a significant parallel shift in research topics, as captured by keywords. While all point estimates are positive – in line with the results on self-references – none of the coefficients is significant. Some of them are large in magnitude, *e.g.*, the annual number of *new* disease-related MeSH terms, and consistently higher than for *old* disease-related MeSH terms used by the scientist.[51]

Table 1.6 shows that the scientists' research became more applied, *i.e.*, relevant for industry. In Panel A, we investigate the share of publications published in journals relevant for clinical research.[52] We find positive significant effects of a magnitude of 3 percentage points. Affected

---

[51]We explore in Table A-19 whether affected scientists use different research methods such as "in vitro" research, *e.g.*, related to cell cultures, other typical "research animals", e.g. rats, monkeys or hamster, or "humans". In all three cases, we do not find any significant changes. The point estimates are close to zero suggesting that there is no differential effect of the supply shortage on other research tools than mice.

[52]We use a classification of journals based on the proportion of published research coming from general hospitals and industry using the publicly available data set provided by Tijssen (2010). This list includes about 5,000 major scientific journals indexed in Web of Science. We link this list to scientific journals appearing in Scopus. A journal is classified as relevant for clinical research, if more than 3% of all publications can be related to a general hospital, but less than 3% to industry. This leaves us with a set of 1,182 journals or around 25% of all journals. The data can be accessed here: https://www.vosviewer.com/journal-application-domain-map [last accessed on March 8, 2021].

Table 1.5: Type of research (novelty) – DiD

Mean/PPML +10 years	(1) Self-References	(2)	(3) New Keywords	(4)	(5) New MeSH	(6)	(7) New Disease MeSH	(8)
	All	USA	All	USA	All	USA	All	USA
Post × affected	−0.223**	−0.201*	0.038	0.213	0.058	0.044	0.131	0.096
	(0.093)	(0.115)	(0.156)	(0.217)	(0.045)	(0.057)	(0.088)	(0.121)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9129	5270	8738	4981	9622	5440	9452	5338
Scientists	537	310	514	293	566	320	556	314
Log-likelihood	−10287	−6134	−8331	−4486	−22339	−12542	−5454	−3124

Notes: Columns (1) to (8) show the estimates of linear regressions with high-dimensional fixed effects. The outcome variable is the annual average number of self-references in Columns (1) and (2), new keywords in Columns (3) and (4), new MeSH terms in Columns (5) and (6), and new disease-related MeSH terms in Columns (7) and (8). Columns with odd numbers show the estimates for sample A and columns with even numbers show the estimated for sample USA. The unit of observation is the individual scientist by year. Standard errors are clustered at the scientist level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

scientists are more likely to publish their research in clinical journals relative to their control group, illustrated in Columns 1 & 5. This is driven by publications using unaffected strains: the average B6 publication becomes more likely to be published in a clinical journal in Columns 3 & 7. The remaining affected mice strain publications, however, do not change their relevance for clinical research in Columns 4 & 8.[53] It is likely that affected scientists turn to this type of research since spared mice were already associated with it before the fire.

We observe similar patterns when exploring the commercial applicability of the research output. Thus, we link patents via references from the front page to scientific publications using the open-access data set from Marx and Fuegi (2020).[54] In Panel B of Table 1.6, we weighted scientific publications by the availability of at least one reference to a patent within 10 years and calculate the share relative to all publications. The share of patent-weighted publications, thus commercially applicable, increases for all publications, presented in Columns 1 & 5. Again, we find that the effect is driven by publications based on spared strains like B6 in Columns 3 & 7 and not by publications based on affected mice strains in Columns 4 & 8.[55] We investigate the content of citing patents by extracting the most frequently used expressions in the patent title (Table A-14). Less than 2% of patents relate directly to the development of

[53]These results hold at the extensive margin. The (mice/B6) publication count in clinical journals is increasing in PPML and linear regressions (Table A-20).

[54]Patent citations to non-patent literature are a commonly used approach to investigate the relevance of scientific research for commercially relevant applications (*e.g.*, Poege et al., 2019; Watzinger and Schnitzer, 2019; Ahmadpoor and Jones, 2017). We link the scientific publications to patents via the Pubmed ID. Each patent-paper linkage is assigned a confidence score. We keep NPL examiner citations with a confidence score greater than 3, that where found on a patent's front page, and have an priority year available. Patents are aggregated to DOCDB Patent Families. To minimize truncation of the data, we include only patents with a priority year 10 (or 5) years after the publications year of the scientific research.

[55]The same holds true at the extensive margin. For instance, we use patent-weighted publication counts and observe a small significantly positive effect for B6 publications (Appendix A-20).

Table 1.6: Type of research (appliedness) – DiD

Panel A	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share/Linear	Sample All: Clinical Y / Y Publ				Sample USA: Clinical Y / Y Publ			
+10 years	Y = All	Y = Mice	Y = B6	Y = Affectd	Y = All	Y = Mice	Y = B6	Y = Affectd
Post × affected	0.030*	0.027*	0.036***	−0.004	0.032	0.030	0.031**	0.000
	(0.016)	(0.016)	(0.011)	(0.010)	(0.020)	(0.020)	(0.013)	(0.013)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9622	9622	9622	9622	5440	5440	5440	5440
Scientists	566	566	566	566	320	320	320	320
Log-likelihood	−878	−1046	1827	2238	−336	−452	1396	1195

Panel B	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share/Linear	Sample All: Patent-Weighted Y / Y Publ				Sample USA: Patent-Weighted Y / Y Publ			
+10 years	Y = All	Y = Mice	Y = B6	Y = Affectd	Y = All	Y = Mice	Y = B6	Y = Affectd
Post × affected	0.024***	0.029***	0.019***	0.004	0.039***	0.040***	0.016**	0.003
	(0.007)	(0.008)	(0.005)	(0.006)	(0.011)	(0.012)	(0.006)	(0.009)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9622	9622	9622	9622	5440	5440	5440	5440
Scientists	566	566	566	566	320	320	320	320
Log-likelihood	5132	3686	7767	5512	2542	1850	4357	2446

Notes: Columns (1) to (8) show the estimates of linear regressions with high-dimensional fixed effects. In the top part of the table, the outcome variable is the share of all/mice/B6/affected mice publications published in a journal that usually publishes more clinical relevant research relative to all/mice/B6/affected mice publications. In the bottom part of the table, the outcome variable is the share of all/mice/B6/affected mice publications that are associated with a patent application (patent-weighted) relative to all/mice/B6/affected mice unweighted publications. A patent-weight is calculated based on the patent family's first application being filed within 10 years from the scientific publication. Columns with odd numbers show the estimates for sample A and columns with even numbers show the estimated for sample USA. The unit of observation is the individual scientist by year. Standard errors are clustered at the scientist level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

new mice strains (such as transgenic mice). Instead, the patents cover methods, treatments, the functioning of cells, and the composition of genes (all of these are more important among B6 publications, except methods). These results are in line with our findings on clinically relevant journal publications. Affected scientists use the newly adopted mice strains to publish research that is more applied.[56]

---

[56]In Table A-20 in the Appendix, we present the results of trajectory indicator weighted publication counts. These estimates are, however, confounded by the overall decrease in the number of publications for affected scientists. Nevertheless, they show an increase in B6 related applied publications. Moreover, our research trajectory results are robust to alternative specifications in Table A-21.

Heterogeneity

We solidify switching costs as the primary mechanism at work by exploiting heterogeneity among scientists. We focus on the scientists' pre-fire intensity of mice usage and seniority as possible moderators of the average treatment effects on mice adoption, research output and research direction.[57] We augment the main specification (Equation 1.2) with a triple interaction to elicit heterogeneous effects.

Figure 1.9 presents differential effects for affected scientists depending on their pre-fire mice usage. Scientists with a high number of mice-related publications relied more heavily and frequently on mice as research tools, managing larger supply and complementary facilities. Accordingly, switching costs for these scientists are likely larger in absolute terms, proportionally to the number of mice they used. Indeed, we find heterogeneous effects on the adoption of mice strains, productivity, and type of research, which mirror the observed heterogeneity between sample A and sample USA. Affected scientists with an above-average pre-fire number of mice publications experience disproportionate decreases in their research output. This is in line with the interpretation that a higher pre-fire reliance on mice makes switching more costly. The imperfect substitutability of strains creates the necessity to adjust research projects. We find evidence for positive effects on the shares of clinical journal publications and patent-cited publications.[58]

Figure 1.9: Heterogeneity by mice dependence
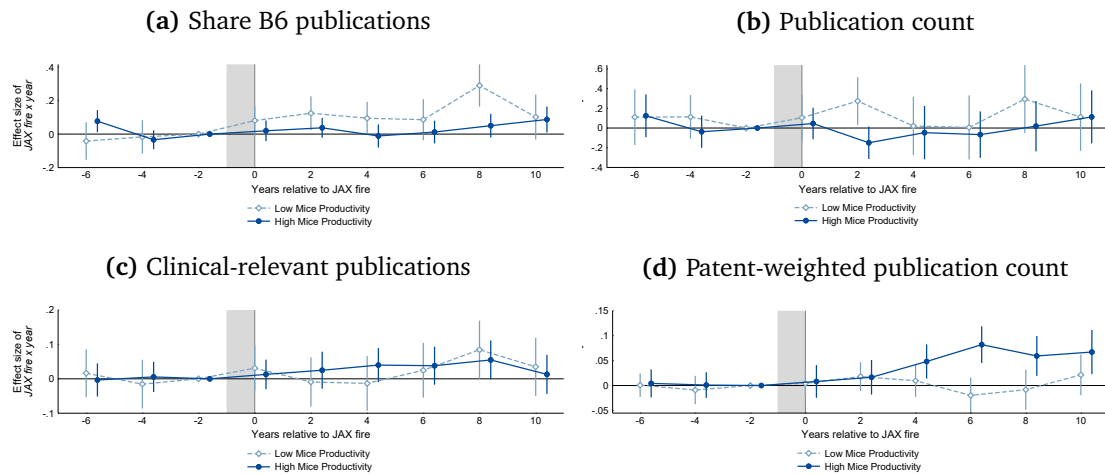
(a) Share B6 publications

(b) Publication count

(c) Clinical-relevant publications

(d) Patent-weighted publication count

Notes: The figures show the event study estimates and the 90 percent confidence bands of regressions following Equation 1.2 with a triple-interaction term using sample A. The triple-interaction term is a scientists' JIF-weighted mice publication record in the 5 years before the fire. Count data is estimated using Poisson pseudo maximum likelihood regressions and shares are estimated using linear regressions. The unit of observation is the individual scientist by year. Standard errors are clustered at the scientist level.

---

[57]Unfortunately, we cannot exploit variation in mice characteristics, because the most common strains in our scientists' portfolios are very similar regarding the number of gene loci, sales, or commercial availability.

[58]Tables in the Appendix Section A.5.9 show the results of regression estimations with interactions that confirm the presented event studies.

When examining potential heterogeneity by scientist age (Figure A-13), we expect more senior scientists to be more path-dependent so that the adoption of new mice strains causes more switching costs and higher adjustments to projects. In fact, older scientists are less likely to adopt new strains such as B6, have fewer publications, and are more likely to change their type of research towards more applied publications. However, these differences are not significantly different from the baseline.

1.7 Discussion and Conclusion

Institutions and markets increasingly play a role in the production and diffusion of research inputs. In this paper, we posit that the existence of switching costs for scientists influences the adoption of research tools and bears important implications for the functioning of such markets. To provide empirical evidence, we exploit the unforeseen and temporary supply shortage of specific mice strains caused by the 1989 Morrell Park fire at JAX as a natural experiment. We find that the supply-shock, albeit temporary, had long-lasting effects on the adoption of different mice strains. This effect is explained by those mice for which switching costs were presumably higher, due to lower levels of supply prior to the fire, the lack of alternative suppliers, and higher uncertainty associated with a more limited understanding of their genetic code. Moreover, the shock favored an increase in the use of mice, among those not affected by the fire, subject to lower uncertainty and characterized by higher versatility and robustness in breeding (*e.g.*, the B6 mouse).

Scientists more exposed to the shock have shifted their adoption from the mice in short supply to these same mice. We further find that scientists incur only temporary losses in productivity, but permanently change the type of their research along several dimensions. This is again consistent with the existence of switching costs and an interpretation of different mice strains as imperfect substitutes. The adoption of new mice strains comes with switching costs that affect productivity in the short run. In the long run, the adjustments to the type of research induced by the adoption of imperfect substitutes create further costs for scientists who consider switching back to the mice they had previously used.

The existence of frictions that potentially undermine the functioning of markets is a central tenant in the literature on innovation diffusion (David, 1985; Dosi, 1982). Our results demonstrate the importance of these frictions for the adoption of research tools, which is of particular importance due to their role in the process of knowledge production (Mokyr, 2002; Stephan, 2012). Moreover, we pin down the role of one key mechanism: demand-side switching costs.

The most relevant implication of these findings is that research tools markets may not naturally converge to any specific or desirable equilibrium. Due to path dependency, idiosyncratic shocks and distortions may have enduring effects. This underscores the importance of institutional interventions and community-wide efforts to steer research tools markets' trajectories. Indeed, the biomedical community is engaged in activities that promote the diffusion of best

practices and knowledge, and laboratories such as JAX take a central part in such activities (*e.g.*, training opportunities, knowledge platforms, coordination, and centralized production). Analogous activities exist also in different research fields and innovation sectors, more broadly. Our evidence provides an economic *raison d'etre* for such efforts that, arguably, are not the natural outcome of pure market mechanisms.

This study does not put into question the potential benefits of markets for research tools or the role of institutions that intend to centralize their production. On the contrary, we posit that they underpin the processes of knowledge sharing and standardization that favor diffusion (Furman and Stern, 2011). If anything, we see that more established research tools (those available from multiple suppliers and accompanied by a broader knowledge base) were resilient to the shock. Self-production alone would be most likely associated with more severe frictions to diffusion and higher risks for scientists (Baruffaldi and Gaessler, 2021). However, testing this broader hypothesis escapes the possibilities of our empirical setting and is left for future research.

2

The Innovation Effect of Drug Approval

2.1 Introduction

Technological change in healthcare and pharmaceutical markets has been a main driver for progress in human health. This applies especially to the development of new pharmaceutical treatments, which are responsible for a substantial increase in life expectancy (Lichtenberg, 2019). A variety of case studies show that many of these health improvements are related to radical innovations (see for discussion, *e.g.*, Kyle, 2020).[1] However, pharmaceutical companies invest also a substantial amount of their resources into incremental follow-on innovations. These are modifications over existing drugs, such as dosage formulations, the discovery of new therapeutic use, or drug combinations (European Commission, 2009).

There is an ongoing economic and political discussion regarding the net consumer surplus derived from incremental innovations in the pharmaceutical industry (Yin, 2017). Some incremental innovations are meaningful improvements to therapeutic treatments for the benefit of society (Arcidiacono et al., 2013; Bokhari and Fournier, 2013), which I define as *"improvement innovations"*. However, other incremental innovations are argued to be trivial modifications to the original drug that have little or no therapeutic benefits. The latter type of innovation activities is conducted, *i.a.,* to strategically enforce patent protection surrounding the focal invention (Amin and Kesselheim, 2012; Frakes and Wasserman, 2020; Gurgula, 2020).[2] In the case of pharmaceuticals, this behavior aims at prolonging market exclusivity periods to

---

[1] Some examples are the development of sulfa drugs (Jayachandran et al., 2010), vaccines (Whitney et al., 2014), new cancer drugs (Howard et al., 2016; Lichtenberg, 2018), or new treatments for cardiovascular disease (Cutler et al., 2007).

[2] This is commonly described as strategic patenting, which can be defined as the use of the patent system to leverage "complementarities between patents in order to attain a strategic advantage over technological rivals" (Von Graevenitz et al., 2007).

limit and delay generic competition (also referred to as life-cycle management). This comes at the expense of longer effective patent lives delaying generic entry and allowing the originator company to earn supracompetitive profits (Budish et al., 2015). I posit that patenting reflects underlying real (although marginal) innovation activities. I define this type of innovation activities as *"enforcement innovations"*.[3]

The incentives of pharmaceutical companies to invest in improvement and enforcement innovations may change over the drug life cycle. In the course of this, the event of drug approval is of capital importance. The transition from pre-approval to post-approval may have two countervailing effects on follow-on innovation activities: on the one hand, entering the market should stimulate originators to protect the approved drug from competition and to improve its features. This would lead to an upsurge in enforcement and improvement innovation activities after the drug's marketing authorization (European Commission, 2009; Sternitzke, 2010, 2013). On the other hand, the marketing authorization impedes the enforceability of follow-on patents that are filed after approval. There are multiple reasons for this, *e.g.*, (trial-related) information disclosed in the course of the drug approval can render an invention obvious and serve as novelty threatening prior art and, thus, increase the patentability standards (Breckenridge and Jacob, 2019; Kallenbach and Vallazza, 2018; Mello et al., 2013).[4] If the marketing authorization, in turn, reduced the effectiveness of follow-on patents filed after approval, this would decrease the incentives of originators to invest in enforcement innovations. In contrast, improvement innovations, more likely to fulfill the requirements of novelty and non-obviousness, should be unaffected by the changes in enforceability. Thus, this study asks the question: is the marketing authorization an effective public institution to reduce enforcement innovations that are presumably without actual therapeutic benefit?

To this end, I collect novel data on pharmaceutical patents that combines detailed information on the focal inventions, follow-on innovation activities, and data on marketing authorizations. I exploit the fact that follow-on innovations can be observed through patent citations (Galasso and Schankerman, 2015). Originator companies, which develop and commercialize a new drug, file first for a "primary" patent covering the new molecular entity (NME). Thus, forward citations to this "primary" NME patent capture all innovation activities building up and being related to the focal invention. Among those patented follow-on innovations, I distinguish between those presumably related to enforcement and those presumably related to improvements. First, I identify "secondary" patents, which cover subsidiary features of the drug such as dosage formulations. These patents are commonly used to delay generic entry (*e.g.*, Amin and Kesselheim, 2012) and, thus, are more likely to be the result of enforcement innovations. In contrast, I associate follow-on "product" patents with improvement innova-

---

[3]Other scholars define the "alteration of one or more technical elements of a product to limit or eliminate competition" as predatory innovation (Schrepel, 2018).

[4]Additionally, in some jurisdictions such as the United States (U.S.), institutional reasons require all patents that protect the focal drug from generic entry to be listed in U.S. Food and Drug Administration's Orange Book *before* the first drug approval (21 U.S.C. §355(b)(1) & (2)).

tions. Second, I identify whether patents are filed within the same disease category and, thus, being more marginal compared to patents filed in different disease categories.

Successful drug candidates are on average more valuable and, hence, more likely to receive follow-on innovations than drug candidates that never receive approval. Therefore, I compare only approved drugs with each other and leverage plausibly random differences in the timing of drug approval relative to the patent filing, called *time to approval* (*e.g.*, Gilchrist, 2016). Focal inventions that have not yet obtained a marketing authorization serve as a control group for focal inventions with early approval. I provide evidence that inventions with early and late marketing authorization do not differ in ex-ante patent or drug characteristics such as the underlying technology, the timing of the patent grant, or the field of application (*e.g.*, disease categories), and exhibit parallel pre-approval trends.

First, I document the presence of considerable follow-on innovations related to improvement and enforcement after the drug approval, both by the originator, vertically related firms, and generic manufacturers. This is in line with prior literature (Howard, 2007; Sternitzke, 2013). While in the beginning of the drug life cycle, follow-on innovation activities are typically conducted by the originator company, this activity shifts towards third parties over time.

Second, I find support for the hypothesis that enforcement innovation activities decrease substantially after the marketing authorization. This decrease is observed for secondary patents and patents in the same narrow field of application. The immediate timing of the decrease in follow-on innovation provides suggestive evidence for the proposed enforceability mechanism. Consistent with this, I do not find similar effects on enforcement innovation for prior milestones in the drug development process, such as the completion of phase II clinical trials, that are presumably not related to increases of patentability standards. Moreover, I illustrate that the decrease in innovation is not only associated with the focal firm, but also with third parties, who are likely in a vertical relationship with the originator.[5] Thus, my findings suggest that the change in incentives provided by the marketing authorization also affects enforcement innovation activities of other firms along the value chain.

Third, I show that improvement innovation activities, which are presumably not affected by the increase in patentability standards at the marketing authorization, remain unaffected by the drug approval.[6] Originators seem to shift their resources towards more novel innovations, such as new compounds and new diseases. Thus, marketing authorization-related increases in patentability standards, which decrease the enforceability of marginal follow-on patents, delineate incremental innovation activities: investments in improvement become relatively more important than investments in enforcement with little therapeutic benefit.

---

[5]It is a common feature in the pharmaceutical industry that smaller biotech companies develop and patent the NME and license them to larger firms with experience in designing clinical trials and commercializing/marketing drugs (Higgins and Rodriguez, 2006).

[6]If at all, the decrease in uncertainty coming from the public validation of the safety and effectiveness of the drug, may make incremental innovations with therapeutic benefit more worthwhile.

Finally, I can exclude a variety of alternative explanations for my findings. I show that the effects are not driven by differences in incentives for competitive entry by holding the market exclusivity period constant. To this end, I exploit the discontinuities in the EU patent term extension regime related to pharmaceutical products. Furthermore, I account for strategic timing decisions, such as accelerated market entry for more valuable drugs. The results suggest that the true effect is likely of a larger magnitude than the estimated effect (lower bound). Furthermore, I show that the decrease in enforcement innovation after the drug approval is also not driven by individual diseases.

I contribute to several strains of literature. First, I extend the literature on *patent fencing* and *evergreening* in the pharmaceutical industry (*e.g.*, Abud et al., 2015; European Commission, 2009; Gupta, 2020; Hemphill and Sampat, 2012; Sternitzke, 2010, 2013) and show that secondary patenting becomes less likely with drug approval. This is presumably related to the obstacles in enforceability of follow-on patents filed after approval (*e.g.*, novelty threatening prior art disclosed at the marketing authorization).

More generally, my findings relate to the literature on the incentives for innovation in the pharmaceutical industry, *e.g.*, market size (Acemoglu and Linn, 2004; Blume-Kohout and Sood, 2013) or intellectual property rights (Gaessler and Wagner, 2020; Kyle and McGahan, 2012). I show that the type of incremental innovation activities carried out changes with drug approval. Originators can streamline the innovation activities by investing less in life-cycle management activities like enforcement innovations.

Lastly, I contribute to the literature on patents and follow-on innovation. It has been established that patents have a negative impact on third-party follow-on innovation, *e.g.*, due to bargaining failures (Galasso and Schankerman, 2015) or by restricting the freedom-to-operate (Gaessler et al., 2019).[7] This study provides evidence that there are additional factors relevant in the relationship between patents and follow-on innovation, such as the impact of product commercialization – here proxied by drug approval.

This paper is structured as following: Section 2.2 describes the drug development process and pharmaceutical patenting. Section 2.3 elaborates on the data, summary statistics and provides descriptive evidence. Section 2.4 details the empirical strategy. Section 2.5 presents the estimation results and Section 2.6 discusses alternative explanations. Section 2.7 concludes.

---

[7]Extensive empirical literature investigates whether and at which magnitude patents decrease follow-on innovation below the socially desirable. It consists of both technology specific studies largely focusing on Biotechnology or Pharmaceuticals (Gaessler and Wagner, 2020; Murray et al., 2016; Sampat and Williams, 2019; Williams, 2013) and of across-technology studies (Gaessler et al., 2019; Galasso and Schankerman, 2015).

2.2 Drug Development and Patenting

Drug development is a costly endeavor. Recent estimates of the average capitalized costs per drug oscillate around \$2.5 billion depending on the disease category (DiMasi et al., 2016). These costs are the result of extensive clinical trials and approval processes with an average development time of about 12 years (Adams and Brantner, 2006). Additionally, pharmaceutical companies spend around \$300 million related to post-approval R&D, which includes enforcement innovations, improvement innovations, and the monitoring of possible adverse reactions and/or new side effects (DiMasi et al., 2016). In total, the pharmaceutical industry invests around 17 percent of its global revenues derived from prescription drugs in R&D (European Commission, 2009).

The drug development process illustrated in Figure 2.1 starts with the discovery of an active ingredient. The initial discovery phase is followed by (pre-)clinical trials once a concrete field of application is found.[8] This period can be described as the pre-approval period. Successful development is verified in the marketing authorization process.[9] Subsequently, pharmaceutical companies (either the originator, vertically related firms, or licensees) are allowed to launch and sell the drug for indications approved by the marketing authorization. In the initial years after the commercialization, the drug is typically protected by various forms of market exclusivity, such as data exclusivity and patents. Thus, the originator can sell the newly developed drug without competition from generic products. After patent expiry, generic products enter the market (European Commission, 2009) and prices fall drastically (Morton and Kyle, 2011).

Intellectual property rights are considered to be the most important incentives in the development of new drugs (Mansfield, 1986).[10] Typically, promising drug candidates (active ingredients/NME) are included in patent filings directly after discovery. These initial patents are considered "primary patents" and have a fixed patent term of 20 years.[11] In some jurisdictions, such as the EU, pharmaceutical companies can extend the patent term by up to 5 years. Besides, pharmaceutical companies file additional patents (called "secondary" patents) covering further subsidiary characteristics related to the drug, such as the different dosage forms

---

[8]Four phases of clinical trials can be differentiated: in phase I, originators assess the toxicity of a new drug candidate. Phase II proves efficacy. Phase III involves long-term trials with large patient groups. This phase is also the origin of development of novel pharmaceutical formulations and dosage forms, which might be used for strategic patenting purposes. Phase IV follows the product launch. Originators are required to monitor possible adverse reactions and/or new side effects (European Commission, 2009). The duration of these clinical trials depends on exogenous characteristics, *e.g.*, related to the firm (Wagner and Wakeman, 2016). However, there are considerable random elements unrelated to firm or drug (Gilchrist, 2016).

[9]Marketing authorization is either granted by the European Medicines Agency or nationally (European Commission, 2009). It verifies that medicines are safe, effective, and of good quality.

[10]Various empirical studies show that drug patents are an important determinant of investments in pharmaceutical R&D (*e.g.*, Kyle and McGahan, 2012; Gaessler and Wagner, 2020). Moreover, various scholars highlight the importance of the design of patent characteristics (Budish et al., 2015; Gilchrist, 2016; Izhak et al., 2020).

[11]Budish et al. (2015) describe why this fixed patent term might distort the incentives for firms to invest in projects with long clinical trials. The residual exclusivity time is too short to recover the investment costs. Therefore, I conduct a robustness check in which the market exclusivity period does not change with the timing of the marketing approval.

(*e.g.*, tablets or capsules) or particular pharmaceutical formulations (*e.g.*, mixtures of active agents), methods of delivery, or new combinations of active substances. The ratio of primary to secondary patents is roughly 1:7 (European Commission, 2009).

Sternitzke (2013) describes that these secondary patents are often filed strategically to extend the market exclusivity time (called *evergreening*) and to broaden the scope of the patent protection (called *patent fencing*). In a recent study, Gupta (2020) shows that secondary patents effectively extend market exclusivity periods. However, the incentives for originators to invest in innovation activities may decrease after the marketing authorization, if the invention's only purpose was enforcement without actual therapeutic improvements. This is because clinical trial documents published in the course of the marketing authorization may create obstacles to meet the patentability requirements "novelty" and "inventive step/non-obviousness". Several decisions of the European Patent Office's (EPO) Board of Appeal show that the disclosure of clinical trial protocols and results, *i.e.*, in scientific journals, was prejudicial to the novelty of secondary patents if they allow a skilled person to derive the follow-on invention with a high degree of certainty (see *e.g.*, EPO Boards of Appeal decisions T 385/07, T 715/03, T 1859/08, T 158/96, T 2506/12, T 1031/00). The same applies to non-obviousness. Disclosure of the trial-related information can render an invention obvious if it increases the likelihood that an average skilled person would solve the technical problem at hand in light of the disclosed information (see *e.g.*, EPO Boards of Appeal decisions T 1745/12, T 1493/09, T 2506/12, T 0385/07). In some of these cases, the disclosure of the trial-related information invalidated secondary patents (Breckenridge and Jacob, 2019; Kallenbach and Vallazza, 2018; Mello et al., 2013).[12] Moreover, in some jurisdictions such as the U.S., institutional reasons require all relevant secondary patents to be listed in the U.S. Food and Drug Administration's (FDA) Orange Book *before* the timing of the first drug approval (21 U.S.C. §355(b)(1) & (2)).[13] Both circumstances create obstacles for the enforceability of follow-on patents filed after approval since they are more likely to be annulled in opposition or subsequent litigation (European Commission, 2009).

---

[12]For further discussions, see:
https://www.iam-media.com/clinical-trial-disclosures-obstacle-patentability-in-europe [last accessed on March 8, 2021] or https://www.lexology.com/library/detail.aspx?g=51697502-e1a1-4639-9c32-f45e298e2675 [last accessed on March 8, 2021].

[13]Following the Hatch-Waxman Act in 1984, a generic manufacturer has to wait until all of the patents listed in the FDA's Orange Book expire or file a "Paragraph-IV challenge" by successfully showing non-infringement or invalidity of all these patents (Branstetter et al., 2016). For further discussions, see:
https://www.finnegan.com/en/insights/articles/requirements-benefits-and-possible-consequences-of-listing-patents-in-fdas-orange-book.html [last accessed on March 8, 2021].

Figure 2.1: Drug development process and patenting
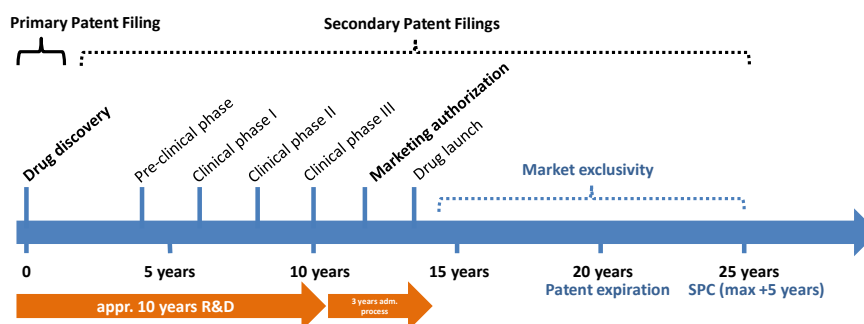
Notes: The figure illustrates the typical drug development and patenting process following the European Commission (2009).

2.3 Data and Summary Statistics

2.3.1 Data

This study aims at investigating whether drug approval decreases enforcement innovations. However, linking inventions (focal patents) and products is a difficult endeavor (de Rassenfosse, 2018). This is because many inventions relate to more than one commercialization attempt and many products are based on more than one technology. The pharmaceutical industry provides a unique setting since few patents protect a drug, and few drugs are protected by the same patent (Cohen et al., 2000). This allows me to clearly link a focal invention – the drug – to its primary patent, the drug approval, and all related incremental innovation activities.

Patent-Drug Link

I use data on primary patents covering a NME and the approved drug(s) from public registers. The data is available because firms can apply for the extension of their patent term by applying for Supplementary Protection Certificates (SPC) in the European Union.[14] A formal requirement is to provide information on the *patent-drug link*. National patent authorities, like the German Patent and Trade Mark Office (DPMA), collect and publish this information.[15]

I collect all patent-drug links at the DPMA until September 2018 (Figure 2.2a). This data includes the primary patent application number, the drug name, the patent applicants, and all relevant dates concerning the drug approval, such as the date of the first EU marketing autho-

---

[14]Available empirical studies that investigate SPC have descriptive character and analyze primarily the determinants of SPC usage as well as application outcomes (Kyle, 2017; Mejer, 2017; Gaessler and Byrski, 2018).

[15]Similar data sets are available for example at the FDA with the Orange Book. The FDA includes all secondary patents related to the drug. In this study, I rely on forward citations instead, since it allows me to differentiate between follow-on innovation activities related to enforcement and improvement.

rization. The full data set comprises 1,405 patent-drug links[16] with 1,135 unique INPADOC patent families.

Originators can wait until the patent expiry day to file for a SPC and, thus, to include the patent-drug link into the database. Therefore, more recent patent cohorts may suffer from truncation and selection problems since only a (non-random) subset of patent-drug links is included. Hence, I restrict the sample using the following steps. First, I account for the truncation issues by excluding all patents that are filed within the last 20 years before data collection (1997+patent term). This has the additional advantage of including only similarly relevant patents and drugs into the analysis. Figure 2.2b shows that the 1997-restricted sample still includes a large share of the primary patents.[17] Second, I only keep those observations that relate to the first drug associated with a primary patent.[18]

The data provided allows me to calculate the *approval-lag*. It is computed between the priority date of the focal drug's primary patent and the date of the first EU marketing authorization.[19] I exclude extreme values in the time to approval. Setting the cut-off point at 90%, which corresponds to 16 years, this restriction allows me to investigate a full five-year post-approval period for all observations.[20]

The final data set eventually comprises 590 unique patent families that correspond to 590 unique drugs. This is a similar magnitude compared to other studies using drug approval data (Gilchrist, 2016). Due to the 1-to-1 relationship, I use the expressions *"(focal) drug"*, *"primary patent"*, and *"(focal) invention"* interchangeably.

Patent Data

I link the resulting data set to further patent information on the primary patent from Patstat via the application number and collect all patent information at the INPADOC patent family ID level.[21] The main dependent variable is the annual number of EPO forward citations, which the primary patent received from subsequent EP patents.[22] Prior literature has argued that

---

[16]The SPC data comprises also plant protection patents, which I exclude in this study.

[17]All results are quantitatively similar when using the larger full sample in Appendix Figure B-11.

[18]Some firms try to apply for more than one SPC for one drug patent, which yields several patent-drug links. These SPC applications are registered in the data set, but in most cases *rejected* (Kyle, 2017; Mejer, 2017; Gaessler and Byrski, 2018), so that I drop duplicative SPC filings.

[19]In the SPC data the approval-lag, which is the basis for the calculation of the patent term extension, is calculated using the national application year. Since forward citations can occur from priority onward, I define the priority date as the "real" starting point of an invention's age. Moreover, the SPC data includes all marketing authorization dates in the EU. I choose the first EU marketing authorization for this study.

[20]The last patent is filed 1997 with a priority year in 1996. Forward Citations are observed until 2017. 1996 plus a maximum of 16 years time to approval leaves a post period of 5 years for the last patent filed in the sample. All results are robust to alternative or no thresholds (Appendix Figure B-11).

[21]The pharmaceutical industry is characterized by a high number of continuations. These continuations are frequently not included in narrower patent family definitions (DOCDB). For a discussion, see Martinez (2010).

[22]Citations are aggregated at the INPADOC family level on both sides: cited INPADOC patent family ID and citing INDPADOC patent family IDs.

Figure 2.2: Distribution of SPC filings and patent priority filings over time

(a) SPC filings - full sample | **(b)** Primary patent priory filings - full sample
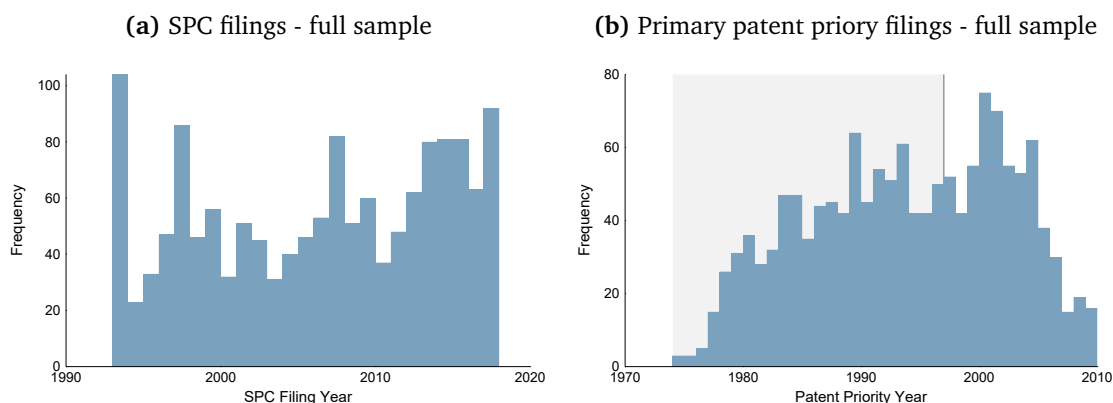
Notes: Figure (a) presents the distribution of all 1,405 SPC filings over time (full sample). Each SPC filing includes a patent-drug link, which is used throughout the analysis. A SPC is typically filed at the end of a patent's lifetime. Figure (b) shows the distribution of the corresponding patent priority filings over time. The grey window shows the restricted sample with 590 unique patent-drug links.

forward citations serve as a proxy for follow-on innovation and knowledge spillovers.[23] If not stated otherwise, the citation counts are log-transformed with one being added before taking the logarithm to include patents with no forward citations in the respective year.

Furthermore, I extract the harmonized firm names curated by the OECD (Harmonised Applicant Names – HAN) of all applicants mentioned in all cited and citing patent publications. I use this information to distinguish between self- and other companies' citations following the approach outlined in Gaessler et al. (2019). I investigate the nature of these third party citations by manually checking the applicant. The 20 largest pharmaceutical firms are related to 50% of the forward citations. Some citations are, however, generated by firms that focus on generic manufacturing. Additionally, I differentiate between examiner-generated and applicant-generated citations and the citation type. To this end, I identify X and Y citations that reference the primary patents as validity-threatening prior art (*e.g.*, Harhoff and Wagner, 2009; Nagler and Sorg, 2020; Wagner and Wakeman, 2016). Moreover, I identify further patent characteristics like the patent family size or the number of applicants, and identify the first patent grant within the patent family.

Drug Data

I complement the publicly available register data by adding drug-level characteristics from the proprietary pharmaceutical Clarivate Analytics' Cortellis Investigational Drugs database (commonly known as Cortellis) at the patent, drug, and firm level. I link cited and citing patents by their INPADOC patent family ID to the Cortellis database. If citing patents are successfully linked, I define the forward citation as being in the "pharmaceutical" domain. Crucially, Cortel-

---

[23]EPO citations are less strategic than citations generated at the United States Patent and Trademark Office. For a discussion, see Gaessler et al. (2019).

lis categorizes patents by their relationship to the respective drug (called "patent type"). This allows me to differentiate citing patents into product patents (*e.g.*, new products, new product derivatives, new macromolecule), process patents, and secondary patents (*e.g.*, new formulations, new dosage forms, new drug combinations, or new use). I use the "product" patent category as a proxy for innovation activities that are more likely to involve actual improvements, and the "secondary" as well as "process" patent categories as proxies for innovation activities that are more likely related to enforcement.

Moreover, I use the Cortellis link to investigate how marginal citing patents are compared to the cited patent. To do so, I identify whether the patents belong to the same narrow field of application – here the drug market is defined by diseases. In Cortellis, each patent is associated with *targeted conditions*, which I assign to the World Health Organization's ICD-9 disease codes using the crosswalk by Dranove et al. (2020).[24] I use this data to identify whether each citing-cited patent dyad overlaps in at least one disease category.

To identify the origin of the follow-on innovation activities, I rely on the drugs' full development history and retrieve information on the timing of clinical trials for each focal patent-drug link.[25] If the citing companies conducted clinical trials that are related to the focal drug, I define their citations as being generated from a vertically related firm.

2.3.2 Summary Statistics

In this section, I summarize the data and investigate whether the approval lag is correlated with observable ex-ante patent and drug characteristics. This analysis as well as Section 2.3.3 inform and motivate the empirical strategy, in which I exploit staggered drug approval.

First, I divide the sample into two equally sized groups, using as a threshold the median of the time to approval, as illustrated in Figure 2.3. It shows inventions that obtained the first marketing authorization within 10 years (early MA) and inventions that obtained the first marketing authorization between 11 to 16 years (late MA) after the priority date. The distribution of the time to approval is smooth between 5 and 16 years suggesting the absence of structural breaks. By definition, a longer time to approval leads to marketing authorizations occurring in later years. These characteristics are in line with Sternitzke (2010), who finds the average approval lag to be 11 years. The market exclusivity period, shown in Appendix Figure B-2, is uniform in the early MA group at 15 years but decreases if the approval lag exceeds ten years. I exploit this discontinuity in Section 2.6 when investigating alternative explanations.

Figure 2.3c illustrates that the annual distribution of patent priority filings is largely uncorrelated with the approval lag. This suggests the absence of cohort effects. Overall, citing patents' priority filings are similarly distributed, as well, with one notable exception (Figure

---

[24]Dranove et al. (2020) had two expert medical coders independently code the concordance between Cortellis indications and ICD-9 codes. I am grateful to Manuel Hermosilla for sharing the mapping.

[25]Some patents are related to more than one drug. Thus, I manually evaluate which of these drugs corresponds most likely with the drug in the SPC data set. To this end, I compare the drug names conditional on being authorized in a similar time period.

Figure 2.3: Distribution of further timing related variables

(a) Time to approval

(b) First EU marketing authorization

(c) Primary patent priority filings

(d) Citing patent filings

(e) Time to patent grant

(f) Primary patent grants
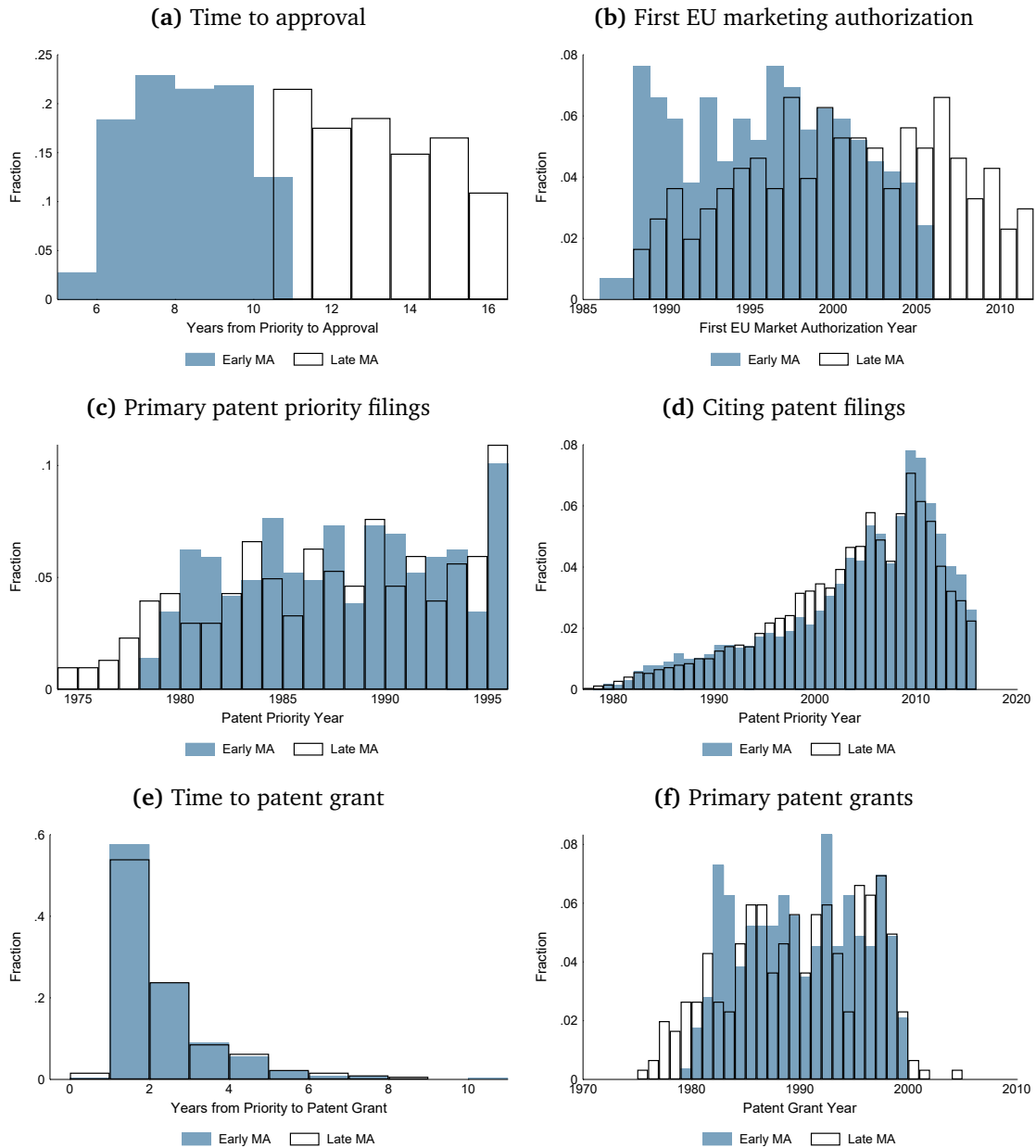
Notes: The figures compare the distribution of timing related variables between early MA and late MA drugs. The main sample of 590 patent-drug links is split in two equally sized groups, using as threshold the median of the approval lag.

2.3d). Primary patents in the early MA group are referenced in fewer follow-on patents from 1995 onward. This is precisely at a time, where the majority of these early MA primary patents received their first marketing authorization.

Additionally, the primary patent's first grant year and the *time to patent grant* are equally distributed (Figures 2.3e & 2.3f). This is important because one could assume that patents

with a late drug approval are associated with an inherently different technology or differential firm incentives.[26] For example, prior research shows that more valuable patents have a shorter patent pendency time at the respective patent offices (Harhoff and Wagner, 2009). However, the time to patent grant does not differ across different drug approval lags.

Second, I compare the same groups of focal inventions along a variety of patent and drug characteristics in Table 2.1. This provides further confidence in the idea that the approval lag is unrelated to any other initial patent and drug indicator that could potentially affect the evolution of forward citations.[27]

The dependent variables, such as self-citation counts or secondary patent citation counts, exhibit a similar magnitude in the first years after the priority filing. The mean levels of forward citations in the initial year are not statistically different from each other. The same applies to the primary patent-related covariates, which are determined at or shortly after the patent filing. Except for a few cases, they are uncorrelated with the approval lag. For example, primary patents have a similar family size, number of applicants, likelihood of being filed internationally, and likelihood of being filed in all three major patent offices (EPO, USPTO, JPO). This indicates an ex-ante similar private value between the patents (see, *e.g.*, Harhoff et al. (1999, 2003) for a discussion). However, early-approval patents are filed more often under the technology classification pharmaceuticals and less often in biotechnology. If the nature of the technology influences the timing of the marketing authorization as well as the forward citation patterns, this could potentially bias the results. Therefore, the empirical strategy accounts for those differences by using patent fixed effects. Drug-related covariates (combination of active ingredients and salts) are equally distributed.

To shed further light on whether the nature of the drugs differs by the timing of the marketing authorization, I exploit the disease categories associated with each invention. At large, pharmaceutical patents with early or late MA also do not differ with regard to the indications listed in the patent applications. The five most prominent disease categories are "infectious and parasitic diseases", "neoplasms", and "endocrine, nutritional and metabolic diseases, and immunity disorders", "diseases of the nervous system and sense organs", and "diseases of the circulatory system" (Table 2.2). The likelihood of being associated with one of these top 5 categories does not differ across patent-drug approval pairs except for "endocrine, nutritional and metabolic diseases, and immunity disorders". This shows that early and late approved drugs occur among similar diseases. However, it might still be the case that unobserved differences at a lower aggregation ICD-9 disease level may drive different dynamics in generating forward citations. Thus, I will account for this in Section 2.6. In a nutshell, focal drugs with early and late MA are similar concerning most ex-ante patent and drug characteristics. This is important because it constitutes the basis for the definition of treatment and control.

---

[26]Prior literature suggests that economic outcomes are related to the timing of the grant decision, *e.g.*, likelihood of licensing (Gans et al., 2008) or the willingness to invest in commercialization (Wagner and Wakeman, 2016).

[27]Summary statistics without the sample split can be found in Appendix Tables B-1 and B-2.

Table 2.1: Mean comparison – ex-ante patent and drug characteristics

Early MA vs Late MA	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Early MA (N = 288)			Late MA (N = 302)				
	Mean	Median	Std. Err	Mean	Median	Std. Err	Diff	p-value
Time to approval	8.26	8	1.3	13.21	13	1.7	4.94	0.000***
Time to patent grant	2.33	2	1.2	2.44	2	1.4	0.11	0.322
Patent priority year	1987.38	1987	5.1	1987.00	1987	5.8	-0.39	0.394
First patent grant year	1989.66	1990	5.4	1989.36	1990	6.3	-0.30	0.533
First MA year	1995.64	1996	5.0	2000.17	2000	6.1	4.53	0.000***
Market exclusivity	15.00	15	0.0	12.88	13	1.7	-2.12	0.000***
Initial forward cit.	0.66	0	1.1	0.64	0	1.1	-0.02	0.812
Initial self cit.	0.27	0	0.5	0.28	0	0.5	0.02	0.684
Initial other cit.	0.40	0	0.9	0.36	0	0.8	-0.04	0.586
Initial same ICD9 cit.	0.25	0	0.5	0.25	0	0.5	0.00	0.967
Initial other ICD9 cit.	0.03	0	0.2	0.04	0	0.2	0.00	0.915
Initial biotech patent cit.	0.00	0	0.0	0.00	0	0.1	0.00	0.329
Initial secondary patent cit.	0.13	0	0.4	0.10	0	0.3	-0.03	0.334
Initial process patent cit.	0.02	0	0.2	0.03	0	0.2	0.00	0.875
Initial product patent cit.	0.19	0	0.4	0.21	0	0.5	0.02	0.544
Pediatric drug	0.07	0	0.2	0.07	0	0.2	0.00	0.990
Drug combination	0.30	0	0.5	0.31	0	0.5	0.01	0.810
Salt of drug molecule	0.14	0	0.4	0.18	0	0.4	0.04	0.229
Size of patent family	26.83	24	16.7	24.77	24	14.6	-2.06	0.112
Number of applicants	1.09	1	0.3	1.10	1	0.3	0.01	0.827
Transn. patent family	0.87	1	0.3	0.85	1	0.4	-0.01	0.630
Triadic patent family	0.52	1	0.5	0.56	1	0.5	0.05	0.268
Tech area organic chem.	0.44	0	0.5	0.50	1	0.5	0.07	0.110
Tech area pharma.	0.47	0	0.5	0.36	0	0.5	-0.10	0.010***
Tech area biotech.	0.06	0	0.2	0.10	0	0.3	0.05	0.035**
Tech area material chem.	0.01	0	0.1	0.01	0	0.1	-0.00	0.616
Applicant country US	0.36	0	0.5	0.33	0	0.5	-0.03	0.497
Applicant country Europe	0.47	0	0.5	0.43	0	0.5	-0.04	0.311

Notes: This table compares ex-ante patent & drug characteristics split at the median approval lag with t-tests. The unit of observation is the unique patent-drug level. *Initial* forward citation counts include all patent references within 12 months from the primary patent's priority filing. Significance levels:* $p<0.1$, ** $p<0.05$, *** $p<0.01$.

2.3.3 Descriptive Analysis

If the marketing authorization decreased the incentives for the originator to invest in incremental innovation activities related to enforcement, the trend in forward citations should diverge from the marketing authorization onward. In the following figures, focal drugs are again split by the median time to approval. Thus, inventions in the early MA group obtain the marketing authorization between year 5 and 10 (median year 8), which is highlighted by the grey bar. Inventions in the late MA group are not associated with a drug approval at this point in time.

Figure 2.4 presents the development of the average yearly number of forward citations. It becomes apparent that the forward citation count develops parallel before any approval and diverges precisely around the timing of the early marketing authorizations (between year 5 and 10). Inventions with early MA start to accumulate fewer forward citations after the drug

Table 2.2: Mean comparison – disease characteristics

Early MA vs Late MA	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Early MA (N = 288)			Late MA (N = 302)				
	Mean	Median	Std. Err	Mean	Median	Std. Err	Diff	p-value
Number of ICD9	2.32	2	2.4	2.26	2	2.4	-0.06	0.790
Infectious/parasitic diseases	0.23	0	0.4	0.26	0	0.4	0.03	0.470
Neoplasms	0.14	0	0.4	0.15	0	0.4	0.00	0.905
Endocrine/immun. disorders	0.12	0	0.3	0.18	0	0.4	0.06	0.066*
Blood diseases	0.03	0	0.2	0.02	0	0.2	-0.01	0.482
Mental disorders	0.05	0	0.2	0.09	0	0.3	0.04	0.061*
Nervous system diseases	0.12	0	0.3	0.11	0	0.3	-0.01	0.633
Circulatory system diseases	0.17	0	0.4	0.15	0	0.4	-0.03	0.435
Respiratory system diseases	0.07	0	0.3	0.05	0	0.2	-0.02	0.321
Digestive system diseases	0.06	0	0.2	0.04	0	0.2	-0.02	0.405
Genitourinary diseases	0.08	0	0.3	0.12	0	0.3	0.04	0.192
Pregnancy/childbirth	0.00	0	0.1	0.00	0	0.1	-0.00	0.952
Skin diseases diseases	0.07	0	0.3	0.02	0	0.2	-0.04	0.017**
Musculoskeletal diseases	0.07	0	0.3	0.09	0	0.3	0.01	0.578
Congenital anomalies	0.00	0	0.0	0.00	0	0.0	0.00	.
Conditions perinatal period	0.04	0	0.2	0.02	0	0.1	-0.02	0.213
Ill-defined conditions	0.09	0	0.3	0.08	0	0.3	-0.02	0.542
Ijury/poisoning	0.08	0	0.3	0.07	0	0.2	-0.01	0.541

Notes: This table compares disease characteristics split at the median approval lag with t-tests. The unit of observation is the unique patent-drug level. A primary patent can be associated with more than one ICD-9 category. Significance levels:* p<0.1, ** p<0.05, *** p<0.01.

approval (Figure 2.4a). This divergence becomes especially apparent when looking at self-citations within the originator company, which are the core of this study (Figure 2.4b). The divergence is permanent.

Secondary patents, which are the result of enforcement innovations, are filed to a substantial degree after the approval. This is fully in line with Howard (2007) or Sternitzke (2013), who show that pharmaceutical companies continue with patent filings after the approval. There is, however, also a substantial drop in self-citations related to secondary patents for the early MA group around the time of their respective drug approval. This divergence is not visible for product patents, which are more likely to involve actual improvements (Figures 2.4c and 2.4d). In the Appendix, I describe a variety of alternative outcome variables (Figures B-3 and B-4). In all cases, early MA and late MA inventions develop parallel in the pre-approval period and patent citations more likely related to enforcement innovation decrease for the early MA group after their approval.

These findings are supported when looking at more narrowly defined groups. In Figure B-5 in the Appendix, I split the early MA group and the late MA group at their respective median time to approval.[28] These subgroups of patents, which still differ by around 2 years in their time to approval, are supposedly even more similar concerning the type of drug/type of

---

[28]I split the early MA group at the median time to approval, which is 8 years, and I split the late MA group at the median time to approval, which is 12 years. All four sub-samples have roughly the same sample size.

technology. In both cases, the subgroups receiving the marketing authorization earlier diverge negatively after the approval.

This graphical analysis provides first descriptive evidence that enforcement innovation activities decrease after drug approval. Importantly, it shows that follow-on innovation activities have parallel trends before the marketing authorization.

Figure 2.4: Evolution of forward citations by *time to approval*
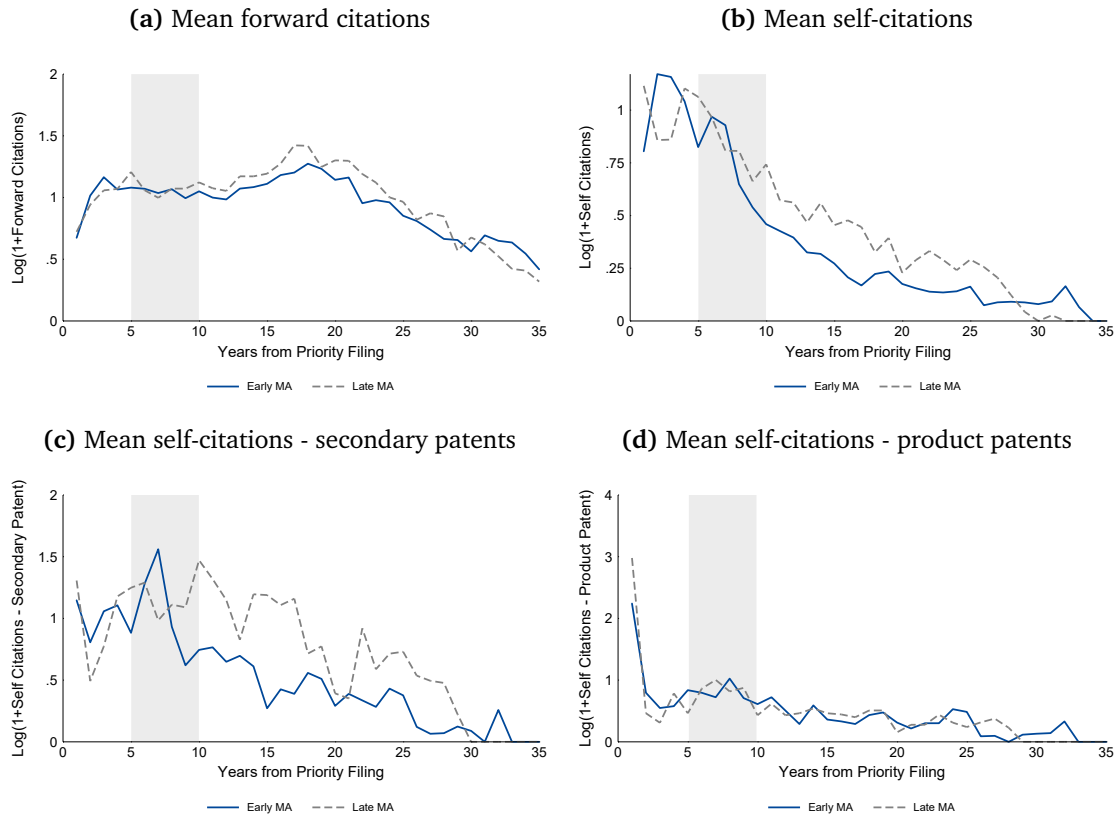
(a) Mean forward citations

(b) Mean self-citations

(c) Mean self-citations - secondary patents

(d) Mean self-citations - product patents

Notes: The figures present the average log-transformed number of forward citations split at the median approval lag (early/late MA) over time. The annual citation count is normalized by the average citation count from the pre-approval period t_0-t_5. The unit of observation is the unique patent-drug level.

2.4 Empirical Strategy

In a perfect experiment, I would be able to compare patented inventions for which a drug gets randomly approved to other inventions for which a drug randomly fails to obtain regulatory approval. However, the approval of drugs or failure in drug development is far from random but is determined by the quality of the underlying invention or strategic R&D and market-related decisions. Therefore, I hold the existence of drug approval constant and leverage the timing instead. It cannot be predicted perfectly at the time of the patent filing whether and when a drug associated with the patent will be on the market. Gilchrist (2016) describes a

variety of reasons why this period involves random elements unrelated to eventual drug and firm characteristics, such as uncertainty in the development process. Strategic delay of marketing authorizations is costly since the product would enter the market later and may lose first-mover-advantages (Lakdawalla, 2018; Robinson et al., 1994). Thus, drugs associated with a late approval serve as a counterfactual scenario for drugs associated with an early approval. This depends on the assumption that forward citations would have developed similarly between those two groups in the absence of the approval. I find support for this assumption given the high similarity between early and late approved drugs in terms of ex-ante primary patent/drug characteristics and the observed parallel trends before approval in Sections 2.3.2 and 2.3.3. Thus, I expect a primary patent's forward citation count to change after drug approval relative to primary patents with *later* drug approval.

For this reason, I employ an event study design, which follows Schmidheiny and Siegloch (2019) and exploits the variation in the timing of the treatment (staggered drug approval) across inventions. It studies the dynamics of an outcome variable during a time span around a treatment. Therefore, I regress the number of forward citations in a year t on a set of nonparametric event variables, which indicate when the marketing authorization has happened relative to t. Formally, I estimate the following event study equation:

$$\mathbf{E}[y_{it}|X_{it}] = \exp[\alpha + \sum_{j=\underline{j}}^{\overline{j}} \beta_j \, MA_{it}^j + \delta_t + \theta_i], \quad (2.1)$$

where y_{it} represents the number of forward citations. δ_t is a citation year fixed effect, and θ_i a patent-drug fixed effect. The patent-drug FE and the citation year FE are capturing together the age of the invention. MA_{it}^j is a treatment indicator for the drug approval happening j periods away from t. I limit the effect window of the drug approval event that is allowed to influence the outcome variable, from $\underline{j} < 0$ periods before to the event to $\overline{j} \geq 0$ periods after the event. The effect on y is assumed to stay constant outside the defined effect window. This is achieved by binning the treatment indicators MA_{it}^j at the endpoints of the effect window. Here, at the endpoints $\underline{j} = -5$ and $\overline{j} = +5$ the treatment indicator is set to 1.[29] This indicates that at the endpoints the treatment for unit i at time t has happened $\underline{j}$ *or more* periods ago, or will happen in $\overline{j}$ *or more* periods into the future. Therefore, the coefficients corresponding to the endpoints cannot be interpreted. However, binning up the endpoints is essential in order to separately identify dynamic treatment effects β_j from secular time trends even when no never-treated units are present (Schmidheiny and Siegloch, 2019). The coefficient β_j of MA_{it}^j with $j < 0$ looks forward in time, $j = 0$ at present time, and $j > 0$ at past time. I cluster robust standard errors at the unique patent-drug level.

In my preferred specification, I control additionally for the time-varying effects of patent grants and patent term extensions (SPC grants) to disentangle further dynamic factors, which

---

[29]The results are robust to using larger effect windows (Appendix Figure B-9).

may impact forward citation patterns:

$$\mathbf{E}[y_{it} | X_{it}] = \exp[\alpha + \sum_{j=\underline{j}}^{\bar{j}} \beta_j \, MA_{it}^j + \sum_{j=\underline{j}}^{\bar{j}} \gamma_j \, patent_{it}^j + \sum_{j=\underline{j}}^{\bar{j}} \eta_j \, SPC_{it}^j + \delta_t + \theta_i], \quad (2.2)$$

where $patent_{it}^j$ is a treatment indicator for the patent grant happening j periods away from t and SPC_{it}^j is a treatment indicator for the patent term extension (SPC grant) happening j periods away from t.

For continuous dependent variables like log-transformed forward citation counts, I run linear regression models with high-dimensional fixed effects. In a robustness check, I estimate Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects for count dependent variables.[30]

2.5 Empirical Results

In this section, I present and discuss the results from the event study analysis outlined in Section 2.4. I evaluate whether a marketing authorization has an impact on enforcement and improvement innovation by originators and third parties.

2.5.1 Main Effect on Originators

First, I provide evidence for a decrease in innovation activities building up on the focal drug following the marketing authorization within the focal firm.

Figure 2.5a depicts event study estimates with log-transformed self-citation counts as the dependent variable. I estimate a linear regression with invention and calendar year fixed effects as well as indicators for alternative events such as patent/SPC grant (Equation 2.2). I do not observe (potentially confounding) differential trends on self-citations before the approval. This shows that related follow-on patents are not always filed shortly before approval such as at the end of clinical trials in phase III. If this was the case, late approved inventions would exhibit different pre-trends in (self)-citations. Thus, the absence of such "bunching" before approval provides suggestive evidence that any effect of the marketing authorization on follow-on innovation activities is not driven by timing effects.

After the drug approval, I find a significant decrease in self-citations. Thus, the focal firm seems to reference the primary patent less often in new patents after the related drug receives its first EU marketing approval. The decrease in self-citations happens immediately and is long-lasting. Figure B-9 in the Appendix shows that there is no recovery in self-citations. Prior literature has emphasized that outcomes of innovation inputs typically become visible only with a 2-3 year time lag (Hall et al., 2010). The immediate timing of the decrease in

---

[30]For linear regressions, I use the *reghdfe* Stata package based on Correia (2016). For PPML regressions, I use the *ppmlhdfe* Stata package as described in Correia et al. (2020).

follow-on innovation is consistent with the interpretation that the missing innovations are of enforcement type rather than actual improvements. This provides first suggestive evidence for the proposed enforceability mechanism.

The effects can be interpreted as semi-elasticities. After the drug approval, self-citations to the focal invention decrease by a magnitude of 4-5%. This decrease is of larger magnitude when looking at Poisson pseudo-maximum likelihood estimates with count data (Figure B-6 and Table B-12 in the Appendix). The effect size seems reasonable given that the enforceability of follow-on patents changes at the intensive and not extensive margin. As shown descriptively in Section 2.3.3, investments in incremental innovation remain an important feature of the pharmaceutical industry also post-approval (*e.g.*, Sternitzke, 2013).

All results are robust to alternative dependent variable constructions, such as winsorizing at the 99% or 95% level to account for outliers (Table B-11), and alternative specifications, such as the replacement of patent-drug fixed effects by patent cohort and ICD-9 category indicators, different control variables, or more restricted samples (Table B-7 in the Appendix). Moreover, they do not change when excluding controls for alternative events like the patent grant (following Equation 2.1 in Figure 2.5b). The findings are coherent with the interpretation that the marketing authorization changes the incentives for firms to conduct incremental innovation activities, which build up on the focal invention. In the next section, I investigate whether this decrease is driven by follow-on innovation activities related to enforcement.

Figure 2.5: Impact of marketing authorization on self-citations
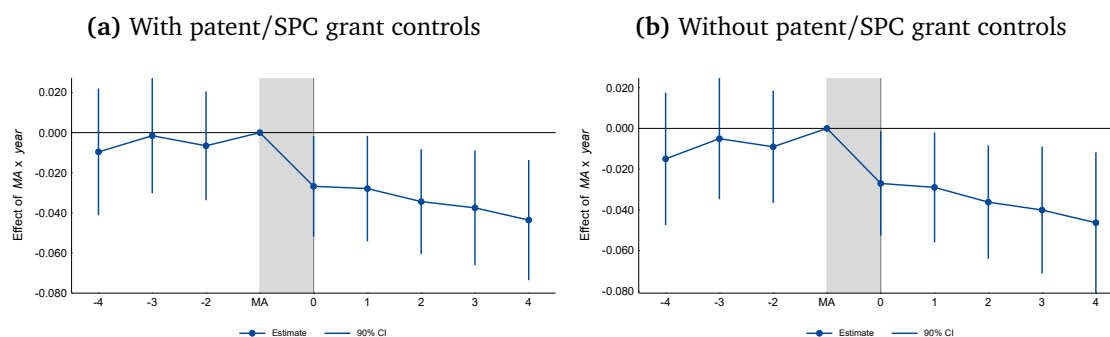
(a) With patent/SPC grant controls **(b)** Without patent/SPC grant controls

Notes: The figures show the event study estimates and the 90 percent confidence bands of linear regressions with high-dimensional fixed effects. The outcome variable is the log-transformed annual self-citation count. Figure (a) follows the preferred specification (Equation 2.2) and Figure (b) follows the specification without controls for patent grant and SPC grant (Equation 2.1). The unit of observation is the unique patent-drug level. Standard errors are clustered at the patent-drug level.

2.5.2   Type of Innovation

As outlined in Section 2.2, different types of incremental innovation activities should respond differently to the marketing authorization. While I expect innovation activities related to enforcement to decrease due to the obstacles to patent enforceability after drug approval, this

should not be the case for improvement innovations. Thus, I split originator forward citations by patent and disease types to distinguish between those likely related to enforcement and those likely related to improvements.

Patent Type

First, I show that the drop in incremental follow-on innovation is related to secondary patents that are more likely to be used solely to prevent generic entry.

Figure 2.6a illustrates that the drop in self-citations differs substantially by patent type. The negative effect is most pronounced in self-references to secondary patents. The effect size is of similar magnitude compared to the total effect discussed in Section 2.5.1. Moreover, I find a decrease in references to process patents, which is considerably smaller. This is consistent with the idea that process patents are more related to the production of the approved drug and less so for patent fencing purposes. In contrast, references to product patents or patents related to more complex biotechnology do not decrease after the focal drug received its first marketing authorization. Thus, incentives for improvement innovation efforts are not negatively affected by the approval.

At the same time, I show in Figure B-8 in the Appendix that the decrease in self-citations is concentrated among examiner citations, who are more likely to include conflicting prior art into the patent documents. Applicant citations, which are set more strategically, do not significantly decrease after the drug approval. Similarly, the effect results both from XY-citations and No-XY citations. This suggests that at least parts of the "missing" self-citations are considering the focal invention as novelty threatening prior art. Wagner and Wakeman (2016) argue that X and Y references "are indicators of legally "weak" patents as they [...] increase the likelihood of a post-grant validity challenge." This is a common characteristic of enforcement innovations, where, *e.g.*, the resulting secondary patents face regular legal challenges and are annulled regularly (European Commission, 2009).

These findings support the hypothesis that the marketing authorization only impacts innovation activities related to enforcement.

Disease Type

I further differentiate between the field of application of the incremental follow-on innovation activities to assess how closely the incremental innovation is related to the field of the primary innovation. This is related to the fact that a marketing authorization is tied to specific indications approved. To this end, I differentiate between self-citations within the same ICD-9 disease category and self-citations in other ICD-9 disease categories.

Figure 2.7 shows that the decrease in self-citations is entirely driven by a reduction of patents applied within the same ICD-9 disease area. The magnitude of the effect resembles the overall effect size. Self-citations related to patents in different ICD-9 disease areas, which

Figure 2.6: Impact of marketing authorization by type of patent
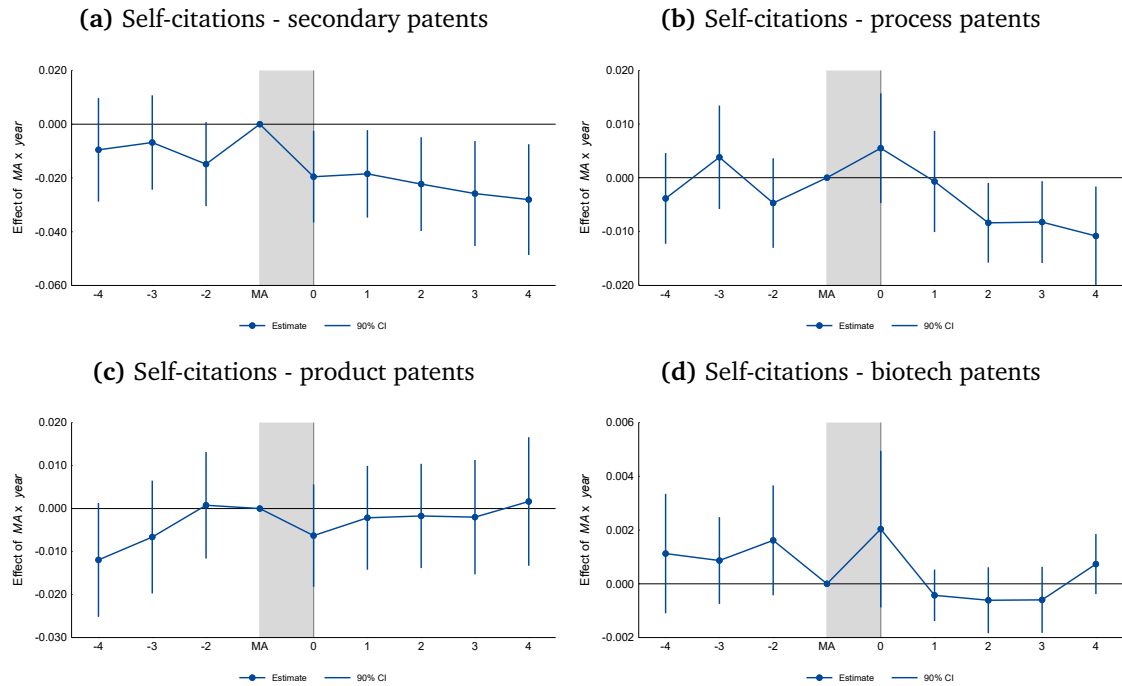
(a) Self-citations - secondary patents

(b) Self-citations - process patents

(c) Self-citations - product patents

(d) Self-citations - biotech patents

Notes: The figures show the event study estimates and the 90 percent confidence bands of linear regressions with high-dimensional fixed effects. The outcome variable is the log-transformed annual self-citation count generated by (a) secondary patents, (b) process patents, (c) product patents, and (d) biotechnology patents. The unit of observation is the unique patent-drug level. Standard errors are clustered at the patent-drug level.

are likely to cover more meaningful inventions in unrelated markets, do not decrease after the approval of the focal drug.

These findings support the notion that the marketing authorization mainly affects inventive activities marginal to the focal invention, and thus more likely to be conducted to prevent generic entry.

Enforceability

If the effect was caused by an increase in patentability standards that threaten the validity of patents, it should not be observable in earlier stages of the development process. Therefore, I investigate the effect of prior milestones in drug development on forward citations, which should not change the enforceability of follow-on patents. To this end, I identify the year of successful phase II clinical trials completion, when only limited novelty threatening information is disclosed, and use these event years similarly as in Equation 2.2.

Figure 2.8 shows the absence of any negative effect on self-citations after earlier events. This also holds true when looking at enforcement innovations (proxied by secondary patents) directly. This is consistent with the interpretation that the decreases in forward citations observed before are related to the proposed enforcement mechanism around the drug approval.

Institutional features surrounding the marketing authorization successfully remove some incentives of originators to invest in enforcement innovation activities.[31]

Figure 2.7: Impact of marketing authorization by type of disease
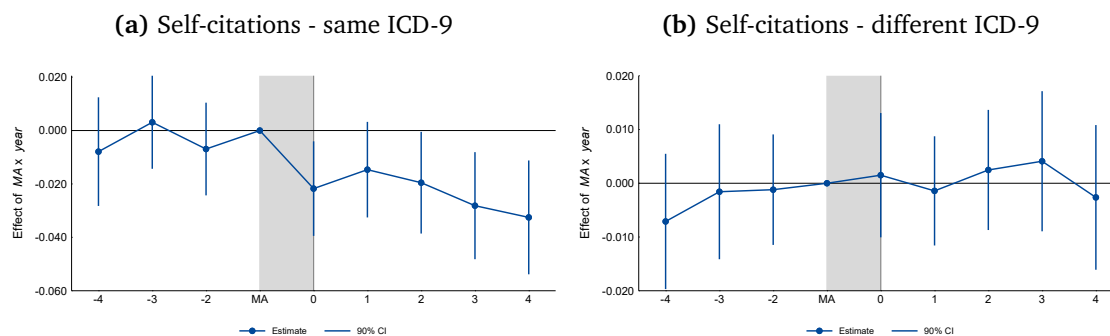
(a) Self-citations - same ICD-9 | **(b)** Self-citations - different ICD-9

Notes: The figures show the event study estimates and the 90 percent confidence bands of linear regressions with high-dimensional fixed effects. The outcome variable is the log-transformed annual self-citation count generated (a) in the same narrow ICD-9 category and (b) in a different ICD-9 category as the primary patent. The unit of observation is the unique patent-drug level. Standard errors are clustered at the patent-drug level.

Figure 2.8: Impact of ending phase II/beginning phase III trials on self-citations
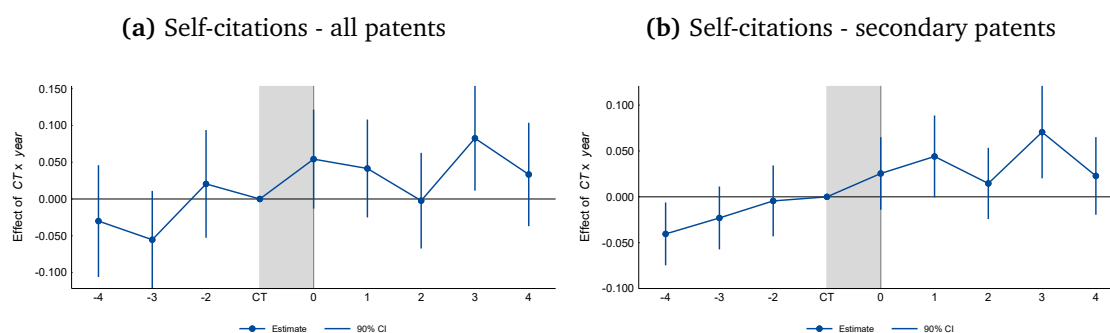
(a) Self-citations - all patents | **(b)** Self-citations - secondary patents

Notes: The figures show the event study estimates and the 90 percent confidence bands of linear regressions with high-dimensional fixed effects. The event is the end of phase II/beginning of phase III clinical trials. The outcome variable is the log-transformed annual self-citation count from (a) all patents and (b) secondary patents. The unit of observation is the unique patent-drug level. Standard errors are clustered at the patent-drug level.

2.5.3 Other Parties

This section is designated to explore the effect of the marketing authorization on other parties. In the pharmaceutical industry, vertical and horizontal relationships are common practice (Higgins and Rodriguez, 2006; Scherer, 2010). Thus, the originator is likely to have several collaborators both in the development of the NME and in the execution of clinical trials. Incremental follow-on innovation activities surrounding the focal drug should therefore not only be

---

[31] I use the beginning year of phase III clinical trials as proxy for the success of phase II clinical trials. This is possible because of the sequential order in the drug development process. I quantify these results in Table B-13 in the Appendix.

conducted by the originators but also by vertically related parties.[32] At the same time, generic companies file for a substantial amount of secondary patents themselves to protect their own exclusivity space once they are allowed to enter the market (Howard, 2007).

First, I show the results of all third parties together. The development of forward citations by other parties follows a similar pattern as for the originator. Figures 2.9a and 2.9b show that the negative effect of other parties' citations is concentrated among secondary patents. There is no decrease in incremental innovation activities resulting in product patents. The same applies for the distinction into "same ICD-9 category" versus "different ICD-9 category" patent citations illustrated in B-7 in the Appendix.[33] Across all variables, the effect happens later and is less precise. The latter hints towards substantial heterogeneity within other parties. The findings are consistent with the idea that the originator is affected most prominently by the marketing authorization, while the effect attenuates towards other firms.

Second, I provide suggestive evidence that this decrease is related to vertically related companies. Unfortunately, I cannot directly distinguish between a patent filed by a vertically related firm and a generic company because many brand-manufactures are also engaged in the generic product market, *e.g.,* one of the largest generic manufacturers *Sandoz AG* is part of *Novartis AG*. However, I observe whether a third party in my sample is engaged in clinical trials related to the focal drug. Since this is more likely to be the case for vertically related companies than for generic manufactures, I interpret this link as vertical relation between the originator and the third party. Figures 2.9c and 2.9d show that the decrease in enforcement innovation activities is related to patents filed by these presumably vertically related firms. Prior literature shows that secondary patent filings of generic firms are happening shortly before the expiry – year 20 to 25 from priority filing – of the primary patent (*e.g.,* Howard, 2007). Consequently, my findings are consistent since these periods are outside of the effect window in my estimation for most inventions.

The findings suggest that due to the vertical relationships in the pharmaceutical development process, shocks to one entity may diffuse to all related firms.

2.5.4 Summary Findings

Table 2.3 summarizes and quantifies all findings.[34] I observe decreases in the number of self-citations driven by secondary patents and within the same ICD-9 disease category. Taken together, this evidence supports the hypothesis that marketing authorization has an impact on the investment activities into incremental innovation related to enforcement. Innovations that are more likely to be related to actual improvements remain unaffected by the drug approval. The information disclosed (published results of clinical trials) in the course of the marketing

---

[32]Moreover, if the focal patent applicant is a smaller pharmaceutical company specialized in development activities but not in commercialization, it might license or sell the focal patent to large pharmaceutical corporations. It goes beyond of the scope of this study to analyse patent transfer.

[33]In Appendix Figure B-7, I illustrate the event study results on other outcome variables related to other parties.

[34]The Appendix quantifies the results on only self-citations in Tables B-8, B-9, and B-10.

Figure 2.9: Impact of marketing authorization on other parties' forward citations
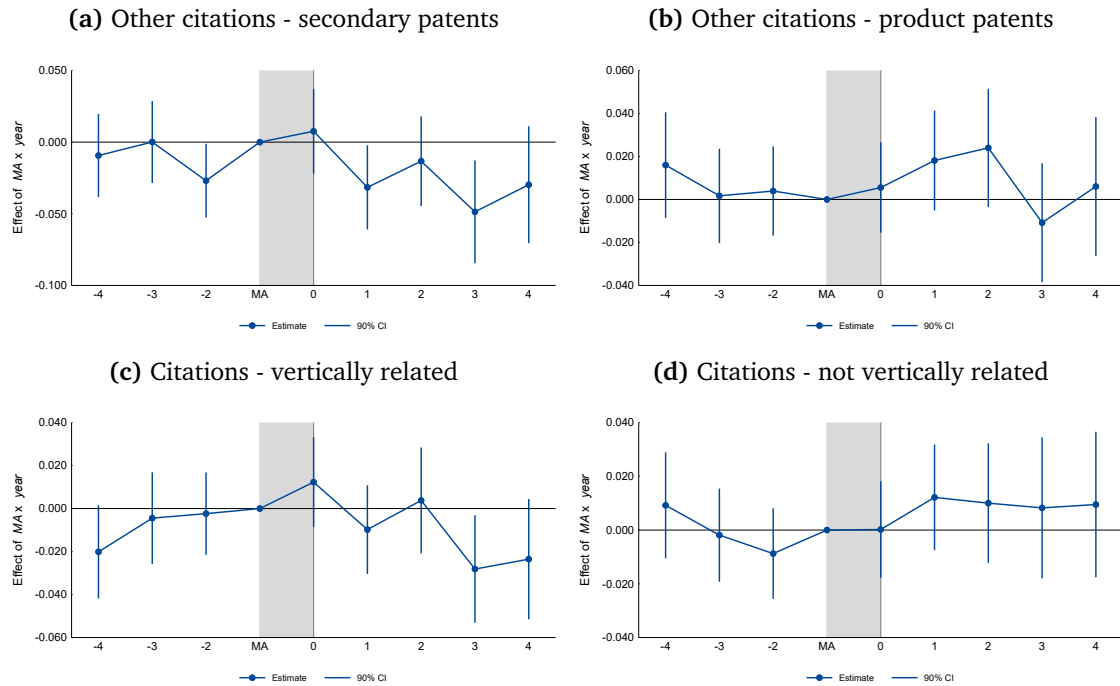
(a) Other citations - secondary patents

(b) Other citations - product patents

(c) Citations - vertically related

(d) Citations - not vertically related

Notes: The figures show the event study estimates and the 90 percent confidence bands of linear regressions with high-dimensional fixed effects. In the top, the outcome variable is the log-transformed annual other parties' citation count generated by (a) secondary patents and (b) product patents. In the bottom, the outcome variable is the log-transformed annual forward citation count by (c) presumable vertically related firms and (d) presumable vertically unrelated firms. The unit of observation is the unique patent-drug level. Standard errors are clustered at the patent-drug level.

authorization increases patentability standards (novelty threatening prior art/obstacles to non-obviousness) and, thus, impedes the enforceability of follow-on patents. Consistent with this proposed enforcement mechanism, I find pharmaceutical firms to decrease their investments in enforcement innovation activities, which have the sole purpose to limit or delay generic entry.

2.6 Alternative Explanations

In this section, I rule out alternative explanations that might explain the decrease in forward citations due to drug approval, such as competitive entry, accelerated approval, or unobserved heterogeneity.

Market Exclusivity and Incentives for Competitive Entry

First, I investigate whether the effect of drug approval is driven by changes in the likelihood of competitive entry. Due to the primary patent's fixed exclusivity term (20 years), it is inherent to the empirical setup that an earlier marketing authorization leads to a longer market exclusivity

Table 2.3: Impact of marketing authorization on forward citations

Log/Linear	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				DV: Log(1+Forward Citations)				
	All	Self	Other	Secondary	Process	Product	= ICD9	$\neq$ ICD9
n years before MA	−0.006	0.006	−0.012	−0.024	−0.018	0.014	−0.026	0.017
	(0.038)	(0.021)	(0.037)	(0.026)	(0.013)	(0.020)	(0.025)	(0.022)
4 years before MA	−0.023	−0.015	−0.010	−0.019	−0.022*	0.003	−0.016	−0.011
	(0.031)	(0.020)	(0.028)	(0.020)	(0.011)	(0.017)	(0.021)	(0.015)
3 years before MA	−0.034	−0.005	−0.035	−0.009	−0.008	−0.005	−0.014	0.008
	(0.028)	(0.018)	(0.027)	(0.020)	(0.011)	(0.015)	(0.018)	(0.015)
2 years before MA	0.003	−0.009	0.009	−0.040**	−0.018*	0.006	−0.010	−0.013
	(0.026)	(0.017)	(0.024)	(0.018)	(0.009)	(0.014)	(0.018)	(0.014)
1 year before MA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Year of MA	−0.023	−0.027*	−0.001	−0.015	0.002	0.003	−0.022	−0.001
	(0.026)	(0.016)	(0.025)	(0.019)	(0.012)	(0.014)	(0.017)	(0.014)
1 year after MA	−0.017	−0.029*	0.006	−0.049***	0.012	0.017	−0.024	0.001
	(0.028)	(0.016)	(0.026)	(0.019)	(0.013)	(0.015)	(0.019)	(0.014)
2 years after MA	−0.024	−0.036**	0.001	−0.036*	−0.010	0.022	−0.014	0.002
	(0.032)	(0.017)	(0.030)	(0.020)	(0.014)	(0.018)	(0.022)	(0.016)
3 years after MA	−0.084**	−0.040**	−0.050	−0.075***	−0.015	−0.011	−0.073***	0.003
	(0.037)	(0.019)	(0.035)	(0.022)	(0.016)	(0.018)	(0.024)	(0.018)
4 years after MA	−0.074*	−0.046**	−0.037	−0.058**	−0.015	0.006	−0.068**	−0.002
	(0.040)	(0.021)	(0.039)	(0.025)	(0.017)	(0.020)	(0.027)	(0.020)
n years after MA	−0.120**	−0.053**	−0.076	−0.090***	−0.019	−0.009	−0.083**	−0.027
	(0.049)	(0.025)	(0.047)	(0.028)	(0.019)	(0.024)	(0.032)	(0.023)
Patent Grant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SPC Grant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Patent-Drug FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Citation Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12390	12390	12390	12390	12390	12390	12390	12390
Cluster	590	590	590	590	590	590	590	590
Log-likelihood	−9340	−1712	−8885	−3982	1515	−1443	−4500	−1257

Notes: Columns (1) to (8) show the estimates of linear regressions with high-dimensional fixed effects. The dependent variable is the annual number of log-transformed forward citations in Column (1), self-citations in Column (2), other parties' citations in Column (3), forward citations generated by secondary patents in Column (4), by process patents in Column (5), by product patents in Column (6), within the same ICD-9 category in Column (7), and in a different ICD-9 disease category in Column (8). The unit of observation is the unique patent-drug level by year. Standard errors are clustered at the patent-drug level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

time (Budish et al., 2015). Prior literature finds that incentives for competitive entry are positively correlated with a longer market exclusivity period. It is associated with higher rates of subsequent entry (Gilchrist, 2016), paragraph IV challenges (Branstetter et al., 2016),[35] and non-infringing imitation activities (Izhak et al., 2020). This is because the longer the delay in generic entry, the higher is the increase in potential competitors' duopoly profits. If earlier drug approval caused more competition, it would increase the incentives to protect the focal

---

[35]If generic manufacturers seek to enter patent-protected markets, they can file an abbreviated new drug application with the FDA (Paragraph IV challenge) either by claiming non-infringement or invalidity of the branded product's patent (Branstetter et al., 2016).

invention by a higher number of secondary patents. Thus, my estimates would underestimate the total effect of drug approval on incremental innovation activities.

I exploit the discontinuity in market exclusivity *extensions* provided by the SPC system in the EU. As illustrated in Figure B-1 in the Appendix, it grants up to five years of patent-like compensation for long approval lags.[36] All focal inventions that are authorized between 5 to 10 years from priority receive a uniform total market exclusivity period of 15 years. For patents with a drug approval after 10 years, there is a negative linear relationship. Thus, for a subset of the drugs in my sample, the market exclusivity is independent of the time to approval and incentives for competitive entry can be considered as constant.

Figure 2.10 shows that the results are robust to this restriction and some point estimates become slightly larger. This supports the notion that the true effect is likely of a larger magnitude than the estimated effect.[37]

Figure 2.10: Constant market exclusivity period
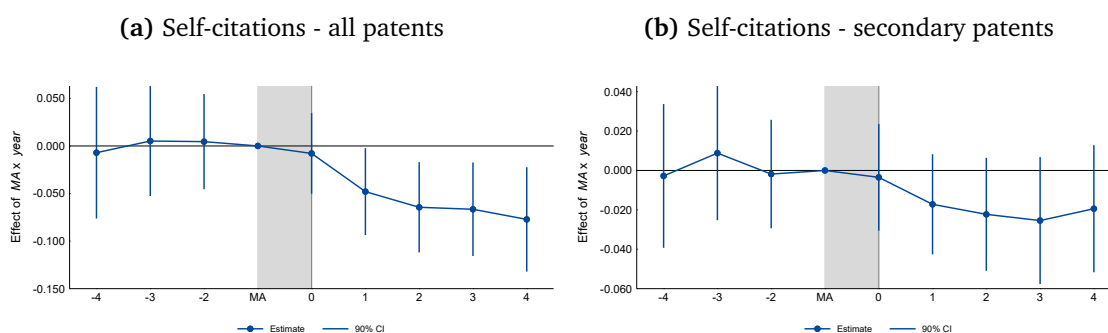
(a) Self-citations - all patents   **(b)** Self-citations - secondary patents

Notes: The figures show the event study estimates and the 90 percent confidence bands of linear regressions with high-dimensional fixed effects. The sample comprises only those 288 patent-drugs links that are associated with a marketing authorization within 10 years. This early MA group has uniform market exclusivity period of 15 years. The outcome variable is the log-transformed annual self-citation count from (a) all patents and (b) secondary patents. The unit of observation is the unique patent-drug level. Standard errors are clustered at the patent-drug level.

Accelerated Approval

Although the time to approval is not correlated with observable (value-related) characteristics, there is still room for concerns regarding unobserved heterogeneity. Dranove and Meltzer (1994) find that the importance of a drug is related to faster development times. If more valuable drugs were approved earlier, this would introduce an upward bias into my event study estimates.

---

[36]The SPC regime aims at compensating for the reduction in the effective patent life of pharmaceuticals caused by the delay when obtaining a marketing authorization. The SPC term is calculated by the difference between the date of the first marketing authorization in the EU and the filing date of the basic patent, less 5 years, and subject to a cap of 5 years.

[37]Further outcome variables can be found in the Appendix in Figure B-10.

Thus, I employ an instrument variables strategy following Gilchrist (2016). He reasons that the time from patent filing to the *beginning of first clinical trials* is uncorrelated with a drug's value. Among other reasons, this is because the science of drug development is highly uncertain and involves substantial spurious elements while the patent system forces firms to file primary patents immediately after a new promising molecule was discovered.[38] I follow his approach and create a cross-section of my data set using the total number of (self)-citations as the outcome variable.[39] A short pre-approval period (early MA) is related to fewer self-citations. I also replicate the main results of the previous analysis using the cross-section in Tables B-3 and B-5 in the Appendix.

Table B-6 shows the results of the IV approach, in which I use the *time to the beginning of clinical trials* as an *instrumental variable* for the *time to approval*. The sample size decreases substantially for reasons of data availability and estimates become noisier. In the second stage, showing the impact of *time to approval* on self citations, the IV estimates increase in magnitude compared to the OLS results.[40] These results suggest that my event study results are more conservative and present the lower bound of the effect.

Field of Application

Lastly, I investigate whether the effect is driven by unobserved differences in focal disease categories. Abud et al. (2015) show that secondary patents are more relevant for some therapeutic classes such as anti-depressants. These differences in importance can have dynamic patterns not controlled for using patent-drug fixed effects. Thus, I estimate leave-one-out regressions for each of the 17 ICD-9 disease categories mentioned in Table 2.2. The results presented in Figure 2.11 exemplify that the negative effect of the marketing authorization on self-citations is not driven by different compositions of disease categories among early and late approved drugs. This rules out the idea that there are different incentives for enforcement innovations across drug types.

---

[38]Gilchrist (2016) provides two further reasons why the time between patent filing and the start of clinical trials is uncorrelated with the invention: agency problems within firms and larger organizational frictions such as merger or acquisition activity may delay development processes.

[39]If the drug approval had an impact on related incremental innovation activities, this should shift the slope of forward citations downward and hence decrease the total citation count. In order to more closely mirror the event study in my analysis, I restrict the sample to forward citations accumulated between year 5 and year 16. This reflects the time period of the first MA (all forward citations after approval) and the last MA (all forward citations before approval).

[40]The instrument marginally passes the Stock et al. (2005) critical values for weak identification tests. The endogeneity-test shows that IV estimates are significantly different from OLS estimates when using the time to the beginning of phase I trials as instrument. In a larger sample using the time to the beginning of phase III trials as an instrument instead, the estimates point to a similar direction.

Figure 2.11: Leave one ICD-9 group out
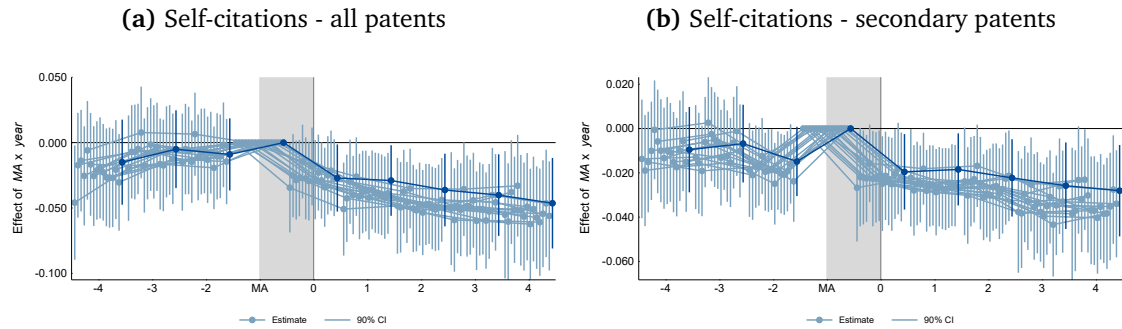
(a) Self-citations - all patents   **(b)** Self-citations - secondary patents

Notes: The Figures show the event study estimates and the 90 percent confidence bands of linear regressions with high-dimensional fixed effects. In each regression, I leave out one of the 17 ICD-9 categories. The outcome variable is the log-transformed annual self-citation count from (a) all patents and (b) secondary patents. The unit of observation is the unique patent-drug level. Standard errors are clustered at the patent-drug level.

2.7 Conclusion

Enforcement innovations are a common feature of modern drug development (Sternitzke, 2010, 2013). However, these innovations often have little or no therapeutic benefit (Frakes and Wasserman, 2020; Gurgula, 2020) so that the gain in consumer surplus is likely to be smaller than the loss resulting from longer market exclusivity periods and higher prices (Yin, 2017). In this study, I show that the incentives for enforcement innovation activities decrease in the course of a drug's marketing authorization. I posit this to be due to obstacles in the enforceability of follow-on patents filed after drug approval. Incremental innovation activities that are more likely to result in therapeutic benefits remain unaffected by the drug approval. This relates to the ongoing debate in the legal and economic literature on how to design more efficient incentives for the development of new drugs. Some scholars argue in favor of strengthening data and market exclusivity mechanisms, which are both unrelated to patents but directly provided by the regulatory health authorities (see *e.g.*, Gaessler and Wagner, 2020). Others suggest "fixing" the existing patenting system, *e.g.*, by strengthening the examination process (Frakes and Wasserman, 2020).

The results of my study show that pharmaceutical companies are self-adjusting their innovation behavior as a reaction to institutional features that potentially impede the enforceability of patents, like information disclosure at the time of drug approval. Thus, strengthening the patenting system might reinforce this type of reaction. For this purpose, Frakes and Wasserman (2020) recommend empowering patent examiners, *e.g.*, by providing them with more examination time. My result suggests to complement this by enhancing the information basis on which examiners assess novelty and non-obviousness. Not all drug approval-related information, which is submitted to the regulatory health authorities in the course of the marketing authorization, is publicly available. Therefore, granting patent examiners access to all documents submitted to the EMA or national health authorities may allow them to properly assess

the degree of novelty involved in secondary patent filings. If this increased patentability standards further, my results suggest that the self-enforcement of pharmaceutical companies will advance.

While this study provides a first explanation, it is of great importance to further disentangle the multiple channels through which the marketing authorization is having an impact on incremental innovation activities in future research. Given that the discussion about information disclosure (*e.g.*, clinical trial results) and the resulting obstacles to patentability in Europe is ongoing in the legal profession, as well, this calls upon collaboration between law and economics in order to fully understand both empirically as well as legally how the patent system can be strengthened. In this study, I considered follow-on patenting as a proxy for innovation. It is left for future research to differentiate between strategic patenting and enforcement innovation activities per se. Moreover, it is important to re-investigate these questions from a welfare perspective. The loss of market exclusivity may discourage pharmaceutical firms to invest in R&D (Budish et al., 2015; Higgins et al., 2020). The possibility, however, to file patents as a result of enforcement innovations may contribute to the development of totally new drugs by guaranteeing the necessary ex-ante incentives and cash flows. These considerations may help policymakers to design more efficient and effective incentives in the pharmaceutical industry that foster innovation activities *with* therapeutic benefits.

<div align="right">

3

</div>

Market Size and Research

Evidence from the Pharmaceutical Industry

3.1 Introduction

What factors drive innovation? This question has actively engaged scholars dating back to Schumpeter (1939). While the importance of this question is obvious for firms and policymakers, the answer is neither singular nor simple. Among others, Mowery and Rosenberg (1979) suggested that both supply-side ("technology push") and demand-side ("demand pull") factors may impact the innovation process. In general, technology push relies on accumulated knowledge from research and development (R&D) activities to drive the introduction of new products. In contrast, demand pull relies on demand characteristics (*e.g.*, population, disposable income, and preferences) to shape the pattern of investments in innovation (Kyle, 2020).

Focusing on demand-side factors, early work by Griliches (1957) and Schmookler (1966) recognized the importance of profit incentives and market size as drivers of innovation. This recognition has been carried forward into more recent work. Two seminal papers have established a causal link between market size and pharmaceutical innovation (Acemoglu and Linn, 2004; Finkelstein, 2004). Since then, a steady stream of empirical studies has examined how demand affects pharmaceutical innovation (*e.g.*, Agarwal and Gaule, 2021; Blume-Kohout and Sood, 2013; Clemens and Olsen, 2021; Dranove et al., 2020; Dubois et al., 2015; Kyle and McGahan, 2012; Lichtenberg and Waldfogel, 2009).[1]

These studies, however, focus on the 'D' as opposed to the 'R' in pharmaceutical R&D. They consider molecules entering (pre-)clinical trials, new drug approvals, or various other defini-

---

[1]A few studies exist which confirm these results for innovation responses following market size shocks outside the pharmaceutical industry (*e.g.*, Aghion et al., 2018).

tions that generally fall within the traditional rubric of 'development' as opposed to 'research'. Heretofore, efforts to establish a link between downstream demand and upstream research have resulted in limited success. Acemoglu and Linn (2004) and Finkelstein (2004), for example, were unable to find a relationship between patenting and demographic-driven expansions[2] or policy-induced expansions[3] in market size, respectively. In contrast, using a similar identification strategy as Acemoglu and Linn (2004), Bhattacharya and Packalen (2011, also referred to as B&P2011) identify some positive relationship between disease prevalence and upstream research, but cannot disentangle the increase in societal importance from pure profit incentives.

While an important first step, the Bhattacharya and Packalen (2011) link is not unexpected since their finding was driven by research conducted at academic medical centers.[4] This is precisely where one would expect clinical studies or 'development' to occur and the corresponding results to be published. Hence, *ex ante* there should be a positive correlation between research conducted at academic medical centers ('research') and clinical trials ('development'). Likewise, while Arora et al. (2018) show a general decline in overall corporate research, there should still be a positive correlation between publications by pharmaceutical firms and their clinical trial activities. Hence, the broader link, if it exists, between downstream demand and upstream scientific research remains elusive.

In this paper, we fill this gap and use the introduction of Medicare Part D to examine the effects of market size on science. For this purpose, we use novel data on scientific publications, patent-paper links, and drug development efforts mapped to disease categories. Moreover, while we use similar measures of disease exposure to Medicare Part D (the Medicare market share) as the extant literature,[5] the use of scientific publications necessitates a categorization of research at the disease level, as opposed to the therapeutic class level.[6] To demonstrate that our data construction choice does not bias the results, we start by replicating the main findings of the prior literature on drug development (*e.g.*, Blume-Kohout and Sood, 2013; Dranove et al., 2020).

We make several important contributions to the existing literature. First, over the decade following the implementation of Medicare Part D, we find no evidence for a causal relationship between market size and research. An increase of one standard deviation in the exposure to Medicare Part D leads only to an overall insignificant increase in scientific publications by

---

[2]They provide a variety of reasons for this result. First, they highlight the imperfection of their patent match. Second, they describe attenuation issues resulting from the delay in the research process. Third, they point towards companies being more responsive to profit incentives in later development stages.

[3]In her paper, Finkelstein argues that "[..] the quick initial response in development suggests the existence of a substantial reservoir of technologically feasible products 'on the shelf'. The decision to begin clinical trials is responsive, on the margin, to increases in the expected economic return to the clinical trial."

[4]Academic medical centers are hospitals that are linked to medical schools and engage in clinical trials.

[5]See, for instance, Blume-Kohout and Sood (2013), Dranove et al. (2020), Duggan and Scott Morton (2010), Hermosilla and Wu (2018), and Krieger et al. (2018).

[6]In the course of the analysis, we account for demographic changes, public research funding, and new research opportunities.

6.9%. This is substantially smaller than any effect on drug 'development' activities found in the prior literature and in our replication (+20.2%). These findings support Finkelstein's (2004) assertion that the link between market size and increases in drug development appears to be driven by a reordering of products already 'on the shelf'. This is also consistent with Dranove et al. (2020) who show that the upsurge in development appears to be driven by clinical trials of less scientifically novel drugs.

Second, there is extensive literature on scientist motivations. For example, scientists are motivated by external funding or rewards (Cohen et al., 2020; Foray and Lissoni, 2010; Hvide and Jones, 2018; Thursby and Thursby, 2011), altruism and prestige (Stern, 2004), recognition by the scientific community (Stephan, 2012), research opportunities and academic freedom (Aghion et al., 2008), public funding (Azoulay et al., 2019; Myers, 2020), and the desire to work on topics useful for society (Merton, 1973). It also appears that some scientists are motivated by monetary incentives (Levin and Stephan, 1991; Stephan, 1996), making them engage in patenting (Lach and Schankerman, 2008; Owen-Smith and Powell, 2001; Thursby et al., 2001). Our findings suggest that these upstream motivations have no direct link to the opportunities created by shifts in downstream demand in the form of market size changes.

Third, given the importance of academic medical centers, Bhattacharya and Packalen (2011) demonstrate that the *type* of research affiliation may matter. Various affiliations have different product market orientations. For example, on one end of the spectrum, the objectives of corporate scientists will be aligned with their firms, while on the other end of the spectrum, scientists at the National Institutes of Health (NIH) may be more interested in basic science research. To explore this variation, we accurately categorize research activities by four different types of affiliation: the NIH, universities, academic medical centers, and corporations. Consistent with our core findings, we illustrate substantial differences in semi-elasticities. Any demand response is concentrated only among corporate research (+22.7% in year 10 after the Part-D introduction), and decreases in effect size by distance to the market (*e.g.*, universities show an insignificant demand response of only +5.8%).

Fourth, we further refine our analysis by focusing not just on the type of affiliation but also the type of research (*e.g.*, applied or basic). With this refinement, we find that Medicare Part D primarily caused an increase in corporate affiliated publications linked to both clinical trials and pharmaceutical products, which are residuals from drug development activities (*i.e.*, applied research). The increase disappears for corporate affiliated publications that are more basic, which is broadly consistent with Arora et al. (2018). Consistently, we do not find any causal relationship between the type of research and market size for universities or academic medical centers. The divergence between our results and those of Bhattacharya and Packalen (2011) suggests that the differences in the *types* of downstream demand matter. In their case, research appears to respond to disease prevalence, while in our case, it does not respond to changes in disposable income within those diseases.

Fifth, not all publications are equal so we generate three different measures of impact. Our first measure are publications weighted by journal impact factors. Second, we map publications to patents to approximate whether scientific research was recognized in commercially relevant applications (Marx and Fuegi, 2020).[7] Finally, we weight the number of publications by the patent family size associated with the publication. Overall, results remain robust with our core findings – changes in downstream demand have no impact on upstream research.

However, there is one exception. In the years directly following the Part D-enactment, a one standard deviation increase in the exposure to Medicare Part D caused an *initial* 15.8 percent increase in corporate-affiliated patent-weighted research. This is consistent with the idea that corporate publishing is used strategically in commercialization activities such as patenting (Della Malva and Hussinger, 2012). Finally, we conclude with a series of robustness tests that redefine the dependent variable, use alternative calculations of Medicare market size, alternative controls, alternative specifications, alternative event windows, and different aggregation levels. In all cases, our core results hold.

Finally, our work has important implications for firms and policymakers. The pharmaceutical industry is highly dependent upon the external market for technologies (Higgins and Rodriguez, 2006), with much of that research emanating from universities (Cockburn and Henderson, 2000). While drug development (*i.e.*, drugs in clinical trials) appears to respond to downstream shifts in market demand, our results suggest that upstream research fails to do so. Firms face the prospect that the flow of research may not meet the kinds of development needs they may require. This problem is even more significant given the slow decline in corporate-level basic science research (Arora et al., 2018). This disconnect suggests that a more active role for policymakers may be needed. For example, in their recent work analyzing the innovation response to COVID-19, Agarwal and Gaule (2021) argue that policymakers may want to complement the market expansion with early-stage research incentives.

The remainder of the paper is organized as follows. In the next section, we describe the empirical setting related to the introduction of Medicare Part D and the empirical strategy. Subsequently, we introduce the data and selected descriptive statistics. We then present our main findings and conclude with a discussion.

3.2 Medicare Modernization Act

In the United States, Medicare is the national health insurance program for the elderly. Prior to 2006, it only covered drugs administered during in-patient hospital stays or at doctor offices, but it did not cover out-patient prescription drugs. In December 2003, U.S. Congress passed the Medicare Prescription Drug, Improvement, and Modernization Act (MMA), which

---

[7]Prior empirical research has shown that the relationship between scientific publications and patents is especially strong in technology fields that are related to the medical sector like chemistry or molecular biology (*e.g.*, Ahmadpoor and Jones, 2017; Poege et al., 2019; Watzinger and Schnitzer, 2019).

implemented the Medicare Part D prescription drug benefit as of January 2006.[8] This coverage is available for U.S. residents with age 65 and older who fulfill the eligibility criteria of Medicare Part A and B.[9] In contrast to other Medicare programs, Part D contracts with private companies that are authorized to sell insurance coverage. However, Medicare Part D is both regulated and subsidized, especially for low-income individuals.[10]

Medicare Part D covers all drugs that are also covered under Medicaid, which is a federal program that assists with medical costs for people with limited income, and that fulfill the following criteria: First, the drug has been approved by the Food and Drug Administration (FDA). Second, it must be available only by prescription. Third, the drug is medically necessary for on-label indications (*e.g.*, this limits off-label usage). Finally, Medicare Part D also covers biological drugs, insulin, smoking cessation drugs, and vaccines. While insurance plans do not have to cover all drugs mentioned above, there are certain "protected classes" for which most drugs are required to be included (*e.g.*, anti-cancer, anti-convulsant, anti-depressants, anti-psychotic, immuno-suppressant, HIV and AIDS drugs). The program excludes, for example, drugs that may be covered under Medicare Part A or B and over-the-counter drugs.[11]

The implementation of Medicare Part D was one of the most significant recent changes in the U.S. healthcare system. It was projected to benefit 29 million people in 2006 and 44 million people in 2015. The expected total public expenditures in the first 10 years were estimated to be $800 billion.[12] This expenditure corresponded to around 0.42% of GDP in 2006, increasing to 0.76% in 2015.[13] The program can be categorized as a demand subsidy.

As expected, Medicare Part D considerably increased the prescription drug use by elderly patients (Duggan and Scott Morton, 2010, 2011). In Figure 3.1 we illustrate this development. At the extensive margin, drug use by Medicare-insured patients increased substantially after 2006, especially in the quartile of diseases most likely to afflict older patients. This implies that previously uninsured elderly are now able to purchase prescription drugs. The same applies when looking at prescriptions for both existing and newly insured patients (Appendix

---

[8]The Medicare Prescription Drug, Improvement, and Modernization Act was introduced by Representative Dennis J. Hastert on June 25, 2003. It was heavily discussed and accumulated 21 roll-call votes until the U.S. House of Representatives agreed on November 22, 2003. Given the close vote, we assume that anticipation of the MMA was not very likely. More details can be found here: https://www.congress.gov/bill/108th-congress/house-bill/1 [last accessed on March 8, 2021].

[9]In fact, the eligibility criteria are broader. Besides the elderly, Medicare Part D is available for U.S. residents who receive Social Security disability payments for at least 2 years or have been diagnosed with end-stage renal disease or kidney failure. See: https://www.healthline.com/health/medicare/medicare-part-d-eligibility#choosing-a-plan [last accessed on March 8, 2021].

[10]See: https://www.medicareadvocacy.org/medicare-info/medicare-part-d/ [last accessed on March 8, 2021].

[11]It further excludes drugs for weight loss or gain, cough and cold preparations, fertility, erectile dysfunction, cosmetic and hair growth, as well as vitamins and minerals. Blume-Kohout and Sood (2013) exploit this set of rules to distinguish between diseases that profit more (*e.g.*, a high number of Medicare-Medicaid dual-eligible beneficiaries and a high number of protected drug classes) and less (*e.g.*, a high number of drugs previously covered by Medicare Part B) from the introduction of Medicare Part D. For more information, see: https://www.medicareadvocacy.org/medicare-info/medicare-part-d/ [last accessed on March 8, 2021].

[12]Own calculations based on the 2006 Medicare Trustees Report using an annual inflation rate of 5%.

[13]https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/ReportsTrustFunds/downloads/tr2006.pdf [last accessed on March 8, 2021].

Figure C-4). Thus, we can see that the MMA differentially increased the market size for drugs that are developed to treat diseases more prevalent among older individuals. This will be a fundamental aspect of our empirical strategy, which we outline in the next section.

Figure 3.1: Evolution of Medicare beneficiaries
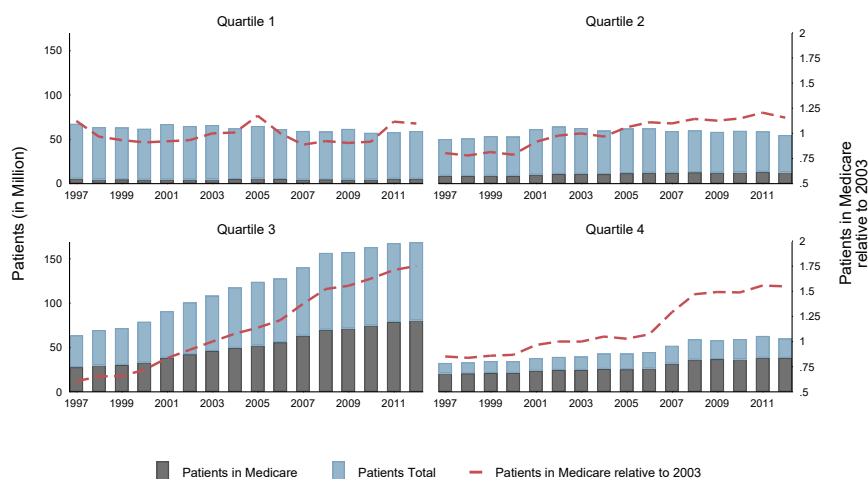
Notes: The figures show the evolution of patient counts of each ICD-9 group aggregated by Medicare market share quartiles. These are patients who received a prescription drug designated to a ICD-9 group disease at least once in a given year. Our exposure measure to Medicare Part D is described in Section 3.3.3. The grey bars display the number of Medicare patients, the blue bars display the non-Medicare patients. Patients are counted multiple times if they appear in more than one ICD-9 group. MMS quartiles are based on the pre-2004 weighted average of patient-based MMS. The red line represents the relative increase in the number of Medicare patients with respect to the baseline year 2003. The figures depict a discrete increase in Medicare patients after 2006 in the highest age quartile.

3.3 Data and Methodology

We are interested in the causal effect of Medicare Part D (*i.e.*, changes in downstream market demand) on upstream research. If upstream research responded similarly positively to market demand pull effects as downstream drug development, we would expect to find an increase in the number of scientific publications, all else equal. Further, we would also expect to see an increase in patent-paper links, which can be viewed as an output of basic science research.

3.3.1 Sample Selection

To create a link between Medicare Part D and R&D activities, we combine data on biomedical scientific publications from the NIH's MEDLINE/PubMed database and Web of Science (WoS), patent information from Patstat, and drug development activities from the Clarivate Analytics' Cortellis Investigational Drugs database (commonly known as Cortellis) database by ICD-9

disease categories and their exposure to Medicare Part D based on the Medical Expenditure Panel Survey (MEPS).

However, matching publications to disease categories requires considerable expertise, because the keywords in biomedical publication databases do not correspond 1-to-1 to standard international disease classifications. We take advantage of an existing crosswalk introduced by Bhattacharya and Packalen (2011), which we update and present in Table C-1. This crosswalk provides a mapping of Medical Subject Headings (MESH) terms with a range of ICD-9 three-digit codes.[14] Since some MeSH terms relate to multiple ICD-9 three-digit codes and vice versa, MeSH and ICD-9 three-digit codes are grouped at the level of mutually exclusive ICD-9 groups that comprise similar diseases.[15] The crosswalk is not exhaustive: it is restricted to ICD-9 categories at the three-digit level, which do not include the words "other" or "unspecified" since those categories typically include various very distinct diseases.[16]

Our final sample includes 129 separate matches at the ICD-9 group level,[17] which corresponds to 272 unique ICD-9 three-digit codes and 192 high-level MeSH terms. We take advantage of the stringent MeSH hierarchy to extend the initial set of MeSH terms by all synonyms and lower-level terms. Eventually, our sample comprises 1,563 MeSH terms. The selected ICD-9 groups constitute the basis for all other independent variables and drug development-related dependent variables.

3.3.2 Empirical Strategy

Figure 3.2 presents our empirical strategy through a case study. We select two disease categories from our panel, one with a very high and one with a very low exposure to Medicare Part D measured by the Medicare market share (MMS): *Alzheimer's disease* and *inflammatory skin diseases,* such as Acne Vulgaris and Seborrhoeic Dermatitis. We show that the trend in the number of scientific publications and number of preclinical trials (drug discoveries) related to either Alzheimer's or inflammatory skin diseases stay parallel before the passage of Medicare Part D in 2003. While the low-MMS inflammatory skin diseases seem to be uncorrelated with

---

[14]MeSH is a hierarchical medical vocabulary administered by the NLM and consists of approximately 30.000 different terms in 2020 (https://meshb.nlm.nih.gov/search) [last accessed on March 8, 2021]. The International Statistical Classification of Diseases and Related Health Problems (ICD) is administered by the World Health Organization. The 9th version (ICD-9) comprises around 13.000 codes (https://www.cdc.gov/nchs/icd/icd9cm.htm) [last accessed on March 8, 2021].

[15]There exists no official publicly administrated crosswalk between ICD-9 and MeSH terms. NIH resources like the UMLS Metathesaurus (https://www.ncbi.nlm.nih.gov/books/NBK9684/ [last accessed on March 8, 2021]) only include selected 1-to-1 matches. The usage of the PubMed search algorithm, which searches for terms in titles and abstracts, is infeasible because it requires the user to search for all possible synonyms. Furthermore, identifying appropriate MeSH terms for each ICD-9 category using the MeSH on demand algorithm (https://meshb.nlm.nih.gov/MeSHonDemand [last accessed on March 8, 2021]) requires an expert assessment in case of multiple results.

[16]The match is further limited to ICD-9 codes that appear more than 100 times in the MEPS data from 2007. It excludes the following ICD-9 categories: pregnancy (class 11), congenital (class 14), perinatal (class 15), symptoms (class 16), injuries (class 17), and services (class V).

[17]In addition to the 127 disease categories in B&P2008, we include two major diseases that they excluded: HIV & Alzheimer. Our results are not sensitive concerning these categories.

the MMA in the science and technology sphere, the high-MMS Alzheimer's disease increases disproportionately from 20 to 60 drug candidates (*i.e.*, drug development). This trend, however, does not occur with publications where the trends remain parallel after 2003.

Figure 3.2: Trends in R&D activities in Alzheimer's and inflammatory skin diseases

(a) Scientific publications $\qquad$ **(b)** Drug discoveries
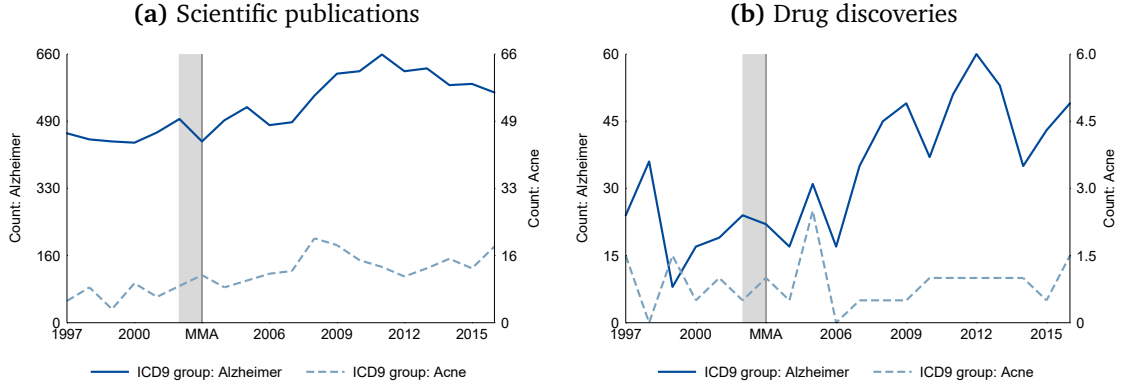
Notes: The left-hand Figure presents the number of publication counts, the right-hand Figure the number of newly discovered new molecular entities for the Alzheimer's disease (ICD-9 code: 331; MMS is 97%) versus inflammatory skin diseases (ICD-9 codes: 690/706; MMS is 4%) between 1997 and 2016.

In our multivariate analyses, we exploit the passage of the MMA by using the variation in exposure measured by the 1997-2003 MMS. To this end, we use a panel data model that observes each disease-related MeSH terms by year. The MeSH terms m are nested within ICD-9 disease groups $i(m)$ (defined as i in the following), which in turn constitute the level of treatment exposure. To capture the dynamics of the effect, we amend the standard Difference-in-Differences framework by replacing the post-period dummy with three-year binned sets of leads from 2004 onward. Hence, we compare research-related outcome variables before and after the introduction of Medicare Part D between more and less affected MeSH terms. The empirical model can be written as follows:

$$
\begin{aligned}
\mathbf{E}[N_{mt}|X_{it}] = \exp[\alpha + & \sum_{t=2004}^{2016} \beta^t \text{ Medicare Market Share}_i \times \mathbf{1}_{\{\delta_t = t\}} \\
& + \mu \left(\sum_{lead=1}^{\bar{f}} M_{i,t+lead} \right) + \lambda \left(\sum_{lag=1}^{\bar{l}} NIH_{i,t\text{-}lag} \right) + \gamma \left(\sum_{t=0}^{\bar{t}} K_{i,t} \right) \qquad (3.1) \\
& + \delta_t + \theta_i],
\end{aligned}
$$

where N_{mt} represents the dependent variable (*e.g.*, the number of publications per MeSH term in year t). The interaction terms $MMS_i \times \mathbf{1}_{\{\delta_t=t\}}$ indicate the exposure to the MMA and whether we are in the pre- or in the post-MMA periods (*i.e.*, $t \geq 2004$). Consistent with prior literature on market size and R&D, our empirical model includes controls for the future demography-driven market size M_{it}, past public research funding NIH_{it}, and research opportunities K_{it}.

Given the differences in the level of R&D activities across disease categories, we either include ICD-9 group fixed effects θ_i or MeSH term fixed effects θ_m. Furthermore, we control for time trends using calendar year effects δ_t. In alternative specifications, we interact the MMS with a full set of two-year binned leads and lags $\sum_{t=1997}^{2016} \beta^t MMS_i \times \mathbf{1}_{\{\delta_t=t\}}$.[18] In this setting, we normalize the coefficient $\beta^{t=2002/3}$ to zero and, hence, express the dynamic treatment effects relative to this pre-treatment year. This also allows us to examine whether the parallel trends assumption is likely fulfilled.

We estimate the relative change of R&D activity using Poisson pseudo-maximum likelihood regressions. Since our dependent variables are count data, Poisson is the preferred econometric model in panels (Hausman et al., 1984). We do not need to use a zero-inflated model, because all MeSH terms have at least one publication. Under our identification assumption, β^t gives us the average causal effect of the MMA in year t. The coefficients can be interpreted as semi-elasticities.[19] We cluster standard errors at the level of the treatment exposure, thus at the ICD-9 group level.

When re-investigating the effect of Medicare Part D on drug development, we turn to an analysis at the ICD-9 three-digit code level N_{ct}, which is the most fine-grained level of observation for our new molecular entities or clinical trials. ICD-9 three-digit code c are again nested in ICD-9 groups $i(c)$. In these specifications, we include either ICD-9 group fixed effects θ_i or ICD-9 code fixed effects θ_c, while all other parts of Equation 3.1 remain the same.

Lastly, we investigate heterogeneous effects by splitting the dependent variable along several categories, such as affiliations, journal types, clinical relevance, and funding sources. We use this battery of dependent variables to explore treatment effects along the entire scientific and innovation value chain.

3.3.3 Medicare Market Share

We exploit variation across disease categories in their pre-Part D MMS, expecting larger increases in scientific research in disease categories with higher MMS. Following previous studies on Medicare Part D (*e.g.*, Blume-Kohout and Sood, 2013; Dranove et al., 2020; Duggan and Scott Morton, 2010; Hermosilla and Wu, 2018; Krieger et al., 2018), we build a measure of a disease category's exposure to Part D based on MEPS.[20] Unlike Blume-Kohout and Sood (2013) and others, we use ICD-9 disease categories as our level of observations and not therapeutic classes because scientific articles are typically indexed by keywords corresponding to diseases. In MEPS, each patient-level drug prescription is associated with a designated disease, an indicator of whether the beneficiary was insured by Medicare, and the patient's age. Using this data, we calculate the MMS for each disease at the ICD-9 three-digit level and the more ag-

---

[18]Our post-treatment period overlaps with the passage/implementation of the Affordable Care Act (ACA) in 2010. However, we do not consider this as a concern because the ACA involved only low reimbursement for pharmaceuticals and, thus, small revenue increases or incentives from market size (Garthwaite et al., 2020).

[19]A one percentage point higher MMS_i leads to a change of the dependent variable N_{mt} of $\beta^t * 100$ percent.

[20]The MEPS data is available here: https://www.meps.ahrq.gov/mepsweb [last accessed on March 8, 2021].

gregated ICD-9 group level. The latter corresponds with the level of our match between ICD-9 and MeSH terms, and, thus, with the level of our empirical analysis on biomedical science.

Researchers could anticipate the market size increase from the authorization of the MMA in December 2003 onward. Thus, we calculate the fraction of patient counts, prescription counts, and quantity weighted prescription counts filled by Medicare-covered individuals compared to all individuals for each disease category as a weighted-average between 1997 and 2003.[21] The categories with the highest MMS are *Alzheimer, Retinal Diseases, Cataract, and Angina Pectoris*. Among the diseases with the lowest MMS are *Hyperkinetic Syndrome of Childhood, Scarlet Fever, Infantile Cerebral Palsy, and Inflammatory Skin Diseases*. The distribution of our MMS at the ICD-9 group level is presented in Figure 3.3, has a mean of 32%, and is in line with other studies (*e.g.*, Dranove et al., 2020; Hermosilla and Wu, 2018). It does not vary with the ICD-9 level, sample selection, or the use of (quantity weighted) prescription counts (Appendix Figure C-3). We are confident in our sample of 129 ICD-9 disease groups as our level of analysis since we are able to replicate the main results of the prior literature (Duggan and Scott Morton, 2010).[22]

Figure 3.3: Distribution of Medicare market shares (1997-2003)
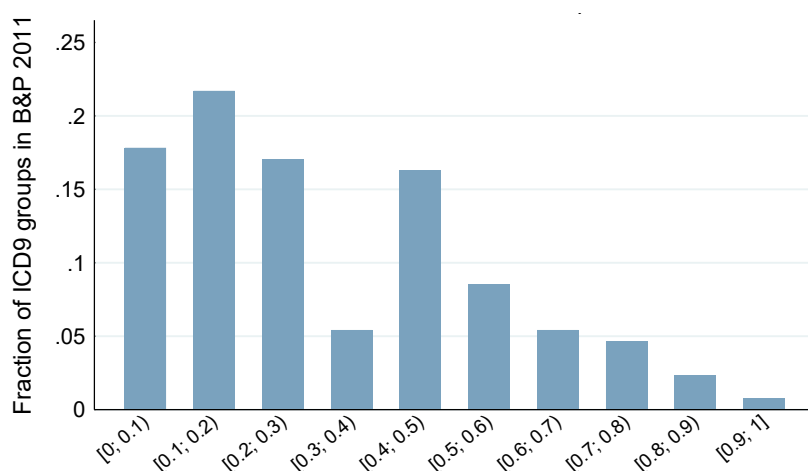
Notes: The figure presents the distribution of MMS scores among ICD-9 groups that are included in the MeSH-ICD-9 crosswalk (B&P2011 sample). We use the patient-weighted average of each year between 1997-2003. The annual MMS are calculated using the total number of patients in Medicare relative to all patients for each ICD-9 group.

---

[21]We weight each survey respondent in MEPS by her representativeness (person-level sampling weight).

[22]Our disease level data allows us to replicate the development of quantities, drug prices, and revenues. We document the positive effect of Medicare Part D on drug consumption (Appendix Figure C-4). Moreover, in line with Duggan and Scott Morton (2010) and Duggan and Scott Morton (2011) drug prices decrease between 2006 and 2009 since patients were able to switch to cheaper insurance plans. However, prices increased after 2009 (holding the 2003-2005 drug basket for each disease constant). Despite the initial price declines, by 2006 revenues increased disproportionately for high MMS diseases. This suggests that the quantity increase outweighs the initial price decline (Appendix Figure C-5).

3.3.4 Dependent Variables

Scientific Publications

NIH's MEDLINE/Pubmed database includes the entire universe of references to journal articles in the biomedical sciences from the early 20th century to the present. We retrieve all publications (also referred to as PMIDs) with at least one of the 1,563 MeSH terms linked to the 129 ICD-9 groups. We restrict our sample to U.S. publications. Next, we match these publications with bibliographic data from Web of Science (WoS) to take advantage of WoS's proper author name and affiliation disambiguation. The WoS data enables us to look at various splits in the publication data, such as the affiliation type (*i.e.*, NIH, university, corporate, academic medical centers),[23] and appliedness of the journal.[24] Moreover, the WoS bibliographic information allows us to add information regarding forward citations and journal impact factors.

The coverage of PMIDs in WoS is high, which gives confidence in capturing all relevant papers related to medical science. Moreover, we extract for each paper all indexed MeSH terms from the Pubmed database to classify whether a publication is related to disease terms that are not in our sample. Since we do not know the exposure of these additional disease-related MeSH to the MMA, we treat them cautiously as potentially confounding and drop them from the sample. Beyond that, we use all indexed MeSH terms to classify whether a publication is related to clinical trials or pharmaceutical products.[25,26]

Next, we measure the extent of the research efforts related to a disease group by counting the number of matched scientific publications at the MeSH term level. Publications might be associated with more than one ICD-9 group, so that we account for this in two alternative ways: we either treat them as simple counts for each disease group separately or weight them by the inverse number of linked diseases that sum up to one across all disease groups. The resulting final data set spans from 1997 to 2016 and includes 449,996 unique publications.

Patents

We use references in patents to the scientific non-patent literature (SNPL) to identify knowledge diffusion between upstream research and more downstream innovation activities, like

---

[23] We infer the affiliation type from the disambiguated Web of Science publication data based on the string name of each affiliation. Academic medical center are identified using the string "hospital". Corporate affiliations are identified using legal forms like "Corp." or "Inc.".

[24] We use a classification of journals based on the proportion of published research coming from a general hospital and industry using the publicly available data set provided by Tijssen (2010). See for more information: https://www.vosviewer.com/journal-application-domain-map [last accessed on March 8, 2021].

[25] The category "clinical trials" includes all MeSH terms that are related to the MeSH ID "D016430" (Clinical Trial) and the entire set of MeSH terms at the hierarchy levels below, such as "Adaptive", "Phase I", "Phase II", "Phase III","Phase IV","Controlled Clinical Trial" or "Randomized Controlled Trial".

[26] The category "pharmaceutical products" includes all MeSH terms that are related to the MeSH ID "D004364" (Pharmaceutical Preparations) and the entire set of MeSH terms at the hierarchy levels below, such as "Dosage Forms", "Drug Combinations", "Drugs, Generic", "Drugs, Investigational", "Pharmaceutic Aids" or "Prescription Drugs".

patenting (following Ahmadpoor and Jones, 2017; Marx and Fuegi, 2020; Poege et al., 2019; Watzinger and Schnitzer, 2019).[27] We locate publications that are directly cited in a patent by matching the Pubmed-patent link constructed by Marx and Fuegi (2020) to our sample of scientific publications.[28] Thus, we weight U.S. scientific publications by an indicator variable that specifies whether a publication was cited as a SNPL within a 5-year window.[29] This applies to 19,891 biomedical scientific publications.[30]

Clinical Drug Development

To measure the impact of Medicare Part D on drug development activities, we use time-series data from Cortellis on all clinical drug development events by disease categories at each stage in the pharmaceutical development process. We link the Cortellis *targeted conditions* to ICD-9 codes using the crosswalk by Dranove et al. (2020)[31] and identify unique new molecular entities (NME) entering Phase I, Phase II, or Phase III clinical testing, as well as being submitted to the FDA for approval. We limit the sample of NMEs to those, which are discovered/tested/approved in the U.S., and include information on the target-based action and at least one designated disease. If a clinical trial occurs more than once for a NME-disease link, we use the first event. To be consistent with our science level analysis, we count the number of NMEs at the more fine-grained ICD-9 three-digit code level, which is nested within our ICD-9 groups.

In total, we identify 9,943 NMEs entering at least one phase of the drug development process. This comprises 201 ICD-9 three-digit codes, and 121 ICD-9 groups (we lose 8 ICD-9 groups that had no NME between 1997 and 2016).

3.3.5 Control Variables

The empirical model includes a set of additional determinants of R&D: projected market size, public research funding, and research opportunities. A detailed description of the control variables construction can be found in the Appendix C.1. First, we control for the projected market size due to U.S. population growth (Acemoglu and Linn, 2004; Blume-Kohout and Sood, 2013). For this purpose, we use demographic (projection) data from the UN World Pop-

---

[27]Patents reference various types of documents that relate to the protected invention by either determining novelty (prior art) or explaining the content of the underlying invention. A subset of these references relates to scientific articles, called SNPL references (Poege et al., 2019).

[28]We use the patent-paper link provided by Marx and Fuegi (2020), which includes patents until 2019. Their link between Microsoft Academics Graphs (and PubMed/MEDLINE) and patents is based on front-page citations to non-patent prior art.

[29]We aggregate all citing patent applications at the DOCDB family level and calculate the 5-year window from the year of the scientific publication to the year of the priority year of the patent family.

[30]Our publication-patent link may suffer from attrition because late publications have not yet been cited in patents. However, we have little reason to expect that the time to patent varies systematically by MMS. This is supported by our finding that the minimum time lag between a scientific article's publication year and a patent's priority year is uncorrelated with the MMS (*i.e.*, the pairwise correlation is −0.0365).

[31]Dranove et al. (2020) had two expert medical coders independently code the concordance between Cortellis indications and ICD-9 codes. We are grateful to Manuel Hermosilla for sharing the mapping with us.

ulation Prospects. Figure 3.4a illustrates the average development of the population-growth driven projected market size before and after the introduction of Medicare Part D. While market size will increase in all disease categories, diseases in the highest quartile of MMS exhibit the sharpest growth due to an aging population.[32]

Second, we control for previous years' public research funding related to each disease category. Therefore, we calculate for each of our 129 ICD-9 groups the exposure to the NIH budget over time by linking each group to the relevant NIH Institute/Center (*e.g.*, ICD-9 162 *malignant neoplasm of trachea, bronchus, and lung* to the NCI *National Cancer Institute*).[33] Figure 3.4b presents the normalized average NIH spending by ICD-9 groups in each MMS quartile. It becomes apparent that diseases in the lowest quartile of MMS are associated with the largest relative increase of NIH funding. However, high MMS diseases receive a substantially higher level of NIH funding.[34]

Third, we account for the *availability of research opportunities*.[35] We develop a direct measure of new research opportunities in a disease area based on the introduction of new terms in the respective branch of the MeSH tree. New terms are added for emerging diseases, a more detailed definition of existing diseases, as well as additional terminology to reflect topical areas that are not well represented in MeSH. Figure 3.4c shows substantial heterogeneity in new research opportunities across ICD-9 groups. High MMS diseases exhibit greater increases in research opportunities around 2000.

3.4 Descriptive Analysis

Our final data set consists of 129 disease groups from 1997 to 2016. This allows us to investigate the possible effects of Medicare Part D on scientific research over a period of 13 post-MMA years. Each of the 129 ICD-9 groups in our sample is associated with, on average, 12.1 disease MeSH terms (1,563 observations). For 121 of these ICD-9 groups, we find drug development activities in around 1.7 ICD-9 three-digit codes per group (201 observations).

Table 3.1 provides summary statistics for the full set of independent and dependent variables in the year 2003. The average ICD-9 group has 204.3 scientific publications or 16.9 per MeSH term. The majority of publications (77.8%) have at least one author with a university affiliation, 13.6% have at least one author with an academic medical center affiliation, 5.4%

---

[32]We also calculate projected market size at the OECD level. Figure C-1 illustrates the evolution.

[33]Since grants are distributed within Institutes primarily by scientific merit (see discussion on NIH funding rules by Azoulay et al. (2019)) and not by allocation to narrower disease categories, we attribute the full Institute's budget to each ICD-9 group. We retrieve NIH spending data (Mechanism Detail by IC, FY 1983-2019) from https://officeofbudget.od.nih.gov/spending_hist.html [downloaded on February 17, 2020].

[34]In an alternative approach we attribute budgets based on the share of all publications in a disease category that acknowledge a specific Institute. Figure C-2 in the Appendix illustrates the evolution.

[35]In similar fashion, Bhattacharya and Packalen (2011) construct a measures of research opportunities based on the content of research inputs and the first appearance of the idea in a scientific publication. The disadvantage of the approach is that it relies on a very narrow set of research inputs that relate primarily to drug-related medical research but not basic science.

Figure 3.4: Control variables over time

(a) Projected market size
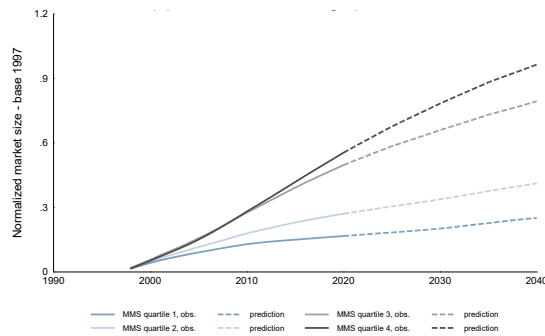
(b) Average NIH funding (normalized)
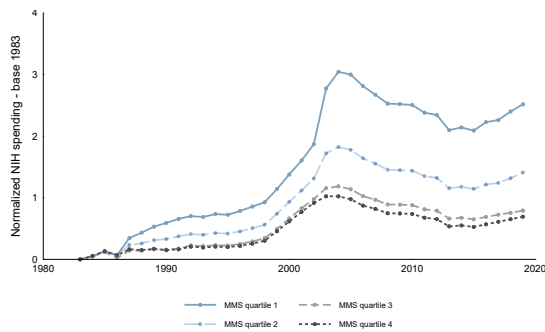
(c) New MeSH terms
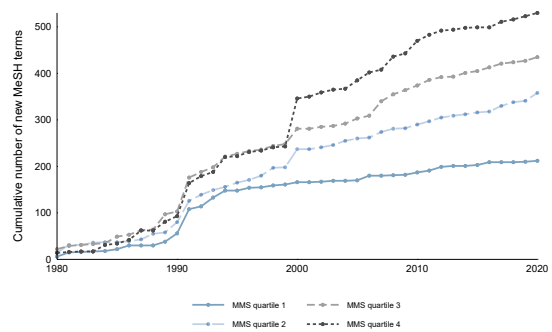
Notes: The three figures present the normalized annual control variables by MMS quartiles over time. In Figure 3.4a, we aggregate the U.S. population-growth driven market size (in 2003 values) of each ICD-9 group. In Figure 3.4b, we average the NIH spending (in 2003 values) of all Institutes/Centers, which are related to our ICD-9 groups. In Figure 3.4c, we accumulate the number of new MESH terms associated with our ICD-9 groups.

have a corporate affiliation, and 4.9% are published with NIH participation. Moreover, 6.2% of these publications are cited in patent applications. This share, however, is substantially larger for corporate publications (14.3%). At the same time, 2.03 NMEs enter pre-clinical trials per ICD-9 group. This number decreases throughout the clinical trial process resulting in 0.09 new drug approvals per ICD-9 group in 2003.

A simple comparison between ICD-9 groups split at the MMS median in Appendix Table C-2 illustrates that both groups are very similar regarding the pre-MMA levels in the majority of dependent variables. This also applies to the distribution of dependent variables (Appendix Figure C-6). An important exception is that diseases prevalent among the elderly are more related to clinically relevant journals and patents. This is not surprising since the overall market size is larger, in levels, before the MMA. Moreover, the MMS is positively correlated with the level of all independent variables. This supports our decision to control for these factors in our multivariate analysis.

Descriptively, the total yearly log-transformed number of scientific publications in low or high MMS ICD-9 disease groups develops in a parallel fashion until 2003. This provides visual

Table 3.1: Summary statistics

	N	Mean	Median	Std. Dev.	Min	Max
ICD9 group level						
MMS (cases)	129	31.81	27.32	23.22	0	97
MMS (prescription counts)	129	34.42	30.21	24.53	0	98
MMS (prescription quantity)	129	35.18	29.28	25.31	0	97
Cumul. US Market Size$_{t \text{ to } t+12}$	129	13575.14	2695.54	40242.85	29	343211
Cumul. NIH funding$_{t-1 \text{ to } -12}$	129	16.72	14.38	10.97	3	46
Cumul. New MeSH terms$_t$	129	0.23	0.00	0.70	0	4
MeSH term level						
Scientific publications	1563	16.86	2.00	69.33	0	1128
Publications - fractional	1563	13.11	1.00	58.95	0	1000
NIH publications	1563	0.82	0.00	4.14	0	70
University publications	1563	13.28	1.00	54.87	0	890
Academic Medical Center publications	1563	2.29	0.00	8.63	0	151
Corporate publications	1563	0.91	0.00	4.41	0	80
NIH grant publications	1563	6.67	0.00	32.87	0	565
Clinical trial university publications	1563	1.27	0.00	5.75	0	110
Clinical trial corporate publications	1563	0.21	0.00	1.16	0	25
Pharmaceutical university publications	1563	0.33	0.00	1.74	0	27
Pharmaceutical corporate publications	1563	0.07	0.00	0.46	0	9
Citation-weighted publications	1563	353.33	16.00	1646.59	0	29351
JIF-weighted publications	1563	60.88	4.28	266.30	0	4518
Patent-weighted publications	1563	1.05	0.00	6.26	0	125
Patent-weighted university publications	1563	0.83	0.00	4.99	0	102
Patent-weighted corporate publications	1563	0.13	0.00	0.75	0	14
Patent family size-weighted publications	1563	9.17	0.00	53.45	0	1082
ICD9 3-digit code level						
Drug discoveries	201	2.03	0.00	5.43	0	47
Phase I clinical trials	201	0.54	0.00	1.32	0	8
Phase II clinical trials	201	0.82	0.00	1.97	0	17
Phase III clinical trials	201	0.26	0.00	0.67	0	4
Drug approval	201	0.09	0.00	0.37	0	3

Notes: This table presents summary statistics linked to the 129 ICD-9 groups in 2003. The unit of observation is at the ICD-9 group level for both the treatment and control variables; at the MeSH term level for the dependent publication variables; and at the ICD-9 three-digit code level for dependent drug development variables.

support for our parallel trends assumption which is needed for our Difference-in-Differences framework (see Figure 3.5a). These trends also hold across other dependent variables, like corporate publications (see Figure 3.5b).[36] Formal tests for parallel trends, *e.g.*, by splitting the pre-period and testing whether there are differential changes in the slope, are employed and found supportive (Appendix Tables C-3 & C-7). After 2004, the log number of publications associated with high-MMS diseases increase disproportionately, but only to a small degree. This divergence is more pronounced for publications from corporate affiliates.

Parallel trends are also supported by the annual count of drug discoveries and approvals in Figures 3.5c and 3.5d, which evolve in a similar fashion for low- and high-MMS diseases

---

[36]Further univariate graphs with other dependent variables can be found in the Appendix in Figure C-7.

until 2003. The number of drug discoveries and drug approvals increases after 2003. As such, we can replicate the prior literature (Blume-Kohout and Sood, 2013; Dranove et al., 2020) descriptively within our sample of ICD-9 groups.

Figure 3.5: Trends in scientific publications and drug development by MMS

(a) All publications

(b) Corporate publications

(c) Drug discoveries

(d) Drug approvals
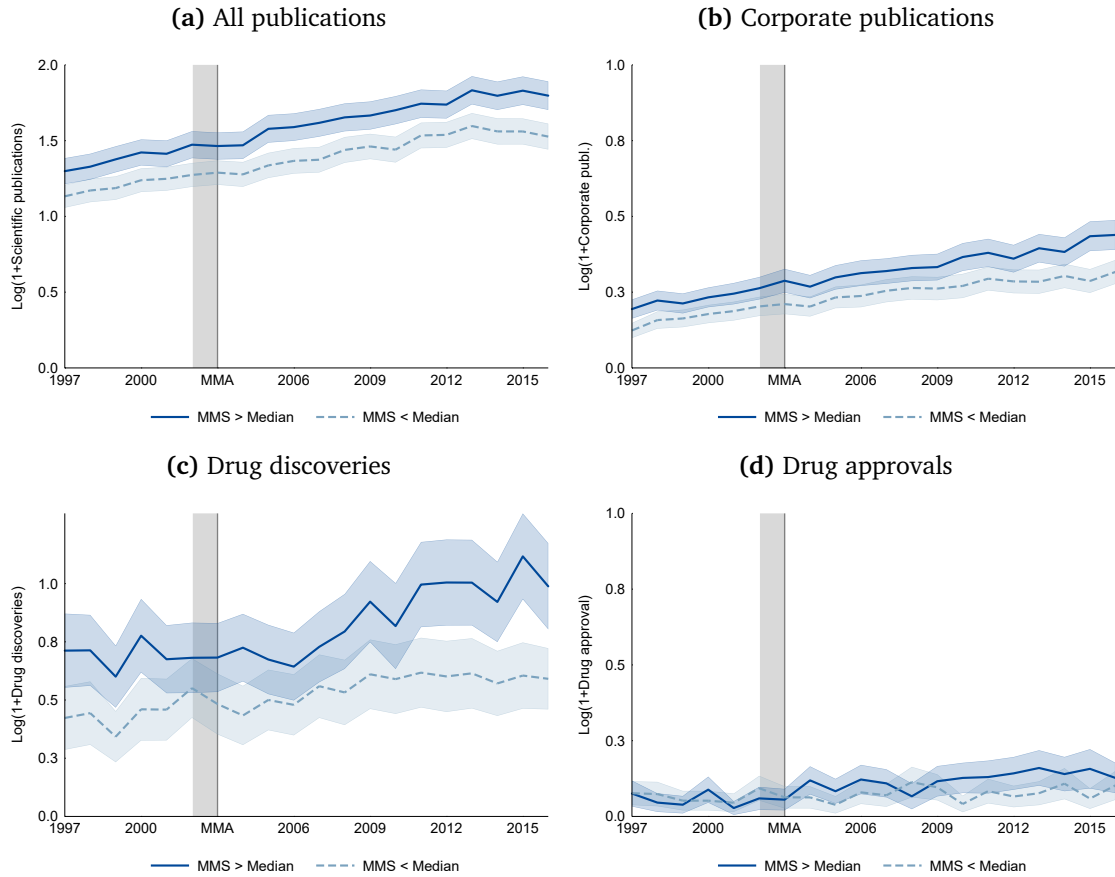
Notes: Figure (a) presents the log-transformed average number of annual publication counts associated with all affiliations and Figure (b) selects only publications from corporate affiliations. Figure (c) displays the log-transformed average number of annual drug discoveries and Figure (d) the log-transformed average number of annual drug approval. In all graphs, the unit of observation is the unique ICD-9 group level.

This descriptive analysis suggests that the introduction of Medicare Part D and, thus, the sudden insurance-induced increase of market size for diseases more prevalent among the elderly, has led to more commercial drug development activities. Upstream research activities, in general, seem to be more resilient to these changes in downstream market size. However, upstream research conducted by corporations appears more elastic in high MMS disease categories after the MMA.

The following multivariate analysis will investigate this pattern in more detail, accounting for other factors like demographic trends, public funding, and new research opportunities that may have an impact on R&D outcomes besides Medicare Part D.

3.5 Empirical Results – Clinical Drug Development

We start by replicating prior results showing the effect of the MMA on clinical drug development. This replication exercise provides validation for our sample selection and variable construction. Moreover, the results of this analysis will enable us to compare the effects among scientific research and drug development activities within our sample. Figure 3.6 shows the event study results similar to Equation 3.1. The dependent variables are the number of newly discovered NMEs and the number of drug approvals, respectively.

Figure 3.6: Event study – drug development
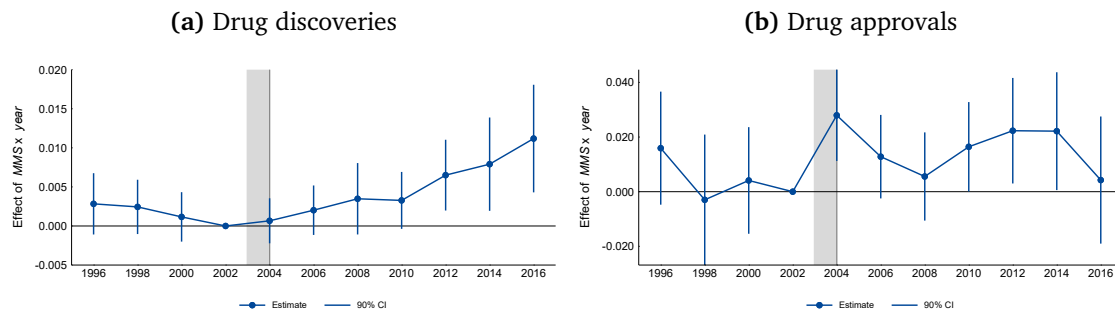
(a) Drug discoveries **(b)** Drug approvals

Notes: The figures show the event study estimates and the 90 percent confidence bands of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects following Equation 3.1. The unit of observation is the ICD-9 three-digit code level, with MMS being calculated based on patient counts at the ICD-9 group level. The sample includes all ICD-9 groups appearing in the B&P2011 crosswalk. Standard errors are clustered at the level of treatment (ICD-9 group level).

In line with our descriptive analysis, the number of new drug discoveries and the number of drug approvals display broadly similar patterns independent of the pre-MMA MMS. This suggests the absence of confounding pre-trends. After the passage of the MMA in 2003, disease categories with higher exposure to Medicare Part D exhibit a relative increase in drug discoveries. The effect becomes larger over time and is statistically significant. The same holds true for drug approvals. However, there is a significant up-tick directly after the introduction of Medicare Part D in 2004 and 2005.[37]

Next, Table 3.2 presents the estimates of our Difference-in-Differences estimation with dynamic post-periods. Columns 1 and 2 present the results on drug discoveries with and without control variables. In Columns 3 and 4 we present NMEs in all stages of the clinical development process. Finally, Columns 5 and 6 present clinical trials in Phase I-III and drug approvals separately. We find a positive and significant effect of a higher MMS on early drug development, accelerating over time. In our preferred specification, in which we control for the future market size, past NIH funding, and research opportunities, the effect becomes significant after 2011. This is consistent with the long discovery process in the pharmaceutical industry. The point estimate in 2015-2016 has a magnitude of 0.87%. This implies that one standard deviation (23.2 percentage points) increase in MMS leads to 20.2% more drug discoveries.

---

[37]We show the event study results of other drug related outcome variables in Figure C-8 in the Appendix.

Importantly, these effects are similar in magnitude compared to Dranove et al. (2020) and slightly smaller than the results of Blume-Kohout and Sood (2013).

These effects are smaller and occur later when looking at drug development in all stages of clinical development or NMEs in Phase I-III clinical trials. However, this is not surprising given the staggered fashion of drug development (from preclinical to clinical trials to approval) and supported by other studies on drug development (Blume-Kohout and Sood, 2013). An exception are drug approvals, which show a positive significant increase immediately after the introduction of the MMA in 2004-2005 by 2.5% per additional percentage point of MMS. These results strongly support Finkelstein (2004) and suggest that pharmaceutical companies reacted by pushing forward advanced drug candidates already in their development pipeline relevant for elderly patients.

Table 3.2: Drug development

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
			Drug Development			
	Early Development		All Development Stages		Phase 1-3	Approval
MMS × 2004-05	−0.0011	−0.0007	−0.0011	−0.0007	−0.0044	0.0256***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.009)
MMS × 2006-08	0.0006	0.0013	0.0000	0.0007	−0.0017	0.0070
	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.008)
MMS × 2009-11	0.0009	0.0020	0.0001	0.0012	−0.0004	0.0138
	(0.002)	(0.003)	(0.002)	(0.003)	(0.004)	(0.010)
MMS × 2012-14	0.0035	0.0053*	0.0024	0.0040	0.0016	0.0198*
	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)	(0.012)
MMS × 2015-16	0.0065*	0.0087**	0.0046	0.0066*	0.0043	0.0191
	(0.003)	(0.004)	(0.003)	(0.004)	(0.006)	(0.012)
Cumul. US Market Size$_{t \text{ to } t+12}$	No	Yes	No	Yes	Yes	Yes
Cumul. NIH funding$_{t-1 \text{ to } -12}$	No	Yes	No	Yes	Yes	Yes
Cumul. New MeSH ids$_t$	No	Yes	No	Yes	Yes	Yes
ICD9 group FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3800	3800	3880	3880	3760	3220
ICD9-codes	190	190	194	194	188	161
ICD9-groups	110	110	114	114	110	87
Log-likelihood	−6733	−6723	−10459	−10449	−6136	−1316

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the ICD-9 three-digit code by year. The dependent variable is the annual number of newly discovered NMEs in Columns (1) and (2), NMEs in all clinical development stages (*i.e.*, preclinical, clinical trials, registrations, approvals) in Columns (3) and (4), NMEs in Phase I-III clinical trials in Column (5), and approved NMEs in Column (6). The control variables are log-transformed. Standard errors are clustered at the level of treatment (ICD-9 group level) and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

3.6 Empirical Results – Biomedical Science

In this section, we pivot and present the results of our empirical strategy outlined in Section 3.3. We evaluate whether Medicare Part D differentially affected scientific publications for disease categories with higher MMS. Importantly, we differentiate by the type of affiliation as well as type of research and investigate the (commercial) impact.

3.6.1 Main Effect

First, we evaluate whether Medicare Part D differentially affected scientific research in MeSH categories associated with high MMS ICD-9 disease groups from all types of affiliations. Our event study results in Figure 3.7 are based on a Poisson pseudo-maximum likelihood regression with the full set of control variables, ICD-9 group and calendar year fixed effects (adapting Equation 3.1). Overall, we see no pre-MMA effect on scientific publications suggesting the absence of confounding pre-trends. After the passage of Medicare Part D, we observe only a slight divergence for the average MeSH category, which is neither significantly different from zero nor large in magnitude.

Figure 3.7: Event study – scientific publications
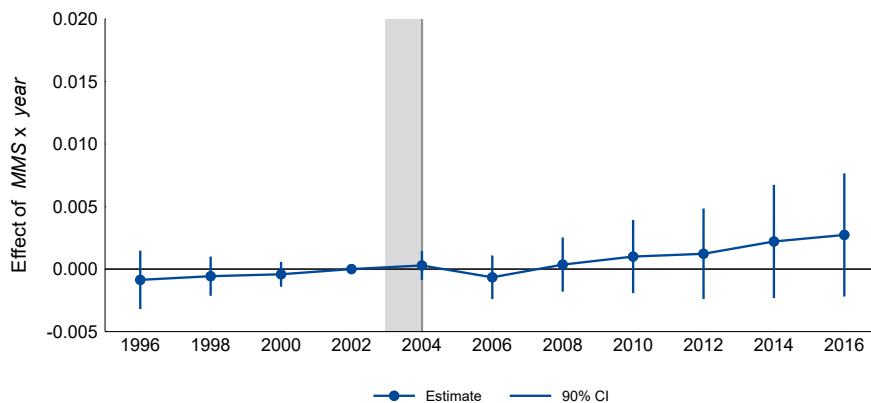
Notes: The figure shows the event study estimates and the 90 percent confidence bands of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects following Equation 3.1. The unit of observation is the MeSH term level, with MMS being calculated based on patient counts at the ICD-9 group level. Standard errors are clustered at the level of treatment (ICD-9 group level).

We quantify the changes in scientific research in response to the Part D demand shock in Table 3.3. In Columns 1 and 2, we report the results from a simple Difference-in-Differences regression. The post-treatment period is defined to start in 2004 and to last until the end of our sample's observation period in 2016. Again, there is no significant effect on scientific publications, independent of the usage of control variables.

In Columns 3 to 6, we estimate the dynamic changes in science. The first time period shows the transitional effect between the passage and the implementation of Medicare Part D. The

post-implementation effects are divided in three-year bins. Column 3 presents the effect of the MMA on MeSH categories related to high MMS diseases without controls. Results are close to zero and insignificant. Adding control variables in Columns 4 to 6 increases the coefficients but not the interpretation of the results. Our preferred specification (Equation 3.1) in Column 6 includes the full set of controls and serves as the baseline for the further analysis.

Under the assumption that there was no relationship between MMS and scientific activity prior to 2003, positive coefficients would indicate that the Part D demand shock led to an increased number of scientific publications in a given time period. This does not seem to be the case. Ten years from the passage of Medicare Part D, the point estimate can be interpreted as one additional percentage point in MMS resulting in 0.3% additional publications. This is considerably below the effect size on drug discoveries.[38] Taking the point estimate at face value, a MMS increase of one standard deviation (23.2 percentage points) leads to only 6.9% additional scientific publications. This can be considered fairly inelastic.

Table 3.3: Scientific publications

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
			DV: Scientific Publications			
MMS × post 2003	0.0003	0.0002				
	(0.001)	(0.001)				
MMS × 2004-05			0.0002	0.0006	0.0006	0.0006
			(0.001)	(0.001)	(0.001)	(0.001)
MMS × 2006-08			−0.0008	−0.0002	−0.0001	−0.0001
			(0.001)	(0.001)	(0.001)	(0.001)
MMS × 2009-11			0.0003	0.0013	0.0013	0.0013
			(0.001)	(0.002)	(0.002)	(0.002)
MMS × 2012-14			0.0004	0.0018	0.0017	0.0017
			(0.001)	(0.002)	(0.003)	(0.003)
MMS × 2015-16			0.0016	0.0032	0.0030	0.0030
			(0.002)	(0.003)	(0.003)	(0.003)
Cumul. US Market Size$_{t \text{ to } t+12}$	No	Yes	No	Yes	Yes	Yes
Cumul. NIH funding$_{t-1 \text{ to } -12}$	No	Yes	No	No	Yes	Yes
Cumul. New MeSH ids$_t$	No	Yes	No	No	No	Yes
ICD9 group FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	31260	31260	31260	31260	31260	31260
MeSH IDs	1563	1563	1563	1563	1563	1563
ICD-group	129	129	129	129	129	129
Log-likelihood	−894474	−894404	−894397	−894371	−894319	−894319

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. The dependent variable is the annual number of scientific publications. The control variables are log-transformed. Standard errors are clustered at the level of treatment (ICD-9 group level) and shown in parentheses. Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

---

[38]The 95th percentile confidence interval rules out an increase greater than 0.9%, which is approximately the effect of drug discoveries in Table 3.2.

3.6.2 Type of Affiliation

We expect the relationship between market size and scientific research to depend on product market orientation. Biomedical scientists with corporate affiliations have direct financial ties to the market for prescription drugs. Their objectives should be aligned with those of the pharmaceutical industry (Henderson and Cockburn, 1996). Moreover, corporate scientists participate in the publication of the results of clinical trials in scientific journals. The latter also applies to scientists and practitioners at academic medical centers, who play an intermediary role between industry and academia (Lander and Atkinson-Grosjean, 2011; Lander, 2013). For scientists at universities, the relationship is more complex since market orientation differs across sub-fields and depends on a variety of factors (elaborated in Foray and Lissoni, 2010), but should be overall less pronounced compared to corporate scientists.

Figure 3.8: Event study – scientific publications by affiliation type
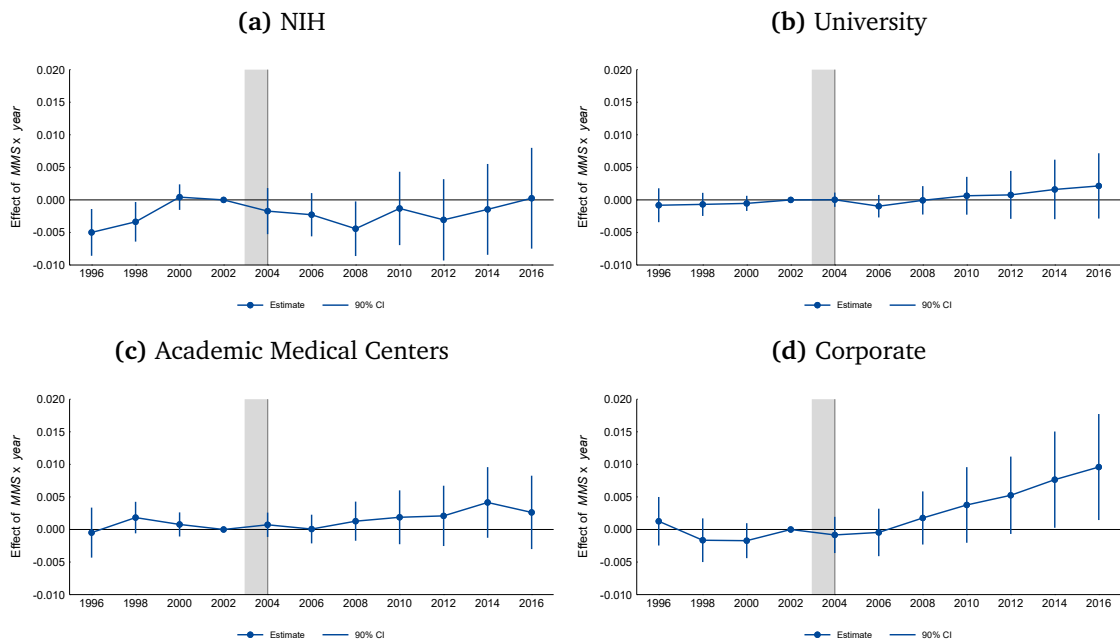
(a) NIH

(b) University

(c) Academic Medical Centers

(d) Corporate

Notes: The figures show the event study estimates and the 90 percent confidence bands of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects following Equation 3.1. The unit of observation is the MeSH term level. The dependent variable is the annual number of scientific publications split by at least one author's affiliation. Standard errors are clustered at the level of treatment (ICD-9 group level).

In the following analysis, we split our dependent variable by whether the publication was coauthored by at least one scientist who was affiliated with the NIH, a university, an academic medical center, or a corporation. Figure 3.8 shows the event study results, displaying the yearly excess publications in high MMS relative to low MMS MeSH categories. While generally not significant, the effects still go along with our predictions. It is least pronounced in the public research sphere at the NIH and most pronounced in the private research sphere at corporations. This is consistent with the idea that the market orientation of scientists matters.

Table 3.4 quantifies the Medicare Part D demand responses by affiliation subgroup. We find that the demand response following the introduction of Medicare Part D is concentrated among scientific publications with at least one corporate affiliation and increases gradually over time. The timing of the effect is consistent with the results of Blume-Kohout and Sood (2013), where the response magnifies through 2009 and appears to stabilize after 2012. In contrast, scientific publications with at least one university scientist have substantially smaller coefficients.[39] The demand response from academic medical centers sits plausibly in-between universities and industry.

The point estimates of our analysis suggest that scientific research at non-corporate affiliations is less responsive to changes in market size compared to drug development within the same sample of ICD-9 groups. For instance, a MMS increase of one standard deviation (23.2 percentage points) leads to only 5.8% (7.9%) of additional scientific publications from university (academic medical center) scientists. The effect on scientific research directly conducted at the NIH is essentially zero. However, the same market size expansion leads to an increase of scientific publications coming from industry by 22.7%. This resembles the magnitude of our findings on drug development from Section 3.5. In the following sections, we will investigate which type of research drives this effect.

As outlined in Section 3.2, private insurance plans, which fall under the scope of Medicare Part D, do not have to cover all approved drugs. However, there are certain "protected drug classes" for which most drugs are required to be included (*e.g.*, anti-cancer, anti-convulsant, anti-depressants, anti-psychotic, immuno-suppressant, HIV and AIDS drugs). Hence, we distinguish between ICD-9 groups that correspond to "unprotected" or "protected" drug classes. Our results in Figures C-12 in the Appendix correspond with our expectations and the previous literature (Blume-Kohout and Sood, 2013; Dranove et al., 2020). However, the only specification that contains statistically significant effects is the model focused on corporate affiliated scientists publishing in "protected" ICD-9 groups.

3.6.3 Type of Research

Given the rise of corporate science in response to Medicare Part D, we investigate the vertical orientation of these research publications. To this end, we differentiate between scientific publications that are related to the development of drugs (*i.e.*, more applied in nature) and more basic science. We add the full set of MeSH terms to each publications and identify those MeSH terms that are related to clinical trials and those that are related to pharmaceutical products. We interpret the residual as fairly basic research. Finally, we split the dependent variable by the appliedness of the journal.

---

[39]These findings are quantitatively similar but less precise when looking at publications from only corporate affiliations. This is consistent with the idea that pharmaceutical industry and university research are interlinked (Henderson and Cockburn, 1996).

Table 3.4: Scientific publications by affiliation type

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
			Scientific Publications			
	NIH	University	No University	AMC	Corporate	All US
MMS × 2004-05	−0.0006	0.0004	0.0007	0.0001	−0.0001	0.0004
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
MMS × 2006-08	−0.0021	−0.0005	−0.0002	−0.0002	0.0010	−0.0004
	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)
MMS × 2009-11	−0.0014	0.0010	0.0014	0.0011	0.0036	0.0009
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
MMS × 2012-14	−0.0018	0.0012	0.0020	0.0020	0.0059	0.0011
	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)	(0.002)
MMS × 2015-16	−0.0001	0.0025	0.0034	0.0030	0.0098*	0.0018
	(0.004)	(0.003)	(0.003)	(0.004)	(0.005)	(0.003)
Cumul. US Market Size$_{t \text{ to } t+12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t-1 \text{ to } -12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. New MeSH ids$_t$	Yes	Yes	Yes	Yes	Yes	Yes
ICD9 group FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	30480	31260	31260	31240	31120	31260
MeSH IDs	1524	1563	1563	1562	1556	1563
ICD-group	111	129	129	128	125	129
Log-likelihood	−55113	−742561	−584310	−156719	−68090	−669330

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. The dependent variable is the annual number of scientific publications split by at least one author's affiliation. In Column (1) at least one author is affiliated with the NIH, in Column (2) with a university, in Column (4) with an academic medical center, and in Column (5) with a firm. Column (3) includes publications that have at least one author not affiliated with a university. In Column (6), we count only publications, for which all authors have U.S. affiliations. The control variables are log-transformed. Standard errors are clustered at the level of treatment (ICD-9 group level) and shown in parentheses. Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

We show in Table 3.5 that the increases in corporate scientific publishing are more articulated among publications that relate to clinical trials and pharmaceutical products but less so in basic science. The latter is supported by the fact that there is no increase in publications, for which the scientists received a NIH grant, a good proxy for the basicness of research. There is also no demand response on clinical trial and pharmaceutical product-related scientific research at universities or academic medical centers (Appendix Table C-18).[40] These results are supported by our event study analyses (Appendix Figure C-13) indicating that the disproportional increases in clinical trials and pharmaceutical product publications for corporation are not driven by pre-existing trends but by the introduction of Medicare Part D. The magnitudes of the effects on corporate applied science are substantially larger compared to all (other) types of research.

These results are consistent with our findings on the appliedness of journals. The demand response is most pronounced for research published in clinical practice, industry practice, and

---

[40]If at all, there is evidence for crowding out in those areas with the strongest increase among corporate publications: pharmaceutical products-related publications.

Table 3.5: Scientific publications by type of research

All ICD9 Groups Count/PPML	(1) Basic	(2) University CT	(3) Pharma	(4) Basic	(5) Corporate CT	(6) Pharma
MMS × 2004-05	0.0008	−0.0010	−0.0044**	−0.0020	0.0045	0.0090**
	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.005)
MMS × 2006-08	−0.0002	−0.0015	−0.0056**	0.0001	0.0042	0.0060
	(0.001)	(0.002)	(0.003)	(0.002)	(0.004)	(0.004)
MMS × 2009-11	0.0013	−0.0006	−0.0056*	0.0008	0.0104*	0.0161***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.006)	(0.006)
MMS × 2012-14	0.0015	0.0010	−0.0052	0.0028	0.0125*	0.0209**
	(0.003)	(0.004)	(0.004)	(0.003)	(0.007)	(0.008)
MMS × 2015-16	0.0029	0.0018	−0.0059	0.0066	0.0162**	0.0236***
	(0.003)	(0.005)	(0.006)	(0.005)	(0.007)	(0.009)
Cumul. US Market Size$_{t\ to\ t+12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t-1\ to\ -12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. New MeSH terms$_t$	Yes	Yes	Yes	Yes	Yes	Yes
ICD9 group FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	31260	30960	30860	31100	30080	28240
MeSH terms	1563	1548	1543	1555	1504	1412
ICD-group	129	124	120	124	109	95
Log-likelihood	−659174	−85517	−32763	−50491	−20014	−7944

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. In Columns (1) to (3), the dependent variable is the number of university scientific publications and in Column (4)-(6) the number of corporate scientific publications, both split by the type of research. Columns (2) and (5) include only publications that are associated with MeSH terms related to clinical trials, and Columns (3) and (6) with MeSH terms related to pharmaceutical products. Column (1) and (4) include the residual. The control variables are log-transformed. Standard errors are clustered at the level of treatment (ICD-9 group level) and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

clinical-industrial journals driven by corporate research activity (Appendix Table C-16) and not by universities (Appendix Table C-15). We view our findings as evidence for the interpretation that a majority of corporate scientific publications, which result from the increase in market size, are related to actual drug development activities (*e.g.*, published clinical trial results) and do not constitute basic research. It supports the notion that demand pull effects are not strong enough to encourage true basic science – not even within industry.

3.6.4 Research Impact

In the last part of the analysis, we explore the impact of scientific research both within the scientific domain and beyond. As outlined in Section 3.3.4 we trace scientific publications to patents. These patent-paper linkages approximate whether scientific research got recognized in commercially relevant applications (Marx and Fuegi, 2020), in our context pharmaceutical and biomedical patents. Thus, we weight scientific publications by the journal impact factor

(which is less affected by truncation compared to forward citations), by the 5-year availability of patents, and by the patent family size.

Our impact-weighted results in Table 3.6 show a similar but more noisy pattern with smaller magnitudes than the overall unweighted publication counts. Within the science domain, there is a disproportional increase in corporate JIF-weighted publications (although not significant), which is substantially smaller for university publications. The magnitude of the corporate effect is around 50% of the effect on simple counts. This suggests that the effects on corporate science are more pronounced at the extensive (quantity) margin and less so when quality-weighted.

Patent-weighted publications increase primarily among corporate affiliations. The strongest effect is found initially after the passage of Medicare Part D. Corporate scientific research that ends up cited in a patent, increases by 0.68% in the years 2004-2005. This corresponds to an increase of 15.8% given a MMS increase of one standard deviation. The effect size decreases in the following years but reaches similar levels after 2011. The quick initial response suggests the existence of a reservoir of scientific research "on the shelf" available for commercialization. This is consistent with prior literature, which suggests that publishing corporate science is used strategically in the patenting process (Della Malva and Hussinger, 2012). An alternative explanation is that industry became more likely to patent ideas from scientific publications as a reaction to the discrete increase in market size.[41]

3.7 Robustness Checks

We conduct a variety of robustness checks, which can be found in the Appendix. The demand response of corporate scientific research to the increase in market size is robust across a variety of changes, unless otherwise stated. The same applies to the inelastic response of all publications, across all other types of affiliation.

First, we *redefine the dependent variable* as a count of the annual number of scientific publications weighted by the inverse number of linked diseases (fractional counts) to account for multiple disease MeSH terms per publication. We also winsorize the dependent variable to deal with outliers (Figure C-10, Table C-8 and C-12).

Second, we calculate *different exposure variables,* for example, MMS based on prescription counts/quantity, binary indicators, and MMS based on 2003 values only (Figures C-11).

Third, we use *alternative control variables* such as the OECD market size or NIH funding calculated as a share of all publications in a disease category that acknowledge a specific Institute. Moreover, we include control variables that do not accumulate future/past periods

---

[41]Family-size weighted publications show a similar pattern, except for an arbitrary decrease in 2015-2016. All results are more pronounced when looking at event studies (Appendix Figure C-14). Using alternative measures for research impact, such as forward citations or SNPL references to drug patents that occur in the Cortellis database in the same ICD-9 group as the scientific publication, point to similar results (Appendix Table C-17).

Table 3.6: Impact/patent-weighted scientific publications

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
	JIF-Weighted		Patent-Weighted		Family Size-Weighted	
	Uni	Corporate	Uni	Corporate	Uni	Corporate
MMS × 2004-05	−0.0001	−0.0002	0.0020	0.0068*	0.0001	0.0049
	(0.001)	(0.003)	(0.002)	(0.004)	(0.003)	(0.005)
MMS × 2006-08	−0.0015	−0.0011	0.0009	0.0016	0.0000	0.0045
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.004)
MMS × 2009-11	0.0004	0.0028	0.0011	0.0025	−0.0003	0.0044
	(0.002)	(0.004)	(0.003)	(0.004)	(0.003)	(0.005)
MMS × 2012-14	0.0006	0.0039	0.0008	0.0058	−0.0013	0.0078
	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)	(0.005)
MMS × 2015-16	0.0017	0.0058	0.0003	0.0062	−0.0019	−0.0013
	(0.003)	(0.005)	(0.004)	(0.006)	(0.004)	(0.007)
Cumul. US Market Size$_{t\ to\ t+12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t-1\ to\ -12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. New MeSH ids$_t$	Yes	Yes	Yes	Yes	Yes	Yes
ICD9 group FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	31260	31120	30700	27980	30700	27980
MeSH IDs	1563	1556	1535	1399	1535	1399
ICD-group	129	125	114	92	114	92
Log-likelihood	−2998960	−268192	−50082	−10335	−329216	−73054

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. In Columns (1) to (2), the dependent variable is the journal impact factor-weighted number of university/corporate scientific publications. In Columns (3) to (4), the dependent variable is the number of university/corporate scientific publications that are associated with at least one patent application (patent-weighted). In Columns (5) to (6), we weight the number of scientific publications by the size of the average patent family associated with the publication. A patent/family size-weight is calculated based on the patent family's first application being filed within five years from the scientific publication. The control variables are log-transformed. Standard errors are clustered at the level of treatment (ICD-9 group level) and shown in parentheses. Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

but only consider year t (Table C-9 for all publications, Table C-13 for university publications, and Table C-14 for corporate publications).

Fourth, we estimate our model using *different specifications, e.g.,* using MeSH term fixed effects and cluster standard errors at the MeSH term level. Additionally, we employ a linear regression with count dependent variables. In this specification, we find a significant increase in overall scientific publications, which suggests that MeSH terms with a high number of pre-MMA publications profit from a larger market in absolute counts but not relatively (Table C-9). The effect size is substantially higher for corporate publications (Table C-14) than for university publications (Table C-13) when comparing the coefficients to the pre-MMA sample mean. Fifth, we *restrict our dependent variable* to publications in which *all* authors are affiliated with universities or firms, respectively. While the effect sizes remain quantitatively similar, the estimation becomes less precise (Figures C-11) since we lose variation in the dependent variables. We also restrict the sample to publications in which all authors have a U.S. affiliation. This does not change the results (Table C-10).

Sixth, we re-investigate the effects including the year of the *MMA implementation (2005) into our baseline period*. This does not change the results regarding the type of affiliations or the type of research. The effects on patent-weighted publications disappear. This, however, is not surprising given that our initial findings showed primarily a disproportionate increase in 2004-2005 (Table C-19).

Seventh, we include *all* publications in our sample regardless of whether they include additional disease-related MeSH terms with unknown MMS. In our default specifications, we drop these PMIDs since we do not know the exposure of these additional disease-related MeSH terms to the MMA (potentially confounding). Including them does not change the results (Table C-20).

Lastly, we chose *different aggregation levels* for our analysis. In Table C-21, we aggregate the dependent variable to the ICD-9 group level. Our results are robust to this aggregation.

Our results on drug development are robust to estimations at the ICD-9 three-digit code level (Table C-4), alternative controls (Table C-5), alternative MMS calculations (Figures C-9), and including the MMA implementation in 2005 into our baseline period (Table C-6).

3.8 Thought Experiment – Incentivizing Upstream Research

We conclude with a thought experiment. For this example and the ensuing back-of-the-envelope calculation, we will take our results at face value. The goal is to explore the magnitude of the linkage between changes in downstream demand and upstream research. We start by assuming that scientific publications are mutually exclusive, meaning they are either categorized as a university, academic medical center, or corporate publication. From our preferred specification, a MMS increase of one standard deviation (23.2 percentage points) in exposure to Medicare Part D, leads to around 1,834 additional publications per year.[42] These additional publications break down broadly as follows: 1,203 authored by scientists with university affiliations, 249 with academic medical center affiliations, 323 with corporate affiliations with the remaining 74 authored by scientists with NIH affiliations. Although the quantitative majority of these publications are focused in the subcategories where one would expect basic science research to occur, it remains open whether this response is meaningful.

The direct costs of Medicare Part D, paid as subsidies to private insurances, during the program's first ten years was expected to be $80 billion annually (Medicare Trustees Report, 2006). When comparing our estimates to these direct costs, our results suggest that a subsidy-driven expansion of market size by $43 million would only lead to one additional scientific publication. This is substantially lower than, for example, the direct benefits of public funding, for which Myers (2020) reports that the average cost per publication is between $344,000 and $665,000 depending on the grant regime. At the midpoint of this range, this suggests that

---

[42]We take the point estimate of our preferred specification and multiply it with the standard deviation in MMS and the 2003 number of scientific publications from Table 3.1.

direct public funding of research would generate about 85 publications for each additional publication from our findings.

In the pre-MMA period, only about 5.4% of publications were authored by scientists affiliated with corporations. This rises to about 17.6% in the post-MMA period. If the whole scientific domain was as responsive as corporations, the "cost" of one additional publication would fall from $43 million to $13.4 million. Direct public funding of research would still generate about 26 publications for each additional publication from our findings. Thus, even considering the most responsive case, it does not appear that changes in downstream demand serve as sufficient incentives for upstream research.

Putting the results from this thought experiment into a broader context, Finkelstein's (2004) assertion appears to be correct – the post-MMA change in development was driven by a reordering of technology already in the development pipeline. Even in the context of our thought experiment, the impact on research appears to be insufficient, especially compared to public funding.

3.9 Conclusion

R&D consists of two separate, but equally important components: research and development. The extant literature has conclusively found a link between changes in downstream market size and drug development – *i.e.*, 'D' – (*e.g.*, Acemoglu and Linn, 2004; Blume-Kohout and Sood, 2013; Dranove et al., 2020; Finkelstein, 2004). Heretofore, however, efforts to extend this linkage back to research – *i.e.*, 'R' – have been met with limited success. Acemoglu and Linn (2004), for example, were unable to find a relationship between demographic-driven expansions in market size and patenting. Using a similar identification strategy, Bhattacharya and Packalen (2011) find a positive relationship between disease prevalence and upstream research conducted at academic medical centers, but do not explicitly account for profit incentives. By exploiting the effects of quasi-experimental variation in market size introduced by Medicare Part D, we do not find a causal relationship between market size and research. We identify one limited exception to our core findings, however, and that involves corporate scientists conducting applied research.

Why does it appear that scientists are not incentivized by these changes in downstream market size? We can only conjecture as to what might explain the inelastic response. First, this type of demand pull incentive may attenuate by market distance so that it never reaches scientists (*e.g.*, Acemoglu and Linn, 2004). For example, if firms respond to demand shocks with 'off the shelf' projects (Finkelstein, 2004), new scientific discoveries may not be necessary to fuel the clinical pipeline. Similarly, the existing knowledge stock may be large enough to accommodate (for some time) the industry's higher demand for scientific discoveries. Our findings, however, suggest no response from university science even more than a decade later. Thus, a disconnect appears to exist between the kind of research industry uses as knowledge

inputs and the kind of research upstream scientists conduct. Such exploration of this disconnect is left for future work.

Second, scientists may react to market size changes only if these changes affect the scientists' incentives through indirect channels, such as altruism, funding, or prestige. This is perhaps more likely the case if the market size increase is due to disease prevalence (*e.g.*, epidemics) as opposed to insurance coverage (*e.g.*, Medicare Part D). Third, scientists may respond to demand pull differently, for example, by providing tacit knowledge through training (junior) scientists for industry (Roach and Sauermann, 2010). Unfortunately, we are unable to explore the above mechanisms with our data. Future research is necessary to understand the inelastic response of scientists. Survey evidence would be most helpful and may shed light on the mechanisms driving the disconnect between research and market size.

Finally, our study has important implications for policymakers. To the extent that there is a disconnect between the direction of academic research and the requirements of downstream markets, additional incentives may be needed to close this gap. NIH grants and public sector funding would appear to be the obvious choice as they are effective tools in fostering scientific research (Azoulay et al., 2019). It may also be the case that an expansion of R&D tax credits could be used to help incentivize companies to re-engage in and reverse the trend away from basic science research (Arora et al., 2018). These issues are left for future research.

A

Appendix to Chapter 1

Fire and Mice

A.1 Sources and Data Construction

Archival Documents

Table A-1: Reference list archival documents

Reference ID	Title	Year
JAX Archive-1	Paigen, K., Statement of Need in Blue Book	1989
JAX Archive-2	JAX, Brown Book 1989	1989
JAX Archive-3	Fire Investigation Report	1989
JAX Archive-4	Mouse News Letter	1989
JAX Archive-5	JAX, Letter to Recipients of Jackson Laboratory Mice	1989
JAX Archive-6	JAX, Annual reports	1988–1992
JAX Archive-7	NIH, Notice of Approval, RFA-OD-90-02	1990
JAX Archive-8	JAX, NIH Construction Grant Application	1990
JAX Archive-9	JAX, NIH application CA-56393-01	1991
JAX Archive-10	JAX, Letter Standel to NIH	1989
JAX Archive-11	JAX, Recovery Plan	1990
JAX Archive-12	JAX, Presentation slides about the fire	1989
JAX Archive-13	JAX, Reply to Charles River Laboratories "Fact Sheet"	1989
JAX Archive-14	JAX, Handwritten note attached to wait list	1990
JAX Archive-15	JAX, Price lists	1989–1990
JAX Archive-16	JAX, Jackson Lab Memorandum	1989
JAX Archive-17	Northeast Research, Survey results	1989
JAX Archive-18	JAX, Case Studies	1990
JAX Archive-19	ILAR, Strains in actual short supply (TJLA Grants)	1989
JAX Archive-20	ILAR, Strains in great demand (August)	1989
JAX Archive-21	JAX, Waiting list	1990
JAX Archive-22	JAX, Top 25 mice strains (sales data)	1988–1989
JAX Archive-23	JAX, Jax Notes 438 (July) on mice in good supply	1989

Notes: All documents were collected in May 2018 from the Joan Staats Library at JAX. A full inventory list of the documents available in the JAX archive can be found in The Jackson Laboratory Archives (2012a,b).

Figure A-1: Strains in short supply (JAX Archive-19) and on wait list (JAX Archive-21)

Figure A-2: Strains in good supply (JAX Archive-23)

Trustees Approve Recovery Plans

At a recent Jackson Laboratory board meeting, the governing trustees approved a two-phased plan to restore Morrell Park's production and distribution capability to its prefire status.

As a result of the May 10 fire, the production facility lost 55 percent of its capacity. By converting some animal holding to breeding space, the loss will be reduced to 40 percent.

"We will be able to reach 60 percent of our prefire capacity by fall," wrote Colin, the interim director and general manager of production and distribution Phil Standel to all JAX customers on June 19.

During the first phase — the interim recovery phase — the Laboratory will install prefabricated modular units, designed especially for breeding and raising genetically defined inbred and mutant mice. Standel estimates that production will reach 70 percent of prefire capacity during the winter and 80 percent during spring, 1990. Currently Lab planners are designing these and meeting with distributors to discuss specifications and costs. Plans for this phase also call for the Laboratory's renovating an existing warehouse to provide sterilization and cleaning facilities. These facilities, however, will remain on line after the interim recovery period.

While these facilities are being acquired and renovated, the Laboratory is also forging ahead with plans for the second phase: the construction of a permanent facility to replace the Morrell Park building.

According to Jackson Laboratory director Ken Paigen, the new facility will be the best not only for raising mice but also for meeting employee needs.

"Like the fire of '47, this fire has come at a time when the Laboratory knows what animal production should be," Paigen said, "and we can design to that."

Although the plans are still in the conceptual stage, lab officials envision that the new permanent structure will be a complex of relatively small, physically separated buildings with backup environmental and support utilities. "Such a design will allow individual rooms or sections to operate independently of one another," Standel said, "and will make the individual sections less vulnerable to disasters such as fire, contamination, or equipment failure.

"Within days of the fire, we began rebuilding those colonies that suffered the most severe losses, and for the next few months will use most of the progeny to expand these colonies. As a result our supplies of some strains will be very limited," added Standel.

All the washing, cleansing, and sterilizing operations needed to insure the genetic quality and purity of these mice are taking place in the Research Animal Facility, which is working multiple shifts.

Additional sterilizing backup will also be provided by equipment in the recently renovated and expanded importation facility in the research laboratory and by equipment in the new Foundation Stocks Facility which is scheduled for completion this fall.

Strains in Good Supply

C57BL/6J	..... both sexes	CSJLF./J	..... both sexes
B6D2F./J	..... both sexes	C3HeB/FeJ	..... males
BALB/cByJ	..... males	SJL/J	..... males
B6AF./J	..... both sexes	B6SJLF./J	..... males
B6C3F./J	..... both sexes	C57BL/KsJ-db	..... females
CB6F./J	..... females	MRL/MpJ-lpr	..... males
CAF./J	..... females		

Other Aids

As a precaution the Lab for a number of years now has been guaranteeing the supply of genetically defined mice by freezing embryos of about 600 strains and by maintaining in other facilities breeding pairs of most of the approximately 1100 strains and stocks available for distribution to researchers at The Jackson Laboratory and elsewhere.

To minimize the disruption to other research while strains are being replenished, the Laboratory has offered to assist scientists elsewhere in setting up their own breeding colonies.

"In addition to supplying breeding pairs, we are in position to advise researchers about animal husbandry techniques and to provide regular consultation," said Standel.

MRL/MpJ-lpr, males are available in good supply

Profile: C57BL/6J

Over the years, The Jackson Laboratory has raised and distributed countless C57BL/6Js. But has anyone wondered why this particular mouse, first bred in 1921 by Jackson Laboratory founder C. C. Little, has been so popular or acquired its name?

Research associate Priscilla Lane calls the C57BL/6J "the mouse of choice" because not only is it a good breeder and long lived but also has a low susceptibility to tumors and few unfavorable characterizations.

During the 1950s and 1960s staff scientist emeritus Elizabeth Russell placed many pigment mutations including those with anemias onto the C57BL/6J background. Later Richard Sidman of Harvard and a member of the board of scientific overseers here along with his coauthors used the C57BL/6J mouse as the model for the Atlas of the Mouse Brain and Spinal Cord. As a result of this study, neurological mutations arising spontaneously on this background were kept on it while many others were placed on it.

Because of its high alcohol preference, this strain has been used by some researchers to examine the mode of inheritance and the effects of alcohol on mice of different ages.

Blacks are also one of the oldest inbred strains. While at Cold Spring Harbor in 1921, Little started a new "family of mice" descended from mice received from Miss Abby Lathrop, a mouse supplier from Granby, Massachusetts.

Sometime after 1900 Lathrop began raising small animals, including a waltzing mouse, as pets. Although initially she planned only to raise two or three hundred, she began receiving orders for hundreds of mice from research institutions and her business expanded. Many of the mice used, for example, in the coat color experiments conducted while Little was an undergraduate at Harvard came from Lathrop.

To start this strain, Little bred female mouse number 57 to male mouse number 52. The result was a black mouse soon designated C57Black. Based on the earlier coding used by Little such as d, b, a for dilute, brown and nonagouti and c for albinos, Lane theorizes that the upper case "C" may have stood for the nonalbino allele (the normal full coloraltele). Today, however, all strain names are expected to adhere to standard nomenclature rules.

But the C57BL/6J mice are not only an old and venerable strain, they are also very lucky. During the May 10, 1989 fire, these mice were in a section of Morrell Park that was behind a fire wall. This JAX strain was not affected by the fire, and consequently stocks of this mouse are at prefire levels.

"Fortunately for researchers the popular C57BL/6J strain is available in large enough quantities to supply both scientists who have traditionally used this mouse and those who may now elect to use it," said Barbara Witham, manager of service and distribution.

JAX Notes

is published for the users of JAX Mice. If you would like to be added to the mailing list to receive JAX Notes, please contact Animal Resources, The Jackson Laboratory, 600 Main Street, Bar Harbor, Maine 04609-0800, or call (207) 288-3371 from 8:30 AM to 4:30 PM (Eastern Time Zone) Mondays through Fridays.

THE JACKSON LABORATORY
BAR HARBOR, MAINE 04609

Figure A-3: Gene-loci strain linkage from (Lyon and Searle, 1989)


```
The laboratory mouse

    Locus   ->    a A A A A   A A A A A   A A A A A   A A A A A   A A A A A
    (read down)     a b c c   d d d d g   h h h h k   k l m m o   o p p p s
                    t p f o   h h h h s   d d r p     p b y y x   x h k o -
                    a - -     - - - -         - - - -   - - - - -   - -   a 1
                        1 1   1 2 3 3       1 2 1 1   2 1 1 2 1   2 1   - s
                                  e t                                     1

    ---------------------------------------------------------------------------------
    Chromosome     2 1 7 1 4   3 U 3 3 X   1 4 1 3 1   4 5 3 3 1   1 2 1 9 1
    (read down)      2           N           2   9                   0     3
    ---------------------------------------------------------------------------------

    BRSUNT/N       a . . . .   . . . . .   d . . . .   . . . . .   . . . . .
    BS             . . . . .   . . . . .   . . . . .   . a . a .   . . . . .
    BUB/BnJ        a a . b a   . . . . h   . b . . b   a a a a 1   . . a b a
    BXSB/MpJ       . . . . .   . . . . .   . . . . b   . . a . .   . . . a .
    C3H/-          + o a a a   a a a a h   b b a a b   a a a a a   a a a b a

    C3H/An         . . . . .   . . . . .   . . . . .   . . . . .   . . . . .
    C3H/Bi         + . . . .   . . . . .   . . . . .   . . . . .   . . . . .
    C3H/BiU        . . . . .   . . . . .   . . . . .   . . . . .   . . . . .
    C3H/Crgl       . . . . .   . . . . .   . . . . .   . . . . .   . . . b .
    C3H/He-        + o a a a   a a a a h   b b a a b   a a a a a   a a a b a

    C3H/HeA        + . . . .   a . a a .   b b . a .   . . . . a   a . . . .
    C3H/HeAf       . . . . .   . . . . .   . . . . .   . . . . .   . . . . .
    C3H/HeDiSnA    + . . . .   a a a a .   b b a . .   . . . . a   a . . . .
    C3H/HeJ        + o a a a   . . . . h   b b a . b   a a a a 1   . . a b a
    C3H/HeOuJ      . . . . .   . . . . .   . . . . b   . . a . .   . . . b .

    C3H/HeSnJ      + . a . .   . . . . .   . . . . b   . . a . .   . . a b .
    C3H/St         + . a . .   . . . . .   . . . . .   . . . . .   . . . . .
    C3H/StWi       + . . . .   . . . . .   . . . . .   . . . . .   . . . . .
    C3HeB/FeJ      + o a a a   . . a . .   b b . . b   a a a a 1   . . a b .
    C3HfB/HeN      + . . . .   . . . . .   . . . . .   . . . . .   . . . . .

    C57BL/6-       a o a a a   a a b b h   b a a b a   b a a a a   a a a a b
    C57BL/6ByA     a . . . .   a a b b .   b a a b .   . . . . a   a . . . .
    C57BL/6ByJ     a . a . a   . . . . .   b . . . a   . . a . .   . . a a .
    C57BL/6J       a o a a a   a . b b h   b a a b a   b a a a a   . . a a b
    C57BL/6JN      a . . . .   . . . . .   b . . . .   . . . . .   . * . . .

    C57BL/10-      a o a . a   a a b b h   b a a b a   b a a . a   a a a a b
    C57BL/10J      a o . . a   . . . . h   b a a . a   b a a . .   . . a a b
    C57BL/10ScSnA  a . . . .   a a b b .   b a a b .   . . . . a   a . . . .
    C57BL/10ScSnJ  a o a . .   . . . . .   b a a b .   . . . . a   a . a a .
    C57BL/AnHf     a . . . .   . . . . .   . . . . .   . . . . .   . . . . .

    C57BL/ImrHeA   a . . . .   a a b b .   b a a b .   . . . . a   a . . . .
    C57BL/KaLwN    a . . . .   . . . . .   . . . . .   . . . . a   . . . . .
    C57BL/KsJ      a o a . a   . . a . .   b b . . a   b a a a 2   . . a a b
    C57BL/LiA      a . a . .   a . b b .   b . . b .   . . . . .   . . . . .
    C57BR/cd-      a o a . a   . . . . h   b a . . a   b a a a a   . . a a b

    C57BR/cdJ      a o a . a   . . . . h   b a . . a   b a a a 1   . . a a b
    C57L/-         a o a . a   . . . . h   b a . . a   b a a a 1   . . a a b
    C57L/J         a o a . a   . . . . h   b a . . a   b a a a 1   . . a a b
    C57P/A         a . . . .   . . . . .   b . . . .   . . . . .   . . . . .
    C58/-          a o a a a   . . a . h   b a a a a   b . a a 2   . . a a a

678        1 = a?        2 = b?
```

A. APPENDIX TO CHAPTER 1

Scopus Queries to Identify Strains

- Search for Jax ("J") mice strains:

    1. Strain-J in title, abstracts or keywords

    2. Exclude strains with same beginning

    3. At least one "mice"-like keyword

```
(TITLE—ABS—KEY("129X1/SvJ ")
AND NOT TITLE—ABS—KEY("129X1/Sv —Lamc2 jeb /Dcr "))
AND KEY(mouse OR mice OR rat OR rats OR strain)
```

- For non-"J" mice strains:

    1. Strain in title, abstracts or keywords

    2. Exclude strains-J nomenclature

    3. . . .

```
(TITLE—ABS—KEY("129X1/Sv ")
AND NOT TITLE—ABS—KEY("129X1/SvJ ")
AND NOT TITLE—ABS—KEY("129X1/Sv —Lamc2 jeb /Dcr "))
AND KEY(mouse  OR  mice  OR  rat  OR  rats  OR  strain)
```

A.2 Figures at Mice Level

Figure A-4: Distribution of known gene-loci by strain in 1989
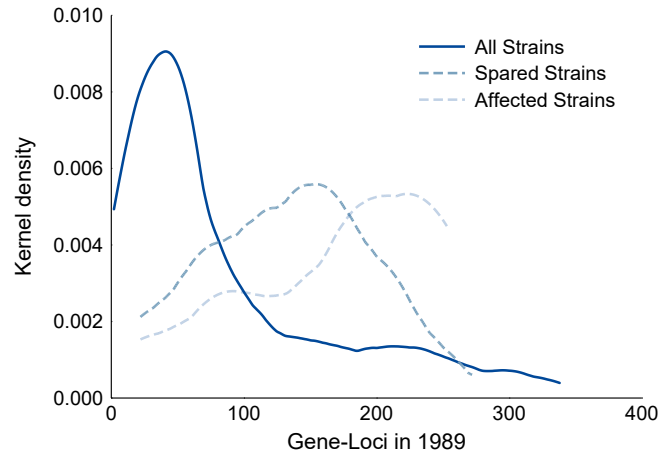
Notes: The figure presents the number of known gene-loci by strain from (Lyon and Searle, 1989). "All strains" include all 246 strains mentioned in the gene-loci strain matrix curated by JAX, "spared strains" include 19 strains that are in our sample of identified spared mice and in the curated matrix, "affected strains" include 26 strains that are in our sample of identified affected mice and in the curated matrix. The unit of observations is the individual mice strain.

Figure A-5: Number of publications B6 vs. 129 strains
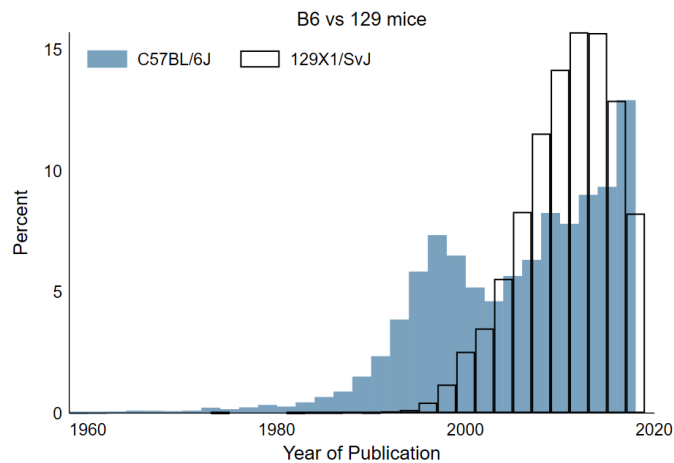
Notes: The figure presents the distribution of publications related to the B6 strain and related to an exemplary 129 substrain, which is in our sample, over time. The unit of observations is the individual mice strain. The figure shows that B6 increases after the fire and much earlier than the transgenic-mice related rise in 129 mice usage.

Figure A-6: Mice usage – publication counts (alternative classification)

Aggregated counts (MGI data)

(a) Publications   **(b)** Publications (JIF-weighted)

Average counts (Scopus data)

(c) Publications   **(d)** Publications (JIF-weighted)

Notes: The top left-hand figure presents the log-transformed annual sums of publications linked to affected and spared mice strains. The top right-hand figure presents the log-transformed annual sums of publications linked to affected and spared mice strains weighted by their respective journal impact factor. Publications linked to affected and spared mice are included twice. The bottom left-hand figure presents the log-transformed average publication counts at mice strain level. The bottom right-hand figure presents the log-transformed average publication counts weighted by their respective journal impact factor. Publications can be linked to multiple mice strains. In both bottom graphs, the unit of observation is the unique mice strain.

Figure A-7: Event studies at mice strain level
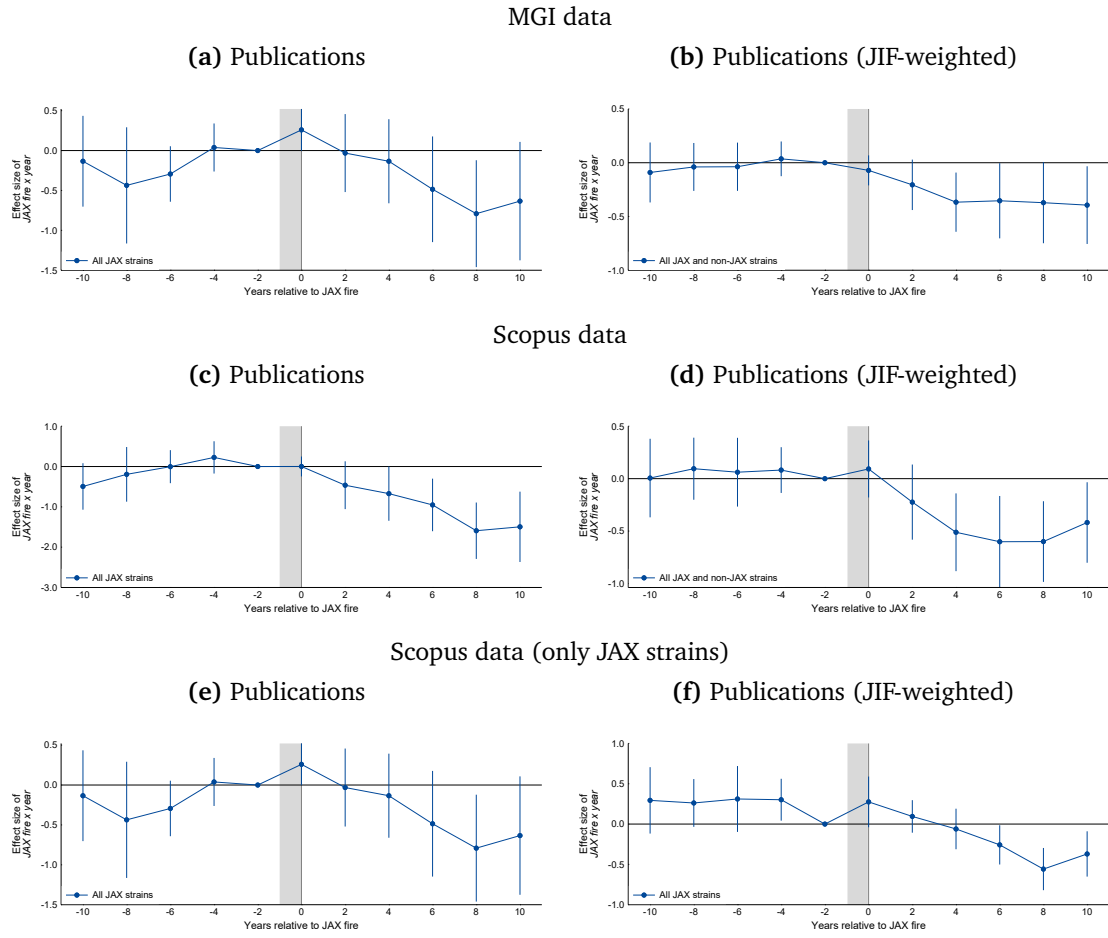
MGI data

(a) Publications

(b) Publications (JIF-weighted)

Scopus data

(c) Publications

(d) Publications (JIF-weighted)

Scopus data (only JAX strains)

(e) Publications

(f) Publications (JIF-weighted)

Notes: The figures show the event study estimates and the 90 percent confidence bands of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects following Equation 1.1 based on MGI data in (a) and (b) and on Scopus data in (c) to (f). The unit of observation is the individual strain by year. Standard errors are clustered at the strain level.

A.3 Figures at Scientist Level

A.3.1 Figures – Summary Statistics/Descriptive Analysis

Figure A-8: Distribution of scientist characteristics
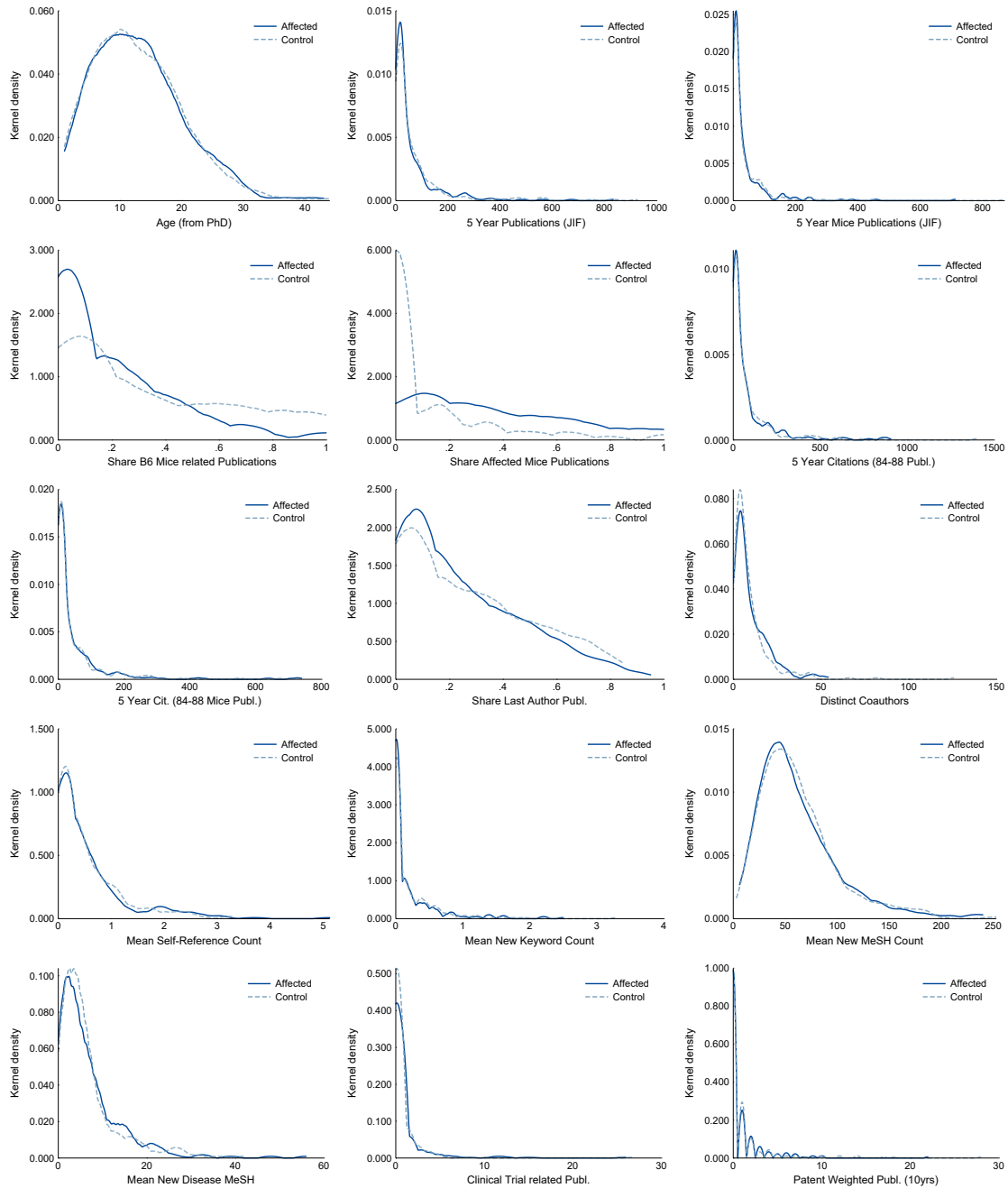
Notes: The figures compare the Kernel density of aggregated pre-fire characteristics of affected and control scientists in the five years before the fire. The unit of observation is the scientist level.

A. APPENDIX TO CHAPTER 1

Figure A-9: Mean comparison – alternative samples

(a) Sample B (strict)   **(b)** Sample C (pre-matching)
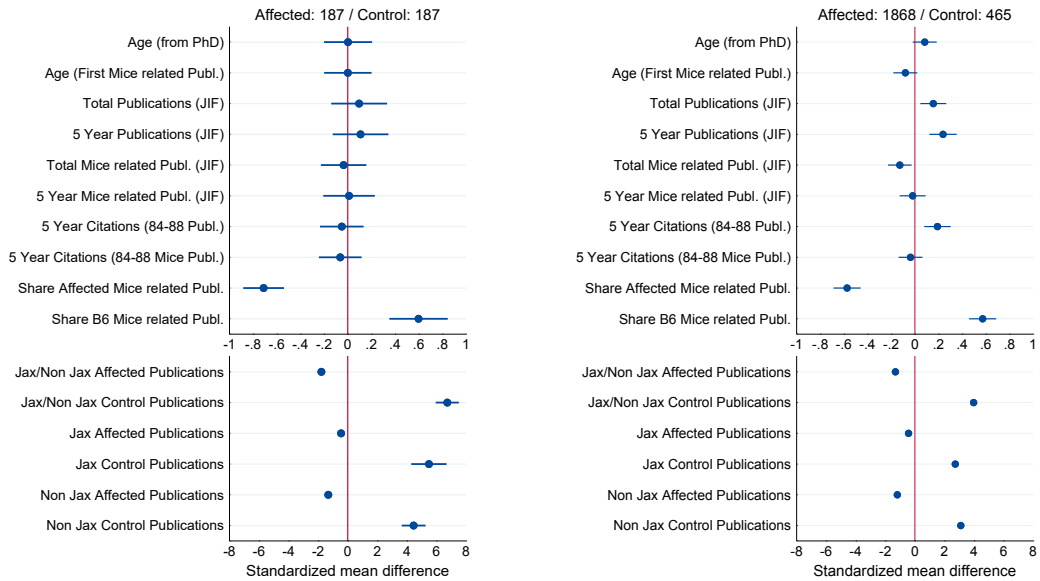
Notes: The left-hand figure presents comparisons of standardized mean differences of key pre-fire characteristics of the scientists in sample B, where additional emphasize was given on a *similar age and similar pre-fire productivity*. The right-hand figures presents sample C with scientists *before matching*. The unit of observation is at the scientist level.

Figure A-10: Comparison of affected/unaffected scientists

(a) Last publication activity - sample A   **(b)** Last publication activity - sample USA
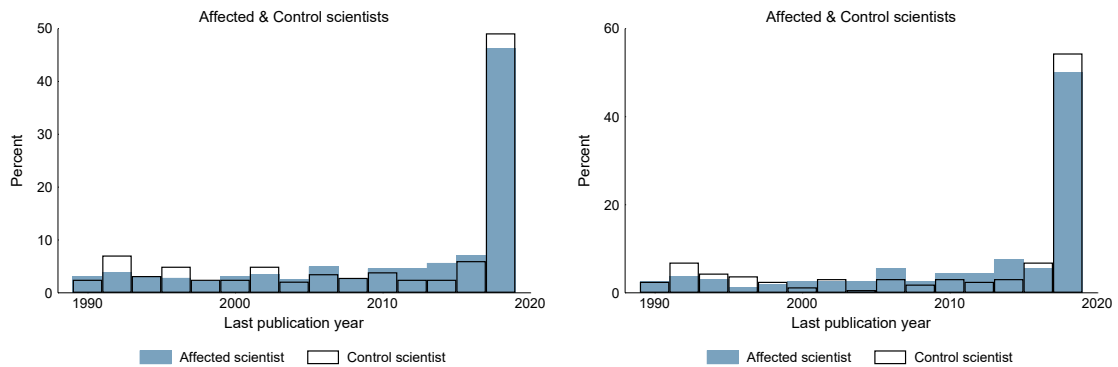
Notes: The figure presents the distribution of years of each scientist's last publication (for the 20-year time frame after the fire) between affected and control scientists for sample A and sample USA. The unit of observation is the individual scientist.

Figure A-11: Scientist publications linked to affected and spared mice strains

(a) Share affected publ. (JIF) - sample A
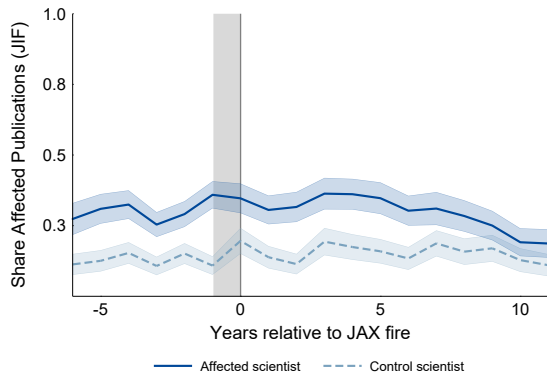
(b) Share affected publ. (JIF) - sample USA
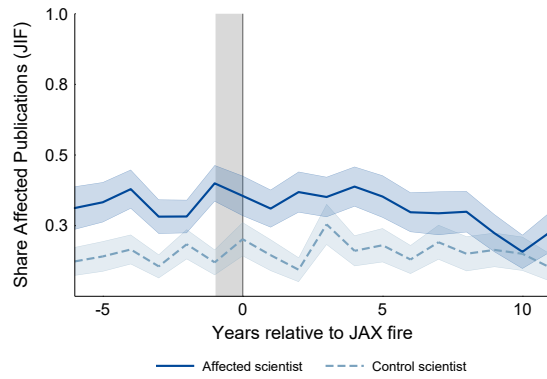
(c) Mice publ. (JIF) - sample A
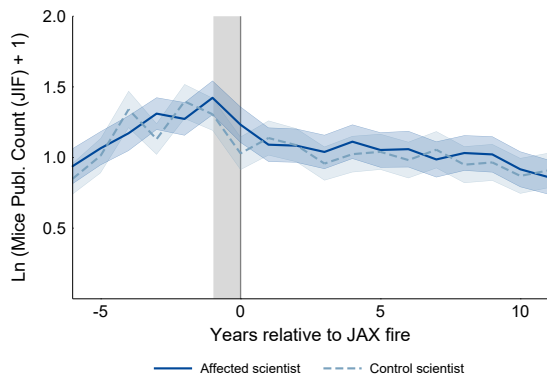
(d) Mice publ. (JIF) - sample USA
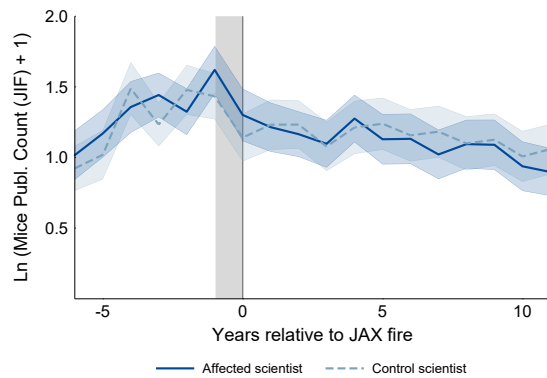
Notes: Figures (a) and (b) illustrate the mean share of affected publications relative to all mice publications (both JIF-weighted) over time for affected and control scientists. Figures (c) and (d) show the mean JIF-weighted mice publication counts. The unit of observation is the scientist level.

A.3.2 Figures – Multivariate Analysis

Figure A-12: Event studies at scientist level

(a) Self-references - sample A   **(b)** Self-references - sample USA
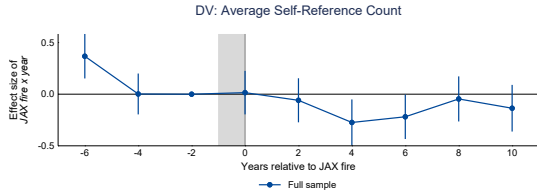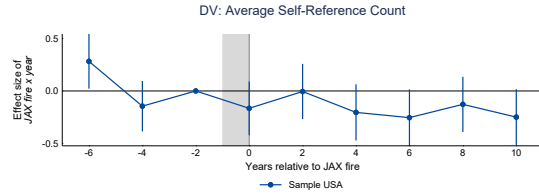
(c) New disease MeSH - sample A   **(d)** New disease MeSH - sample USA
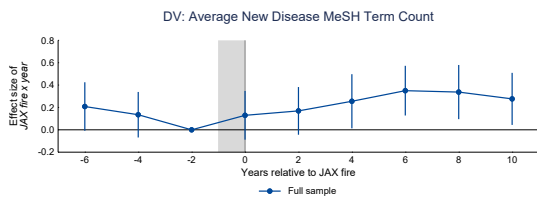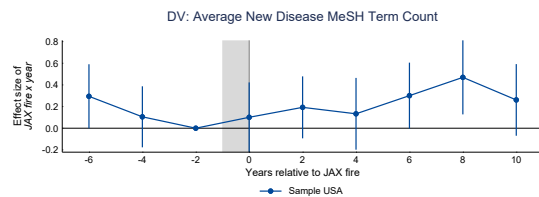
(e) Clinical journal - sample A   **(f)** Clinical journal - sample USA
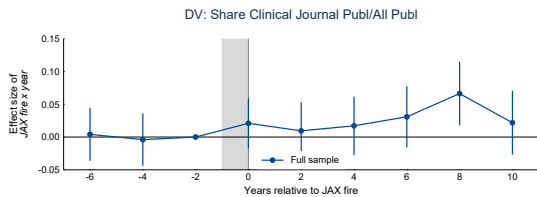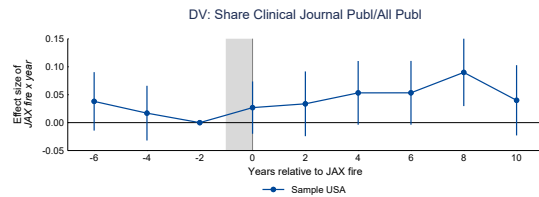
(g) Patent-weighted - sample A   **(h)** Patent-weighted - sample USA
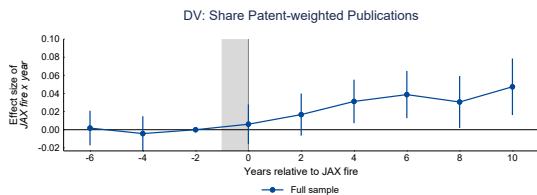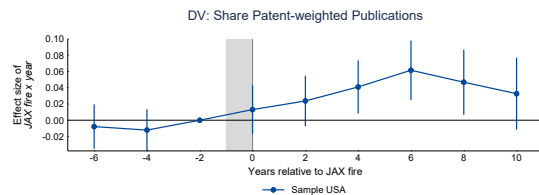
Notes: The figures show the event study estimates and the 90 percent confidence bands of linear regressions with high-dimensional fixed effects following Equation 1.2. The outcome variable is the mean number of self-references in (a) and (b), the mean number of new disease related MeSH terms in (c) and (d), the share of publications published in a journal that usually publishes more clinical relevant research in column (e) and (f), and the share of publications that is associated with a patent application (patent-weighted). In both samples, the unit of observation is the individual scientist by year. Standard errors are clustered at the scientist level.

A.3.3 Figures – Heterogeneity Analysis

Figure A-13: Heterogeneity by age
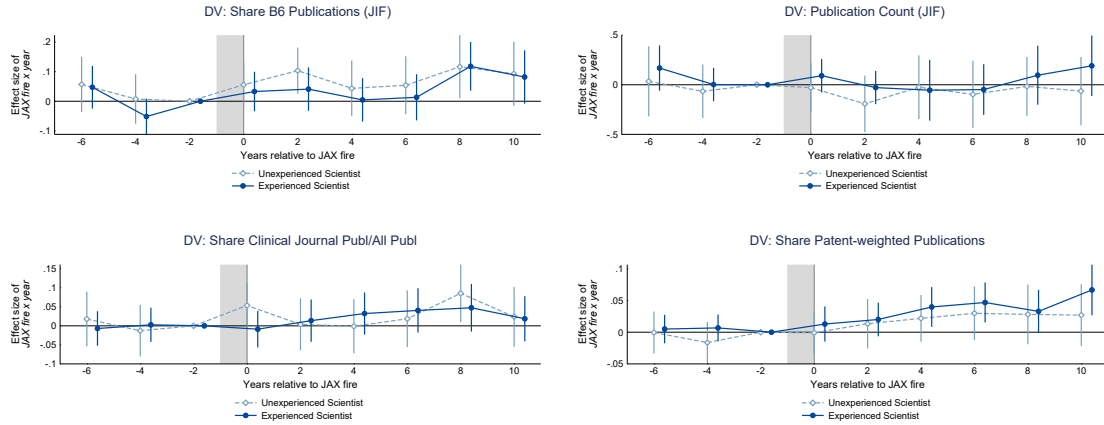
Notes: The figures show the event study estimates and the 90 percent confidence bands of regressions following Equation 1.2 with a triple-interaction term using sample A. The triple-interaction term is the age (since PhD) of a scientists at the time of the fire. Count data is estimated using Poisson pseudo maximum likelihood regressions and shares are estimated using linear regressions. The unit of observation is the individual scientist by year. Standard errors are clustered at the scientist level.

A.3.4 Figures – Patents

Figure A-14: Most often used words in patents associated with scientists' publications
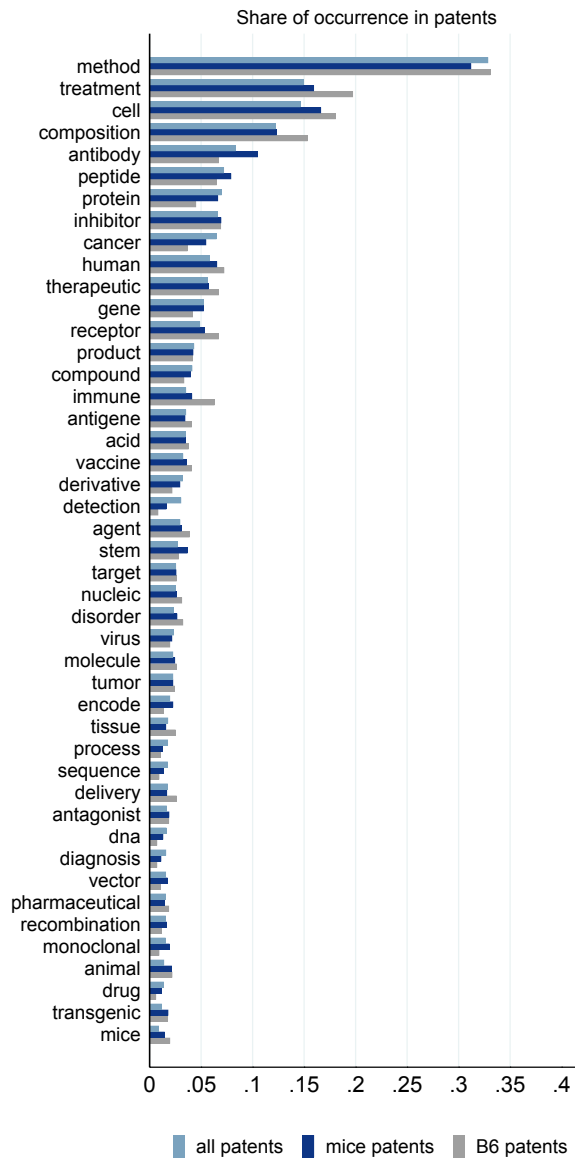
Notes: The figure presents the share of occurrences of the 45 most often used words in the titles of U.S. patent applications related to the 566 scientists in our sample A.

A.4   Tables at Mice Level

Table A-2: Distribution of JAX mice by sector and region in 1988

| | USA | | CAN | | Rest | |
Sector	N	%	N	%	N	%
Government agencies	140,822	7.3	176	0.0	775	0.0
Hospitals	238,502	12.4	30,245	1.6	644	0.0
Industry	503,668	26.2	7,428	0.4	15,358	0.8
Medical schools and universities	873,127	45.4	50,284	2.6	63,753	3.3
Total	1,756,179	91.2	88,133	4.6	80,530	4.2

Notes: Source of data: JAX Archive-2.

Table A-3: JAX mice strains

JAX strain id	Strain name	Strain type	Group	Competitor
JAX:001137	129P1/ReJ	inbred strain	mouse control	
JAX:000690	129P3/J	inbred strain	mouse control	
JAX:000691	129X1/SvJ	inbred strain	mouse control	
JAX:000140	A.By/SnJ	congenic/mutant strain	mouse control	
JAX:000471	A.SW/SnJ	congenic/mutant strain	mouse control	
JAX:000645	A/HeJ	inbred strain	mouse control	
JAX:000646	A/J	inbred strain	mouse treated	
JAX:000647	A/WySnJ	inbred strain	mouse control	
JAX:100001	AKD2F1/J	F1 hybrid strain	mouse treated	
JAX:000648	AKR/J	inbred strain	mouse treated	Yes
JAX:000820	AKR/J-nu^stn	congenic/mutant strain	mouse control	
JAX:000467	B10.A(5R)/SgSnJ	congenic/mutant strain	mouse control	
JAX:000469	B10.A/SgSnJ	congenic/mutant strain	mouse treated	
JAX:000465	B10.BR/SgSnJ	congenic/mutant strain	mouse treated	
JAX:000463	B10.D2/nSnJ	congenic/mutant strain	mouse treated	
JAX:000461	B10.D2/oSnJ	congenic/mutant strain	mouse treated	
JAX:000457	B10.RIII(71INS)/SnJ	congenic/mutant strain	mouse control	
JAX:001060	B6.C-H-2^bml/ByJ	congenic/mutant strain	mouse control	
JAX:100002	B6AF1/J	F1 hybrid strain	mouse control	
JAX:100010	B6C3F1/J	F1 hybrid strain	mouse control	
JAX:100011	B6CBAF1/J	F1 hybrid strain	mouse treated	
JAX:100006	B6D2F1/J	F1 hybrid strain	mouse control	Yes
JAX:100012	B6SJLF1/J	F1 hybrid strain	mouse control	
JAX:001026	BALB/cByJ	inbred strain	mouse control	Yes
JAX:000711	BALB/cByJ- nu	congenic/mutant strain	mouse control	Yes
JAX:000651	BALB/cJ	inbred strain	mouse treated	Yes
JAX:000740	BXSB/MpJ	inbred strain	mouse treated	
JAX:100004	C3D2F1/J	F1 hybrid strain	mouse treated	Yes
JAX:000438	C3H.Sw/SnJ	congenic/mutant strain	mouse treated	
JAX:000659	C3H/HeJ	inbred strain	mouse treated	Yes
JAX:000509	C3H/HeJ- bg^J	congenic/mutant strain	mouse control	
JAX:000635	C3H/OuJ	inbred strain	mouse treated	
JAX:000658	C3HeB/FeJ	inbred strain	mouse control	
JAX:000665	C57BL/10J	inbred strain	mouse treated	Yes
JAX:000666	C57BL/10SnJ	congenic/mutant strain	mouse treated	
JAX:001139	C57BL/6ByJ	inbred strain	mouse control	
JAX:000664	C57BL/6J	inbred strain	mouse control	Yes
JAX:000629	C57BL/6J-bg^J	congenic/mutant strain	mouse treated	
JAX:000819	C57BL/6J-nu	congenic/mutant strain	mouse treated	
JAX:000632	C57BL/6J-ob	congenic/mutant strain	mouse treated	
JAX:000160	C57BL/6J-sl^d	congenic/mutant strain	mouse control	
JAX:000049	C57BL/6J-w^v	congenic/mutant strain	mouse control	
JAX:000662	C57BL/KsJ	inbred strain	mouse control	

Table A-4: JAX mice strains (*continued*)

JAX strain id	Strain name	Strain type	Group	Competitor
JAX:000642	C57BL/KsJ-db	congenic/mutant strain	mouse control	
JAX:000667	C57BR/cdJ	inbred strain	mouse control	
JAX:000668	C57L/J	inbred strain	mouse treated	
JAX:000669	C58/J	inbred strain	mouse treated	
JAX:100003	CAF1/J	F1 hybrid strain	mouse control	Yes
JAX:100007	CB6F1/J	F1 hybrid strain	mouse control	Yes
JAX:000655	CBA/CaH-T6J	congenic/mutant strain	mouse treated	
JAX:001011	CBA/CaHN-xid/J	inbred strain	mouse treated	
JAX:000654	CBA/CaJ	inbred strain	mouse treated	Yes
JAX:000656	CBA/J	inbred strain	mouse treated	Yes
JAX:100009	CByB6F1/J	F1 hybrid strain	mouse control	
JAX:000657	CE/J	inbred strain	mouse control	
JAX:100019	CSJLF1/J	F1 hybrid strain	mouse control	
JAX:000670	DBA/1J	inbred strain	mouse treated	
JAX:001140	DBA/1LacJ	inbred strain	mouse control	
JAX:000671	DBA/2J	inbred strain	mouse treated	Yes
JAX:000643	DW/J	congenic/mutant strain	mouse control	
JAX:000673	HRS/J	inbred strain	mouse treated	Yes
JAX:000674	I/LnJ	inbred strain	mouse control	
JAX:100005	LAF1/J	inbred strain	mouse treated	
JAX:000676	LP/J	inbred strain	mouse treated	
JAX:000677	MA/MyJ	inbred strain	mouse treated	
JAX:000486	MRL/MpJ-+	inbred strain	mouse control	Yes
JAX:000485	MRL/MpJ-lpr	congenic/mutant strain	mouse control	Yes
JAX:001976	NOD	inbred strain	mouse treated	Yes
JAX:000684	NZB	inbred strain	mouse treated	
JAX:100008	NZBWF1/J	F1 hybrid strain	mouse treated	
JAX:001058	NZW/LacJ	inbred strain	mouse treated	
JAX:000680	PL/J	inbred strain	mouse treated	
JAX:100299	(PL/J F x SJL/J M)F1	F1 hybrid strain	mouse treated	
JAX:000726	RBF/DnJ	inbred strain	mouse control	
JAX:000682	RF/J	inbred strain	mouse treated	
JAX:000683	RIIIS/J	inbred strain	mouse treated	
JAX:000687	SM/J	inbred strain	mouse control	
JAX:000688	ST/bJ	inbred strain	mouse control	
JAX:000689	SWR/J	inbred strain	mouse treated	
JAX:100525	WB/ReJ	inbred strain	mouse control	
JAX:000692	WB/ReJ-w	inbred strain	mouse control	
JAX:100410	WBB6F1/J-w/w^v	congenic/mutant strain	mouse control	
JAX:000693	WC/REJ-sl	congenic/mutant strain	mouse control	
JAX:100401	WCB6F1/J-sl/sl^d	congenic/mutant strain	mouse control	

Notes: The table describes the 84 mice strains used in the analysis.

Table A-5: Inbred strains and research application from Lutz et al. (2012)

Parent strain	Strain abbreviation	Research applications
129P3/J	129P	Spontaneous testicular teratomas, targeted mutagenesis
129/S1/SvImJ	129S	Spontaneous testicular teratomas, targeted mutagenesis
A/J	A	Widely used in cancer and immunology research; low-incidence cleft palate
AKR/J	AK	High incidence of leukaemia
BALB/c	Cby	General purpose immunology
C3H/HeJ	C3	General purpose strain in a wide variety of research areas including cancer, infectious disease, sensorineural and cardiovascular biology research
C57BL/6J	B6	General purpose, cardiovascular biology research, background strain for most mice carrying transgenes, spontaneous or targeted mutations
C57BL10/J	B10	General purpose
DBA/1J	D1	Widely used as a model for rheumatoid arthritis; in response to challenge, mice develop immune-mediated nephritis
DBA/2J	D2	General purpose, show low susceptibility to developing atherosclerotic aortic lesions; used in glaucoma research
NZW/LacJ	NZW	Type 1 diabetes
NZB/B1NJ	NZB	Autoimmunity
SJL		Cancer (reticulum cell sarcomas), autoimmunity (experimental allergic encephalomyelitis, EAE)
SWR	SW	General purpose; ageing mice exhibit a high incidence of lung and mammary gland tumours. Highly susceptible to experimental allergic encephalomyelitis

Notes: The figure provides examples for the heterogeneity in application between mice strains.

Table A-6: Research output (strain level) – no transgenic mice

Strains (no transgenic) Sample:	(1) MGI JAX strains	(2)	(3) Scopus JAX strains	(4)	(3) Scopus all strains	(4)
+10 years	Publ	Publ (JIF)	Publ	Publ (JIF)	Publ	Publ (JIF)
Post × affected	−0.664*	−0.866**	−0.194*	−0.388***	−0.248	−0.449
	(0.384)	(0.404)	(0.110)	(0.122)	(0.209)	(0.300)
Strain FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1344	1344	693	693	1239	1239
Strains	64	64	33	33	59	59
Log-likelihood	−2257	−11031	−1335	−3388	−3167	−9385

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the individual mice strain by year. All models include common trend interactions. The sample excludes publications with the MeSH term "transgenic". Standard errors are clustered at the mice strain level and shown in parentheses. Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A-7: Research output (strain level) – heterogeneity

Strains Sample:	(1) Scopus all strains	(2)	(3)	(4)	(3)	(4)
+10 years	Publ	Publ (JIF)	Publ	Publ (JIF)	Publ	Publ (JIF)
Post × affected	−0.035	−0.127	−0.058	−0.134	−0.027	−0.097
	(0.130)	(0.175)	(0.132)	(0.177)	(0.128)	(0.171)
× low JAX sales	−0.934***	−0.950***				
	(0.222)	(0.367)				
× no other suppliers			−0.789***	−0.889**		
			(0.226)	(0.353)		
× few gene loci					−0.885***	−1.286**
					(0.307)	(0.627)
Strain FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1239	1239	1239	1239	1239	1239
Strains	59	59	59	59	59	59
Log-likelihood	−2870	−7793	−2863	−7615	−2860	−7644

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the individual mice strain by year. All models include common trend interactions (*e.g.*, *low JAX sales x Year FEs*). Standard errors are clustered at the mice strain level and shown in parentheses. Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A-8: Research output (strain level) – author age

Strains Sample:	(1)	(2)	(3)	(4)	(5)	(6)
			Scopus all strains			
	Age of authors (avg)		Age of authors (min)		Age of authors (max)	
	Young	Old	Young	Old	Young	Old
+10 years	Publ (JIF)	Publ (JIF)	Publ (JIF)	Publ (JIF)	Publ (JIF)	Publ (JIF)
Post × affected	−0.299	−0.660**	−0.378	−0.547**	−0.394	−0.594*
	(0.288)	(0.292)	(0.313)	(0.275)	(0.266)	(0.316)
Strain FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1218	1218	1197	1239	1218	1218
Strains	58	58	57	59	58	58
Log-likelihood	−5762	−7305	−5025	−7851	−6173	−6780

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the individual mice strain by year. All models include common trend interactions. Age is defined at the team level. In Columns (1) and (2), we define team age at the mean, in Columns (3) and (4) at the age of the youngest team member, and in Columns (5) and (6) at the age of the oldest team member. In all these specifications, the sample is split at the median in "young" and "old". Standard errors are clustered at the mice strain level and shown in parentheses. Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

A.5 Tables at Scientist Level

A.5.1 Tables – Summary Statistics

Table A-9: Summary statistics (scientist level in sample A)

Affected vs Control	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Affected (N = 283)			Controls (N = 283)				
	Mean	Median	Std. Error	Mean	Median	Std. Error	Diff.	p-value
Age (from PhD)	12.92	12.00	7.26	12.85	12.00	7.27	−0.07	0.908
Age (First Mice Publ.)	9.72	8.00	6.31	9.53	8.00	6.05	−0.19	0.714
Total Publications (JIF)	131.82	49.57	212.08	147.13	57.14	255.02	15.31	0.438
5 Year Publications (JIF)	62.01	25.82	100.36	72.54	29.23	125.19	10.53	0.270
Total Mice Publications (JIF)	73.62	25.24	135.55	69.80	24.13	125.79	−3.82	0.729
5 Year Mice Publications (JIF)	37.63	13.95	70.81	39.50	12.83	79.14	1.87	0.767
5 Year Citations (84-88 Publ.)	82.01	27.00	140.93	81.10	29.00	144.64	−0.92	0.939
5 Year Cit. (84-88 Mice Publ.)	46.95	13.00	93.04	43.05	14.00	77.90	−3.90	0.589
Share Affected Mice Publications	0.35	0.25	0.31	0.12	0.00	0.22	−0.23	0.000***
Share B6 Mice related Publications	0.19	0.11	0.23	0.32	0.21	0.33	0.14	0.000***
Jax/Non Jax Affected Publications	2.82	2.00	1.45	0.35	0.00	0.80	−2.46	0.000***
Jax/Non Jax Control Publications	0.14	0.00	0.38	2.87	2.00	2.09	2.73	0.000***
Jax Affected Publications	0.59	0.00	1.03	0.17	0.00	0.47	−0.42	0.000***
Jax Control Publications	0.07	0.00	0.28	1.61	1.00	2.21	1.53	0.000***
Non Jax Affected Publications	2.31	2.00	1.63	0.19	0.00	0.63	−2.12	0.000***
Non Jax Control Publications	0.07	0.00	0.25	1.28	1.00	1.37	1.21	0.000***
Affiliation Rank	179.22	172.00	157.37	192.62	155.00	181.09	13.40	0.348
Top 5 Percent Publications	1.34	0.00	3.60	1.69	0.00	4.13	0.35	0.288
Share Last Author Publ.	0.24	0.17	0.23	0.26	0.20	0.24	0.02	0.276
Share Last Author Mice Publ.	0.21	0.15	0.24	0.22	0.15	0.25	0.01	0.662
Distinct Coauthors	9.87	6.20	9.78	9.68	6.00	12.34	−0.19	0.836
Distinct New Coauthors	5.70	4.00	5.52	5.65	3.40	7.34	−0.05	0.927
Mean Self-Reference Count	0.53	0.28	0.72	0.50	0.30	0.61	−0.03	0.611
Mean New Keyword Count	0.19	0.00	0.38	0.20	0.00	0.38	0.00	0.900
Clinical Journal Publ.	4.70	3.00	9.23	5.25	3.00	7.11	0.55	0.426
Patent Weighted Publ. (10yrs)	1.03	0.00	2.24	1.17	0.00	3.00	0.14	0.525
Mean New MeSH Count	62.26	52.00	39.80	62.02	53.00	37.29	−0.24	0.940
Mean New Disease MeSH	6.61	4.00	7.41	6.40	4.00	6.72	−0.22	0.713
In Vitro related Publications	3.86	2.00	5.17	3.42	1.00	5.47	−0.44	0.328
Other animal related Publications	2.52	1.00	7.21	4.34	1.00	8.32	1.81	0.006***
Human related Publications	7.28	2.00	12.15	5.67	2.00	11.11	−1.61	0.101
Diversity MeSH	377.23	312.00	277.50	385.24	326.00	290.00	8.02	0.737

Notes: The table compares pre-fire scientists characteristics by treatment group with t-tests and displays the mean, median and the standard errors for each group using sample A. If not stated otherwise, observations are aggregated in the period 1984-1988 (5 years pre-fire). The unit of observation is the individual scientist. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A-10: Summary statistics (scientist level in sample USA)

Affected vs Control	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Affected (N = 160)			Controls (N = 160)				
	Mean	Median	Std. Error	Mean	Median	Std. Error	Diff.	p-value
Age (from PhD)	12.97	12.00	7.31	12.93	12.00	7.32	−0.04	0.957
Age (First Mice Publ.)	9.54	8.00	6.24	9.24	8.00	5.82	−0.30	0.657
Total Publications (JIF)	166.56	65.77	250.77	178.24	71.94	295.96	11.68	0.704
5 Year Publications (JIF)	78.80	32.81	120.21	91.33	38.85	146.96	12.53	0.405
Total Mice Publications (JIF)	87.61	28.87	156.35	81.60	26.03	147.02	−6.01	0.723
5 Year Mice Publications (JIF)	46.26	18.09	84.27	47.69	15.03	92.39	1.43	0.885
5 Year Citations (84-88 Publ.)	107.31	42.00	171.93	99.86	43.50	146.82	−7.44	0.677
5 Year Cit. (84-88 Mice Publ.)	60.89	17.50	115.39	54.16	17.50	93.20	−6.73	0.567
Share Affected Mice Publications	0.37	0.31	0.31	0.14	0.00	0.22	−0.23	0.000***
Share B6 Mice related Publications	0.16	0.07	0.21	0.26	0.13	0.31	0.10	0.001***
Jax/Non Jax Affected Publications	2.68	2.00	1.36	0.35	0.00	0.83	−2.33	0.000***
Jax/Non Jax Control Publications	0.18	0.00	0.43	2.79	2.00	2.35	2.61	0.000***
Jax Affected Publications	0.83	0.00	1.16	0.17	0.00	0.49	−0.66	0.000***
Jax Control Publications	0.10	0.00	0.32	1.45	1.00	2.42	1.35	0.000***
Non Jax Affected Publications	1.93	2.00	1.53	0.19	0.00	0.62	−1.74	0.000***
Non Jax Control Publications	0.08	0.00	0.27	1.34	1.00	1.33	1.26	0.000***
Affiliation Rank	162.21	104.00	163.30	181.49	132.00	173.75	19.28	0.307
Top 5 Percent Publications	1.92	0.00	4.56	2.21	0.00	4.82	0.29	0.584
Share Last Author Publ.	0.23	0.17	0.23	0.25	0.20	0.24	0.02	0.444
Share Last Author Mice Publ.	0.20	0.13	0.25	0.21	0.12	0.25	0.01	0.734
Distinct Coauthors	10.79	6.70	10.44	10.48	6.60	12.07	−0.31	0.804
Distinct New Coauthors	6.43	4.20	6.15	6.16	3.90	7.55	−0.27	0.731
Mean Self-Reference Count	0.58	0.28	0.82	0.55	0.35	0.66	−0.03	0.702
Mean New Keyword Count	0.15	0.00	0.33	0.19	0.00	0.34	0.04	0.288
Clinical Journal Publ.	4.72	2.00	7.57	5.46	3.00	7.61	0.74	0.385
Patent Weighted Publ. (10yrs)	1.35	0.00	2.76	1.54	0.00	3.68	0.19	0.606
Mean New MeSH Count	66.49	54.50	43.53	65.40	60.50	37.80	−1.09	0.812
Mean New Disease MeSH	7.26	5.00	8.21	6.68	4.00	7.35	−0.57	0.509
In Vitro related Publications	4.44	2.00	5.44	4.19	2.00	6.41	−0.24	0.714
Other animal related Publications	2.88	0.50	8.54	4.83	2.00	8.39	1.94	0.041**
Human related Publications	8.04	3.00	12.33	6.38	2.00	12.50	−1.67	0.230
Diversity MeSH	390.89	325.50	281.98	410.17	355.50	289.51	19.28	0.547

Notes: The table compares pre-fire scientists characteristics by treatment group with t-tests and displays the mean, median and the standard errors for each group using sample USA. If not stated otherwise, observations are aggregated in the period 1984-1988 (5 years pre-fire). The unit of observation is the individual scientist. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

A.5.2 Tables – Parallel Trends

Table A-11: Test parallel trends assumption

Sample A Pre-Period: 1983-1988	(1) Publ	(2) Mice Publ	(3) B6	(4) Affected
86-88 × affected	0.003	−0.004	−0.158	0.146
	(0.056)	(0.068)	(0.134)	(0.130)
Scientist FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes
Observations	3396	3390	2268	2082
Scientists	566	565	378	347
Log-likelihood	−6412	−4903	−2010	−1819
Sample USA Pre-Period: 1983-1988	(1) Publ	(2) Mice Publ	(3) B6	(4) Affected
86-88 × affected	−0.046	−0.008	−0.194	0.187
	(0.067)	(0.081)	(0.187)	(0.162)
Scientist FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes
Observations	1920	1914	1188	1266
Scientists	320	319	198	211
Log-likelihood	−3667	−2787	−1004	−1099

Notes: Columns (1) to (4) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The time period of these regressions is the pre-fire period 1983-1988. We employ a placebo test (1986-1988 x affected scientists with 1983-1985 as the baseline period) to test the parallel trends assumption. The outcome variable is the annual number of all/mice/B6/affected mice publications. The top table shows the estimates for sample A, and the bottom table the estimates for sample USA. The unit of observation is the individual scientist by year. Standard errors are clustered at the scientist level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

A.5.3 Tables – Adoption of Mice

Table A-12: Adoption of mice – alternative dependent variables

Share/Linear +5 years	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Sample A				Sample USA			
	B6	Affected	B6 (JIF)	Affected (JIF)	B6	Affected	B6 (JIF)	Affected (JIF)
Post × affected	0.031**	−0.017	0.024	−0.006	0.016	−0.004	0.013	−0.013
	(0.015)	(0.016)	(0.015)	(0.016)	(0.018)	(0.020)	(0.018)	(0.020)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6792	6792	6792	6792	3840	3840	3840	3840
Scientists	566	566	566	566	320	320	320	320
Log-likelihood	−262	−570	−511	−682	138	−357	23	−461

Count/PPML +10 years	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Sample A				Sample USA			
	B6	Affected	B6 (JIF)	Affected (JIF)	B6	Affected	B6 (JIF)	Affected (JIF)
Post × affected	0.299**	−0.195	0.002	−0.293	0.203	−0.174	−0.165	−0.496*
	(0.132)	(0.155)	(0.200)	(0.219)	(0.187)	(0.213)	(0.267)	(0.269)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7616	7480	7514	7174	4216	4386	4131	4250
Scientists	448	440	442	422	248	258	243	250
Log-likelihood	−5804	−5593	−17251	−15234	−3166	−3328	−10286	−10150

Log+1/Linear +10 years	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Sample A				Sample USA			
	B6	Affected	B6 (JIF)	Affected (JIF)	B6	Affected	B6 (JIF)	Affected (JIF)
Post × affected	0.065***	−0.047**	0.049	−0.064*	0.034	−0.047	−0.002	−0.125**
	(0.022)	(0.022)	(0.038)	(0.038)	(0.028)	(0.030)	(0.052)	(0.056)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9622	9622	9622	9622	5440	5440	5440	5440
Scientists	566	566	566	566	320	320	320	320
Log-likelihood	−2962	−2733	−8923	−8433	−1627	−1664	−5233	−5221

Notes: In the top part of this table, Columns (1) to (8) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The outcome variable is the share of (JIF-weighted) B6/affected mice publications. In the middle/bottom part of this table, Columns (1) to (8) show the estimates of linear regressions with high-dimensional fixed effects. The outcome variable is the (log-transformed) number of B6/Affected mice related publications. The unit of observation is the individual scientist by year. Standard errors are clustered at the scientist level and shown in parentheses. Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

A. APPENDIX TO CHAPTER 1

Table A-13: Adoption of mice – alternative specifications

Sample A Share/Linear	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				+10 years				
DV: Share B6 (JIF)		Age FE		Cluster Level		Treatment Variable		
Post × affected	0.040***	0.040***	0.040***	0.040**	0.040			
	(0.015)	(0.015)	(0.015)	(0.019)	(0.045)			
× affected (cont.)						0.046***		
						(0.017)		
× affected (mice publ)							0.015	
							(0.015)	
× affected (all publ)								0.034**
								(0.015)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age (squared)	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age (3rd poly)	No	No	No	Yes	Yes	Yes	Yes	Yes
Observations	9622	9622	9622	9622	9622	9622	9622	9622
Scientists	566	566	566	566	566	566	566	566
Cluster	Scientist	Scientist	Scientist	Sc. × Year	Strain	Scientist	Scientist	Scientist
Log-likelihood	−679	−679	−676	−676	−676	−676	−682	−678

Sample USA Share/Linear	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				+10 years				
DV: Share B6 (JIF)		Age FE		Cluster Level		Treatment Variable		
Post × affected	0.020	0.020	0.020	0.020	0.020			
	(0.018)	(0.018)	(0.018)	(0.019)	(0.039)			
× affected (cont.)						0.028		
						(0.019)		
× affected (mice publ)							0.005	
							(0.018)	
× affected (all publ)								0.022
								(0.018)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age (squared)	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age (3rd poly)	No	No	No	Yes	Yes	Yes	Yes	Yes
Observations	5440	5440	5440	5440	5440	5440	5440	5440
Scientists	320	320	320	320	320	320	320	320
Cluster	Scientist	Scientist	Scientist	Sc. × Year	Strain	Scientist	Scientist	Scientist
Log-likelihood	8	10	10	11	11	11	10	11

Notes: Columns (1) to (8) show the estimates of linear regressions with high-dimensional fixed effects (the top table uses sample A, the bottom table uses sample USA). The outcome variable is the share of B6 mice publications. Treatment is defined binary based on the pre-fire share of treated mice publications relative the all treated & control mice publications in Columns (1) to (5), relative to all mice publications in Column (7), and relative to all publications in Column (8). Treatment is defined continuously in Column (6). The unit of observation is the individual scientist by year. Standard errors are shown in parentheses. They are clustered at the scientist level in Columns (1) to (3) as well as (6) to (8), at the scientist times year level in Column (4), and at the mice strain level (most often used by the scientists) in Column (5). Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

A.5.4 Tables – Research Output

Table A-14: Research output – alternative dependent variables

Count/PPML +5 years	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Sample A				Sample USA/Canada			
	Publ	Mice	Publ (JIF)	Mice (JIF)	Publ	Mice	Publ (JIF)	Mice (JIF)
Post × affected	−0.052	0.040	−0.059	−0.041	−0.139**	−0.002	−0.137	−0.120
	(0.059)	(0.068)	(0.078)	(0.104)	(0.071)	(0.085)	(0.093)	(0.125)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6792	6792	6792	6792	3840	3840	3840	3840
Scientists	566	566	566	566	320	320	320	320
Log-likelihood	−14132	−10168	−35208	−28197	−8035	−5807	−22811	−17972

Count/Linear +10 years	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Sample A				Sample USA			
	Publ	Mice	Publ (JIF)	Mice (JIF)	Publ	Mice	Publ (JIF)	Mice (JIF)
Post × affected	−0.232	0.051	−0.538	−0.185	−0.829**	−0.039	−2.288	−0.871
	(0.302)	(0.131)	(1.238)	(0.810)	(0.364)	(0.185)	(1.781)	(1.250)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9622	9622	9622	9622	5440	5440	5440	5440
Scientists	566	566	566	566	320	320	320	320
Log-likelihood	−24537	−18367	−40564	−37791	−13734	−10719	−23737	−22213

Log+1/Linear +10 years	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Sample A				Sample USA			
	Publ	Mice	Publ (JIF)	Mice (JIF)	Publ	Mice	Publ (JIF)	Mice (JIF)
Post × affected	−0.018	0.025	0.013	0.025	−0.092	−0.001	−0.093	−0.080
	(0.047)	(0.039)	(0.065)	(0.060)	(0.061)	(0.051)	(0.093)	(0.085)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9622	9622	9622	9622	5440	5440	5440	5440
Scientists	566	566	566	566	320	320	320	320
Log-likelihood	−7738	−6958	−11753	−11856	−4307	−3957	−6826	−6976

Notes: In the top part of this table, Columns (1) to (8) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. In the middle/bottom part of this table, Columns (1) to (8) show the estimates of linear regressions with high-dimensional fixed effects. In the top/middle tables, the outcome variable is the number of (JIF-weighted) publications/mice publications, in the bottom table the outcome variable is log-transformed $\ln(+1)$. The unit of observation is the individual scientist by year. Standard errors are clustered at the scientist level and shown in parentheses. Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A-15: Research output – alternative specifications

Sample A Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				+10 years				
DV: Publ (JIF)		Age FE		Cluster Level			Treatment Variable	
Post × affected	−0.030	−0.030	−0.035	−0.039	−0.039			
	(0.089)	(0.089)	(0.089)	(0.104)	(0.065)			
× affected (cont.)						0.017		
						(0.094)		
× affected (mice publ)							−0.061	
							(0.091)	
× affected (all publ)								−0.038
								(0.088)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age (squared)	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age (3rd poly)	No	No	No	Yes	Yes	Yes	Yes	Yes
Observations	9622	9622	9622	9622	9622	9622	9622	9622
Scientists	566	566	566	566	566	566	566	566
Cluster	Scientist	Scientist	Scientist	Sc. × Year	Strain	Scientist	Scientist	Scientist
Log-likelihood	−51599	−51599	−51296	−51237	−51237	−51242	−51231	−51238

Sample USA Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				+10 years				
DV: Publ (JIF)		Age FE		Cluster Level			Treatment Variable	
Post × affected	−0.131	−0.134	−0.151	−0.151	−0.151**			
	(0.105)	(0.106)	(0.103)	(0.122)	(0.064)			
× affected (cont.)						−0.116		
						(0.102)		
× affected (mice publ)							−0.110	
							(0.112)	
× affected (all publ)								−0.074
								(0.109)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age (squared)	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age (3rd poly)	No	No	No	Yes	Yes	Yes	Yes	Yes
Observations	5440	5440	5440	5440	5440	5440	5440	5440
Scientists	320	320	320	320	320	320	320	320
Cluster	Scientist	Scientist	Scientist	Sc. × Year	Strain	Scientist	Scientist	Scientist
Log-likelihood	−33406	−33384	−32986	−32981	−32981	−33015	−33013	−33029

Notes: Columns (1) to (8) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects (the top table uses sample A, the bottom table uses sample USA). The outcome variable is the JIF-weighted number of publications. Treatment is defined binary based on the pre-fire share of treated mice publications relative the all treated & control mice publications in Columns (1) to (5), relative to all mice publications in Column (7), and relative to all publications in Column (8). Treatment is defined continuously in column (6). The unit of observation is the individual scientist by year. Standard errors are shown in parentheses. They are clustered at the scientist level in Columns (1) to (3) as well as (6) to (8), at the scientist times year level in Column (4), and at the mice strain level (most often used by the scientists) in Column (5). Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A-16: Research output – forward citation-weighted

Sample A +10 years	(1)	(2) Count/PPML	(3)	(4)	(5) Mean/PPML	(6)
	Citations All	Citations Mice	Citations B6	Citations All	Citations Mice	Citations B6
Post × affected	−0.109	−0.054	0.336	0.043	0.040	0.350
	(0.143)	(0.158)	(0.253)	(0.080)	(0.107)	(0.216)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9622	9622	7616	9622	9622	7616
Scientists	566	566	448	566	566	448
Log-likelihood	−1027444	−722581	−273824	−226919	−257519	−171279

Sample USA +10 years	(1)	(2) Count/PPML	(3)	(4)	(5) Mean/PPML	(6)
	Citations All	Citations Mice	Citations B6	Citations All	Citations Mice	Citations B6
Post × affected	−0.240	−0.037	0.306	0.045	0.004	0.320
	(0.162)	(0.192)	(0.326)	(0.101)	(0.131)	(0.294)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5440	5440	4216	5440	5440	4216
Scientists	320	320	248	320	320	248
Log-likelihood	−667977	−469538	−164438	−138378	−154737	−100227

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The outcome variable in Columns (1) to (3) is the lifetime citation weighted number of publications per scientists per year. The outcome variable in Columns (4) to (6) is the average number of lifetime citations that a scientist received for her publications in a given year. The top part of this table uses sample A, the bottom part uses sample USA. The unit of observation is the individual scientist by year. Standard errors are clustered at the scientist level and shown in parentheses. Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A-17: Research output – organizational indicator-weighted

Sample A +10 years	(1) Count/PPML	(2)	(3) Mean/PPML	(4) Share/Linear	(5)	(6)	(7)
	Coauth.	New Coauth.	Mean Coauth.	Last	Last Mice	First	First Mice
Post × affected	−0.063	−0.054	0.004	−0.005	−0.002	0.004	0.000
	(0.073)	(0.076)	(0.042)	(0.017)	(0.018)	(0.016)	(0.015)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9622	9622	9622	9622	9622	9622	9622
Scientists	566	566	566	566	566	566	566
Log-likelihood	−41650	−34241	−19479	−456	−1355	−683	−681

Sample USA +10 years	(1) Count/PPML	(2)	(3) Mean/PPML	(4) Share/Linear	(5)	(6)	(7)
	Coauth.	New Coauth.	Mean Coauth.	Last	Last Mice	First	First Mice
Post × affected	−0.217***	−0.239***	0.036	−0.001	0.008	0.001	−0.011
	(0.084)	(0.087)	(0.053)	(0.022)	(0.023)	(0.021)	(0.019)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5440	5440	5440	5440	5440	5440	5440
Scientists	320	320	320	320	320	320	320
Log-likelihood	−23517	−19654	−10887	−158	−747	−406	−300

Notes: Columns (1) to (3) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The outcome variable in Column (1) to (2) is the number of distinct (new) coauthors in each year. The outcome variable in Column (3) is the average number of coauthors per publication in each year. Columns (4) to (7) show the estimates of linear regressions with high-dimensional fixed effects. The outcome variable is the share of last/first authored (mice) publications in each year. The top part of this table uses sample A, the bottom part uses sample USA. The unit of observation is the individual scientist by year. Standard errors are clustered at the scientist level and shown in parentheses. Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

A.5.5 Tables – Type of Research

Table A-18: Type of research – all keywords (incl. old keywords)

Mean/PPML +10 years	(1) References	(2)	(3) Keywords	(4)	(5) MeSH	(6)	(7) Disease MeSH	(8)
	All	USA	All	USA	All	USA	All	USA
Post × affected	−0.031	−0.028	0.043	0.188	0.002	0.013	0.026	0.042
	(0.072)	(0.095)	(0.156)	(0.216)	(0.036)	(0.048)	(0.066)	(0.091)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9622	5440	8738	4981	9622	5440	9469	5355
Scientists	566	320	514	293	566	320	557	315
Log-likelihood	−64296	−37089	−9015	−4843	−33996	−18986	−7995	−4566

Notes: Columns (1) to (8) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The outcome variable is the annual average number of all-references in Columns (1) and (2), all keywords in Columns (3) and (4), all MeSH terms in Columns (5) and (6), and all disease related MeSH terms in Columns (7) and (8). Columns with odd numbers show the estimates for sample A and columns with even numbers show the estimated for sample USA. The unit of observation is the individual scientist by year. Standard errors are clustered at the scientist level and shown in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A-19: Type of research – MeSH terms related to alternative research tools

Share/Linear +10 years	(1)	(2)	(3)	(4)	(5)	(6)
		Sample All			Sample USA	
	In Vitro	Animal	Human	In Vitro	Animal	Human
Post × affected	0.013	0.009	−0.013	0.013	0.015	0.005
	(0.013)	(0.012)	(0.016)	(0.013)	(0.017)	(0.022)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9622	9622	9622	9622	5440	5440
Scientists	566	566	566	566	320	320
Log-likelihood	1134	2368	−694	1134	1276	−413

Notes: Columns (1) to (6) show the estimates of linear regressions with high-dimensional fixed effects. The outcome variable is the share of "In Vitro" (defined by MeSH terms) related publications in Column (1) and (4), other non-mice but "animal" (defined by MeSH terms) related publications in Column (2) and (5), and "Human" (defined by MeSH terms) related publications in Column (3) and (6). Columns (1) to (3) show the estimates for sample A and Columns (4) to (6) show the estimated for sample USA. The unit of observation is the individual scientist by year. Standard errors are clustered at the scientist level and shown in parentheses. Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A-20: Type of research – alternative dependent variables

Count/PPML +10 years	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Self-References		New Keywords		New MeSH		New Disease MeSH	
	All	USA	All	USA	All	USA	All	USA
Post × affected	−0.289**	−0.449***	0.132	0.100	−0.063	−0.166**	−0.026	−0.125
	(0.132)	(0.141)	(0.150)	(0.202)	(0.055)	(0.068)	(0.089)	(0.113)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9129	5270	8738	4981	9622	5440	9452	5338
Scientists	537	310	514	293	566	320	556	314
Log-likelihood	−32736	−19567	−23637	−13091	−53726	−30386	−13917	−8059

Count/PPML +10 years	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Sample All: Clinical Y Publ				Sample USA: Clinical Y Publ			
	Y = All	Y = Mice	Y = B6	Y = Affectd	Y = All	Y = Mice	Y = B6	Y = Affectd
Post × affected	−0.026	0.147	0.660***	−0.089	−0.244**	0.058	0.663***	−0.128
	(0.104)	(0.111)	(0.187)	(0.235)	(0.109)	(0.141)	(0.244)	(0.329)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9180	8330	4709	4386	5168	4743	2550	2601
Scientists	540	490	277	258	304	279	150	153
Log-likelihood	−11174	−6483	−2141	−2083	−6174	−3606	−1101	−1181

Count/PPML +10 years	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Sample All: Patent-Weighted Y Publ				Sample USA: Patent-Weighted Y Publ			
	Y = All	Y = Mice	Y = B6	Y = Affectd	Y = All	Y = Mice	Y = B6	Y = Affectd
Post × affected	0.197	0.268	0.953***	−0.199	0.159	0.211	0.898*	−0.160
	(0.144)	(0.166)	(0.368)	(0.345)	(0.176)	(0.194)	(0.538)	(0.414)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6936	5593	2278	2771	4267	3349	1360	1887
Scientists	408	329	134	163	251	197	80	111
Log-likelihood	−4935	−3311	−775	−1117	−3186	−2113	−456	−784

Notes: Columns (1) to (8) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. In the top part of this table, the outcome variable is the number of self-references in Columns (1/2), new keywords in Columns (3/4), new MeSH terms in Columns (5/6), and new disease related MeSH terms in Columns (7/8). In the middle part of this table, the outcome variable is the share of all/mice/B6/affected mice publications published in a journal that usually publishes more clinical relevant research relative to all/mice/B6/affected mice publications. In the bottom part of the table, the outcome variable is the share of all/mice/B6/affected mice publications that are associated with a patent application (patent-weighted) relative to all/mice/B6/affected mice unweighted publications. A patent-weight is calculated based on the patent family's first application being filed within 10 years from the scientific publication. The unit of observation is the individual scientist by year. Standard errors are clustered at the scientist level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A-21: Type of research – alternative specifications

Sample A Share/Linear	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				+10	years			
DV: Patent-Weighted (10yrs)		Age FE		Cluster	Level	Treatment		Variable
Post × affected	0.024***	0.025***	0.024***	0.024***	0.024***			
	(0.007)	(0.007)	(0.007)	(0.008)	(0.006)			
× affected (cont.)						0.026***		
						(0.008)		
× affected (mice publ)							0.017**	
							(0.007)	
× affected (all publ)								0.018**
								(0.008)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age (squared)	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age (3rd poly)	No	No	No	Yes	Yes	Yes	Yes	Yes
Observations	9622	9622	9622	9622	9622	9622	9622	9622
Scientists	566	566	566	566	566	566	566	566
Cluster	Scientist	Scientist	Scientist	Sc. × Year	Strain	Scientist	Scientist	Scientist
Log-likelihood	5130	5130	5132	5132	5132	5132	5128	5129

Sample USA Share/Linear	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				+10	years			
DV: Patent-Weighted (10yrs)		Age FE		Cluster	Level	Treatment		Variable
Post × affected	0.039***	0.039***	0.039***	0.039***	0.039***			
	(0.011)	(0.011)	(0.011)	(0.010)	(0.012)			
× affected (cont.)						0.038***		
						(0.012)		
× affected (mice publ)							0.025**	
							(0.011)	
× affected (all publ)								0.021*
								(0.011)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age (squared)	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age (3rd poly)	No	No	No	Yes	Yes	Yes	Yes	Yes
Observations	5440	5440	5440	5440	5440	5440	5440	5440
Scientists	320	320	320	320	320	320	320	320
Cluster	Scientist	Scientist	Scientist	Sc. × Year	Strain	Scientist	Scientist	Scientist
Log-likelihood	2541	2541	2542	2542	2542	2540	2536	2535

Notes: Columns (1) to (8) show the estimates of linear regressions with high-dimensional fixed effects (the top table uses sample A, bottom table uses sample USA). The outcome variable is the share of patent application-weighted publications (within 10 years from publication) in each year. Treatment is defined binary based on the pre-fire share of treated mice publications relative the all treated & control mice publications in Columns (1) to (5), relative to all mice publications in Column (7), and relative to all publications in Column (8). Treatment is defined continuously in Column (6). The unit of observation is the individual scientist by year. Standard errors are shown in parentheses. They are clustered at the scientist level in Columns (1) to (3) as well as (6) to (8), at the scientist times year level in Column (4), and at the mice strain level (most often used by the scientists) in Column (5). Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

A.5.6 Tables – Alternative Classification of Mice

Table A-22: Alternative classification of mice publications

Sample A Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Strain Types		Affected Strains		Spared Strains	
+10 years	Inbred	Transgenic	Mutant	Jax	Non Jax	Jax	Non Jax
Post × affected	−0.127	−0.161	0.392	−1.669***	−1.564***	2.198***	1.899***
	(0.139)	(0.445)	(0.330)	(0.369)	(0.298)	(0.482)	(0.293)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9282	2772	5406	3332	5916	3995	4675
Scientists	546	198	318	196	348	235	275
Log-likelihood	−27709	−10004	−13454	−4051	−7979	−6194	−6064
Sample USA Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Strain Types		Affected Strains		Spared Strains	
+10 years	Inbred	Transgenic	Mutant	Jax	Non Jax	Jax	Non Jax
Post × affected	−0.332*	0.000	−0.067	−1.683***	−2.074***	1.685***	1.534***
	(0.178)	(0.506)	(0.425)	(0.416)	(0.403)	(0.527)	(0.336)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5253	1764	3162	2261	3247	2227	2890
Scientists	309	126	186	133	191	131	170
Log-likelihood	−17296	−6990	−8434	−2905	−4686	−3735	−4157

Notes: Columns (1) to (7) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. In Columns (1) to (3), the outcome variable is the number of publications that include an inbred, transgenic or mutant mice MeSH term. In Columns (4) to (7), the outcome variable is the number of publications that mention a certain affected or spared Jax/non Jax mice strain. The top part of this table uses sample A, the bottom part uses sample B. The unit of observation is the individual scientist by year. Standard errors are clustered at the scientist level and shown in parentheses. Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

A.5.7 Tables – Alternative Counts

Table A-23: Scientific *articles* with fewer than 20 authors

Sample A +10 years	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Count/PPML				Share/Linear		
DV: Articles	All	All (JIF)	Mice	Mice (JIF)	B6	B6 (JIF)	Affected	Affected (JIF)
Post × affected	−0.044	−0.019	0.030	−0.005	0.048***	0.042***	−0.035**	−0.024*
	(0.068)	(0.086)	(0.079)	(0.108)	(0.015)	(0.016)	(0.015)	(0.015)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9622	9622	9622	9622	9622	9622	9622	9622
Scientists	566	566	566	566	566	566	566	566
Log-likelihood	−20028	−48685	−14107	−40074	−434	−738	−528	−708

Sample USA +10 years	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Count/PPML				Share/Linear		
DV: Articles	All	All (JIF)	Mice	Mice (JIF)	B6	B6 (JIF)	Affected	Affected (JIF)
Post × affected	−0.186**	−0.137	−0.018	−0.099	0.026	0.021	−0.030	−0.038*
	(0.079)	(0.100)	(0.103)	(0.134)	(0.018)	(0.018)	(0.019)	(0.020)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5440	5440	5440	5440	5440	5440	5440	5440
Scientists	320	320	320	320	320	320	320	320
Log-likelihood	−11425	−31047	−8152	−25707	104	−29	−383	−524

Notes: Columns (1) to (4) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The outcome variable is the number of (mice related) articles in scientific journals (JIF weighted) with fewer than 20 coauthors. Columns (5) to (8) show the estimates of linear regressions with high-dimensional fixed effects. The outcome variable is the share of B6/affected mice related articles in scientific journals (JIF weighted) with fewer than 20 coauthors relative to all mice related articles in scientific journals (JIF weighted). The top part of this table uses sample A, the bottom part uses sample B. The unit of observation is the individual scientist by year. Standard errors are clustered at the scientist level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A-24: Research output weighted by inverse of coauthor counts

Sample A +10 years	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Count/PPML				Share/Linear		
DV: Fractions	All	All (JIF)	Mice	Mice (JIF)	B6	B6 (JIF)	Affected	Affected (JIF)
Post × affected	−0.086	−0.076	0.050	0.021	0.046***	0.039**	−0.037**	−0.027*
	(0.068)	(0.091)	(0.082)	(0.109)	(0.015)	(0.015)	(0.015)	(0.015)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9622	9622	9622	9622	9622	9622	9622	9622
Scientists	566	566	566	566	566	566	566	566
Log-likelihood	−10460	−19578	−6424	−13914	−378	−646	−551	−687

Sample USA +10 years	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Count/PPML				Share/Linear		
DV: Fractions	All	All (JIF)	Mice	Mice (JIF)	B6	B6 (JIF)	Affected	Affected (JIF)
Post × affected	−0.186**	−0.163	0.004	−0.073	0.024	0.019	−0.031	−0.038**
	(0.080)	(0.104)	(0.103)	(0.135)	(0.018)	(0.018)	(0.019)	(0.019)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5440	5440	5440	5440	5440	5440	5440	5440
Scientists	320	320	320	320	320	320	320	320
Log-likelihood	−6059	−12525	−3722	−8759	143	30	−378	−506

Notes: Columns (1) to (4) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The outcome variable is the number of (mice related) publications (JIF weighted) divided by the number of coauthors (fractional count). Columns (5) to (8) show the estimates of linear regressions with high-dimensional fixed effects. The outcome variable is the fractional share of B6/affected mice related publications (JIF weighted) to all mice related publications (fractional/JIF weighted). The top part of this table uses sample A, the bottom part uses sample B. The unit of observation is the individual scientist by year. Standard errors are clustered at the scientist level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

A.5.8   Tables – Alternative Samples

Table A-25: Alternative sample B

Sample B +10 years	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Research Output			Impact/Organization		Trajectories	
	All	Mice	B6	Affected	Citations	Coauthors	Clinical	Patents
Post × affected	−0.107	−0.007	0.051**	−0.033*	−0.007	−0.140	0.022	0.023***
	(0.109)	(0.130)	(0.020)	(0.018)	(0.097)	(0.087)	(0.019)	(0.009)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6358	6358	6358	6358	6358	6358	6358	6358
Scientists	374	374	374	374	374	374	374	374
Log-likelihood	−33097	−26399	−497	−355	−154732	−26574	−522	3429

Sample B +5 years	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Research Output			Impact/Organization		Trajectories	
	All	Mice	B6	Affected	Citations	Coauthors	Clinical	Patents
Post × affected	−0.068	0.036	0.038*	−0.014	−0.066	−0.117	0.015	0.018**
	(0.094)	(0.121)	(0.020)	(0.019)	(0.118)	(0.077)	(0.019)	(0.009)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4488	4488	4488	4488	4488	4488	4488	4488
Scientists	374	374	374	374	374	374	374	374
Log-likelihood	−22665	−18278	−371	−390	−93769	−16681	−317	2795

Notes: This table shows a summary of the regressions results using sample B, which is more strict concerning age and scientific productivity. Columns (1), (2), (5) and (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The outcome variables are the number of (mice related) publications (JIF weighted), mean number of citations, and number of distinct coauthors. Columns (3), (4), (7) and (8) show the estimates of linear regressions with high-dimensional fixed effects. The outcome variables are the share of (no) B6 publications (JIF weighted), share of clinical journal publications, and the share of patent-weighted publications. The unit of observation is the individual scientist by year. Standard errors are clustered at the scientist level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

A. APPENDIX TO CHAPTER 1

Table A-26: Alternative sample C

Sample C +10 years	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Research Output			Impact/Organization		Trajectories	
	All	Mice	B6	Affected	Citations	Coauthors	Clinical	Patents
Post × affected	−0.016	−0.126	0.040***	−0.074***	0.092*	−0.018	0.016*	0.015***
	(0.059)	(0.088)	(0.010)	(0.014)	(0.048)	(0.046)	(0.009)	(0.004)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	39661	39627	39661	39661	39661	39661	39661	39661
Scientists	2333	2331	2333	2333	2333	2333	2333	2333
Log-likelihood	−208570	−165362	−469	−19079	−839063	−174336	−2296	20584

Sample D +5 years	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Research Output			Impact/Organization		Trajectories	
	All	Mice	B6	Affected	Citations	Coauthors	Clinical	Patents
Post × affected	−0.038	−0.125	0.023**	−0.057***	0.047	−0.048	0.005	0.013***
	(0.053)	(0.077)	(0.010)	(0.015)	(0.051)	(0.039)	(0.010)	(0.004)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	27996	27972	27996	27996	27996	27984	27996	27996
Scientists	2333	2331	2333	2333	2333	2332	2333	2333
Log-likelihood	−141886	−114373	−101	−13627	−517145	−108763	−1303	16997

Notes: This table shows a summary of the regressions results using the pre-matching sample C. Columns (1), (2), (5) and (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The outcome variables are the number of (mice related) publications (JIF weighted), mean number of citations, and number of distinct coauthors. Columns (3), (4), (7) and (8) show the estimates of linear regressions with high-dimensional fixed effects. The outcome variables are the share of (no) B6 publications (JIF weighted), share of clinical journal publications, and the share of patent-weighted publications. The unit of observation is the individual scientist by year. Standard errors are clustered at the scientist level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

A.5.9 Tables – Heterogeneity

Table A-27: Heterogeneity by mice dependence

Sample A +10 years	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Research Output				Impact/Organization		Trajectories	
	All	Mice	B6	Affected	Citations	Coauthors	Clinical	Patents
Post × affected	0.072	0.311	0.080***	−0.029	0.067	0.005	0.038	0.014*
	(0.157)	(0.212)	(0.022)	(0.021)	(0.121)	(0.122)	(0.025)	(0.008)
× Mice Publ.	−0.127	−0.372	−0.078**	0.004	−0.029	−0.095	−0.015	0.021
	(0.186)	(0.244)	(0.030)	(0.029)	(0.157)	(0.152)	(0.031)	(0.014)
Post × Mice Publ.	−0.154	−0.180	0.085***	−0.030*	−0.071	0.086	0.056***	0.018*
	(0.133)	(0.196)	(0.024)	(0.017)	(0.112)	(0.104)	(0.021)	(0.010)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9622	9622	9622	9622	9622	9622	9622	9622
Scientists	566	566	566	566	566	566	566	566
Log-likelihood	−51153	−40815	−662	−676	−226854	−41641	−869	5144

Notes: This table shows a summary of the regressions results following Equation 1.2 with a triple-interaction term. Columns (1), (2), (5) and (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The outcome variables are the number of (mice related) publications (JIF weighted), mean number of citations, and number of distinct coauthors. Columns (3), (4), (7) and (8) show the estimates of linear regressions with high-dimensional fixed effects. The outcome variables are the share of (no) B6 publications (JIF weighted), share of clinical journal publications, and the share of patent-weighted publications. The triple-interaction term is a scientists JIF-weighted mice publication record in the 5 years before the fire. The unit of observation is the individual scientist by year. Standard errors are clustered at the scientist level and shown in parentheses. Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A-28: Heterogeneity by age

Sample A +10 years	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Research Output				Impact/Organization		Trajectories	
	All	Mice	B6	Affected	Citations	Coauthors	Clinical	Patents
Post × affected	−0.048	0.063	0.046**	−0.023	0.074	−0.209*	0.019	0.019*
	(0.133)	(0.158)	(0.021)	(0.019)	(0.138)	(0.116)	(0.022)	(0.011)
× Experience	0.014	−0.134	−0.012	−0.008	−0.053	0.215	0.024	0.012
	(0.174)	(0.215)	(0.031)	(0.029)	(0.166)	(0.148)	(0.031)	(0.015)
Post × Experience	0.046	0.221	0.012	0.014	0.000	−0.097	0.003	−0.017
	(0.132)	(0.183)	(0.031)	(0.023)	(0.133)	(0.107)	(0.025)	(0.013)
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9622	9622	9622	9622	9622	9622	9622	9622
Scientists	566	566	566	566	566	566	566	566
Log-likelihood	−51233	−40896	−676	−679	−226908	−41621	−877	5133

Notes: This table shows a summary of the regressions results following Equation 1.2 with a triple-interaction term. Columns (1), (2), (5) and (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The outcome variables are the number of (mice related) publications (JIF weighted), mean number of citations, and number of distinct coauthors. Columns (3), (4), (7) and (8) show the estimates of linear regressions with high-dimensional fixed effects. The outcome variables are the share of (no) B6 publications (JIF weighted), share of clinical journal publications, and the share of patent-weighted publications. The triple-interaction term is the age (since PhD) of a scientists at the time of the fire. The unit of observation is the individual scientist by year. Standard errors are clustered at the scientist level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

B

Appendix to Chapter 2

The Innovation Effect of Drug Approval

B.1 Patent Term Extensions

Figure B-1: Patent term extensions by SPC
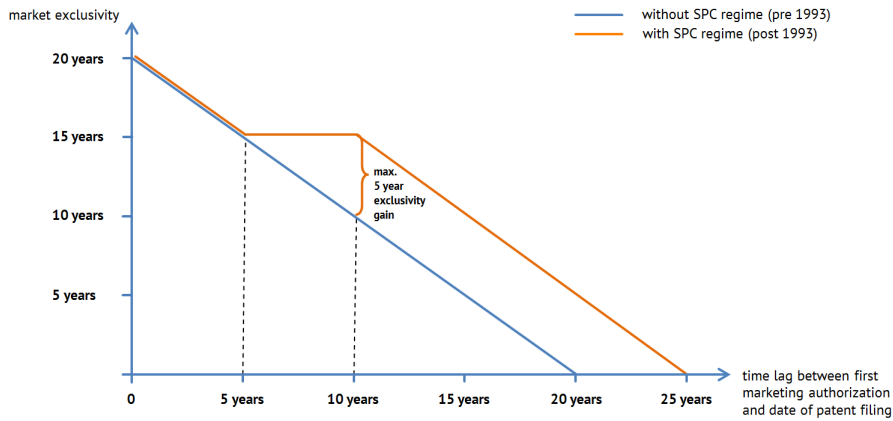
Notes: The figure illustrates the EU SPC regime introduced in 1993. All patents in the sample were retrieved from the SPC data and are, thus, subject to the SPC regime. The patent term extension is capped at 5 years so that market exclusivity can be seen as constant for patents with a approval-lag between 5 and 10 years.

Figure B-2: Distribution – expected market exclusivity
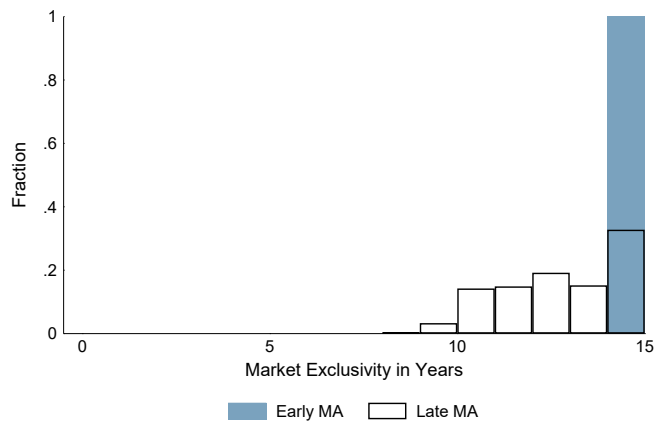
Notes: The figure presents the distribution the expected market exclusivity, which is calculated from first EU marketing authorization until patent expiry account for the patent term extension granted by SPC. The sample of 590 patent-drug links is split in two equally sized groups, using as threshold the median of the approval lag.

B.2 Figures

B.2.1 Figures – Descriptive Analysis

Figure B-3: Evolution of various citation counts by *time to approval*
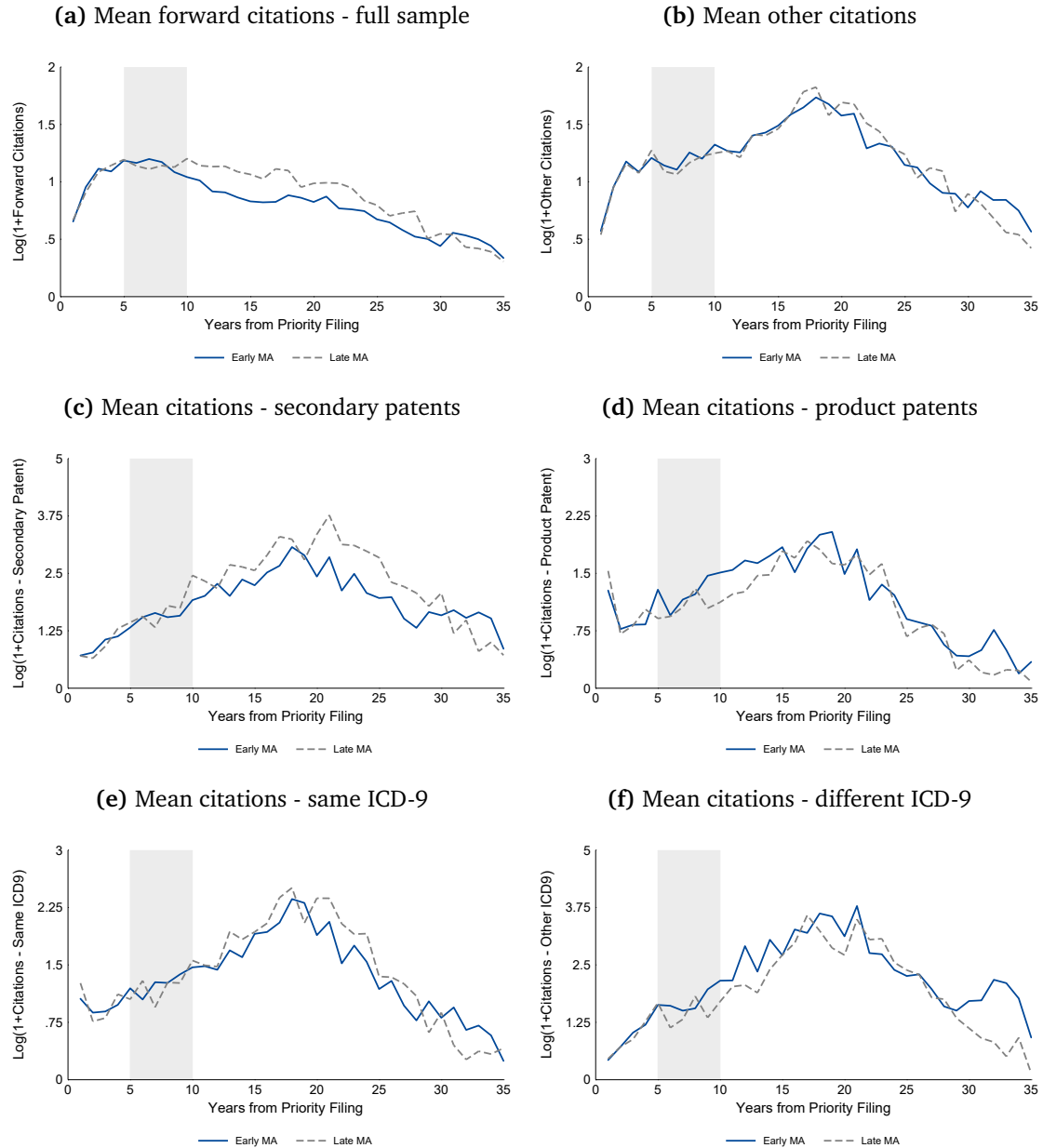
(a) Mean forward citations - full sample

(b) Mean other citations

(c) Mean citations - secondary patents

(d) Mean citations - product patents

(e) Mean citations - same ICD-9

(f) Mean citations - different ICD-9

Notes: The figures present the average log-transformed number of forward citations split at the median approval lag (early/late MA) over time. The annual citation count is normalized by the average citation count from the pre-approval period t_0-t_5. Figure (a) includes the "full" sample with all 1,405 SPC filings. The other figures are based on the "main" sample of 590 patent-drug links. The unit of observation is the unique patent-drug level.

Figure B-4: Evolution of various citation counts by *time to approval* (cont.)
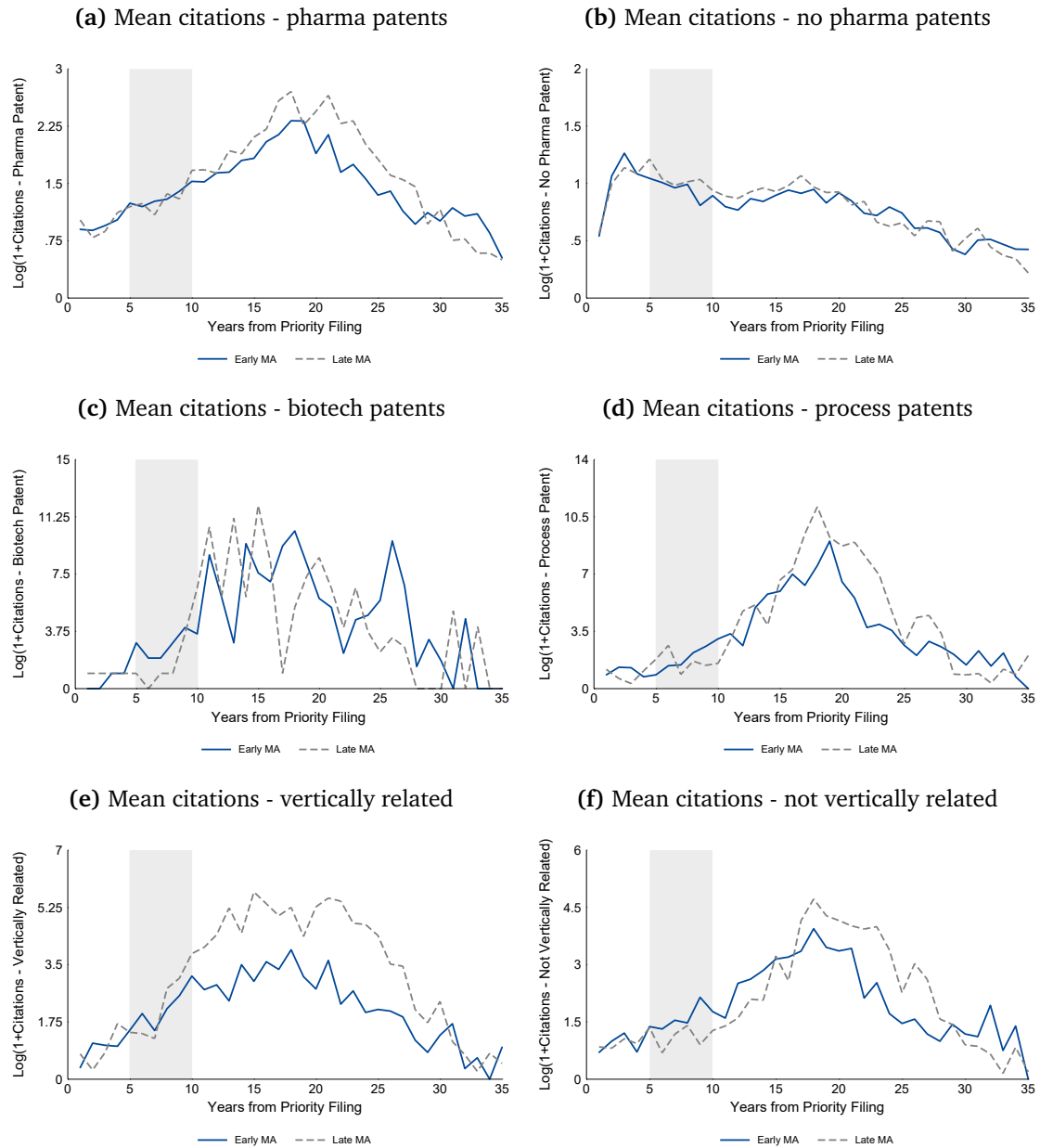
(a) Mean citations - pharma patents

(b) Mean citations - no pharma patents

(c) Mean citations - biotech patents

(d) Mean citations - process patents

(e) Mean citations - vertically related

(f) Mean citations - not vertically related

Notes: The figures present the average log-transformed number of forward citations split at the median approval lag (early/late MA) over time. The annual citation count is normalized by the average citation count from the pre-approval period t_0-t_5. Figures are based on the "main" sample of 590 patent-drug links. The unit of observation is the unique patent-drug level.

Figure B-5: Evolution of citations by *time to approval* – early/late MA

(a) Mean self-citations - early MA

(b) Mean self-citations - late MA

(c) Mean secondary patent citations - early MA

(d) Mean secondary patent citations - late MA

(e) Mean product patent citations - early MA

(f) Mean product patent citations - late MA
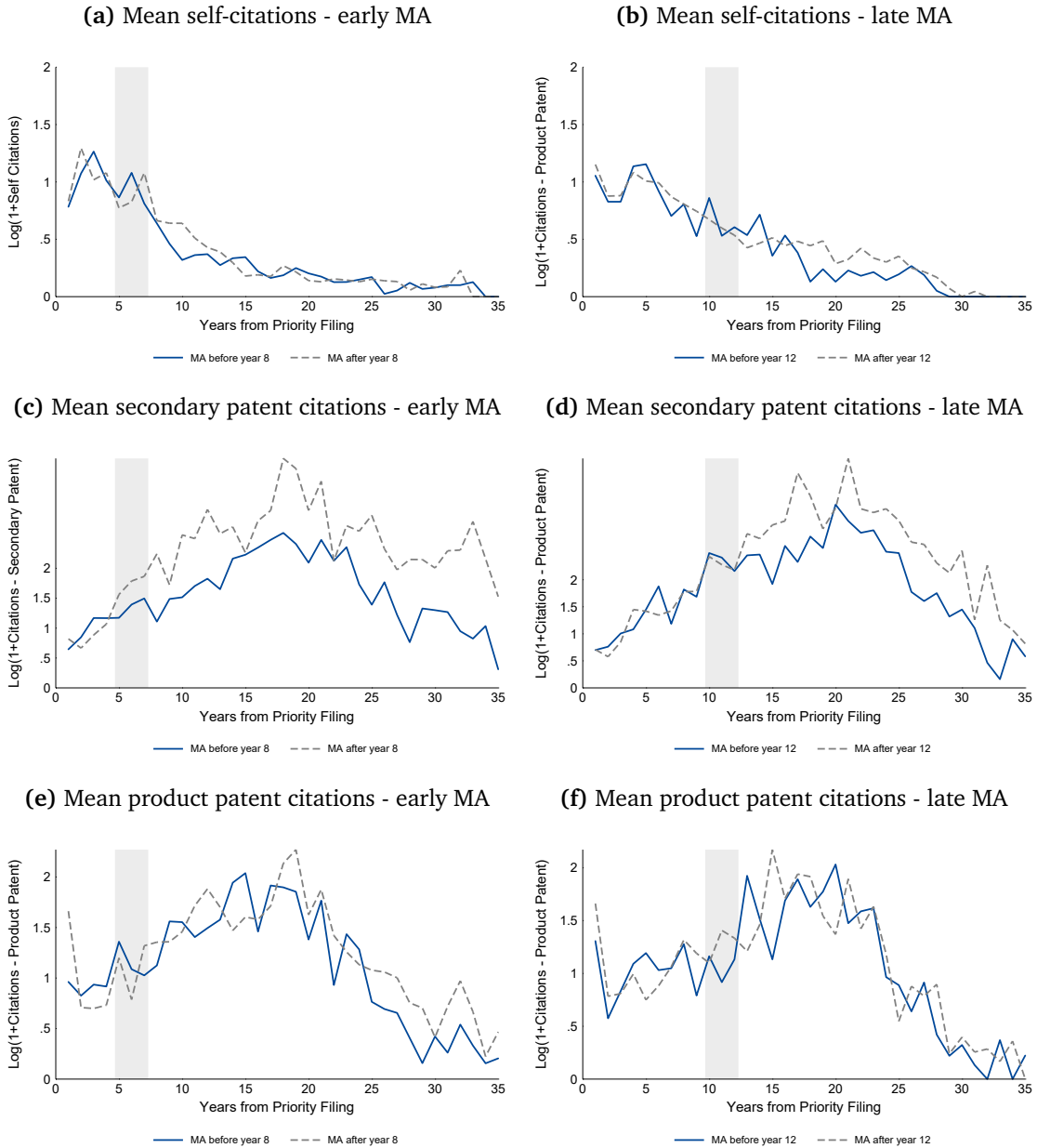
Notes: Figures (a), (c), (e) present the average log-transformed number of forward citations split at the median approval lag within the sample of early MA drugs (approval between year 5 and 10 from priority filing). Figures (b), (d), (f) present the average log-transformed number of forward citations split at the median approval lag within the sample of late MA drugs (approval between year 10 and 16 from priority filing). The annual citation count is normalized by the average citation count from the pre-approval period t_0-t_5. The unit of observation is the unique patent-drug level.

B.2.2   Figures – Multivariate Analysis

Figure B-6: Impact of marketing authorization on self citations – alternative specifications

(a) Without patent/SPC grant controls

(b) Winsorized

(c) PPML without patent/SPC grant controls

(d) PPML with patent/SPC grant controls
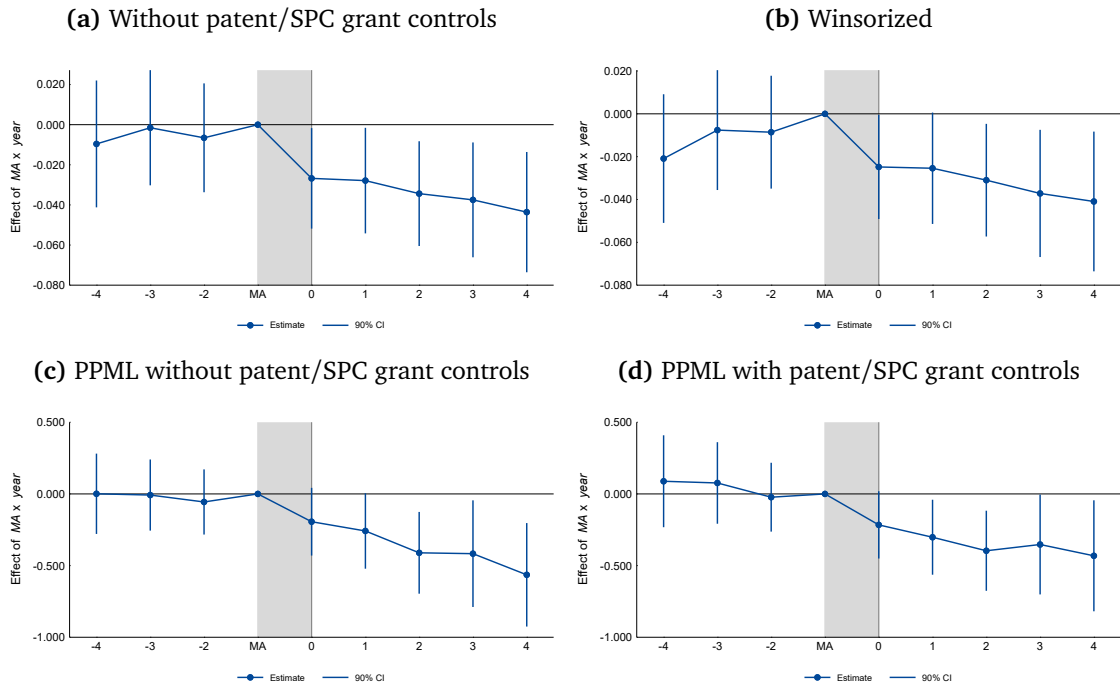
Notes: Figures (a) and (b) show the event study estimates and the 90 percent confidence bands of linear regressions with high-dimensional fixed effects. The outcome variable is the log-transformed annual self-citation count. Figure (a) follows the specification without patent grant and SPC grant controls (Equation 2.1). Figure (b) follows the preferred specification (Equation 2.1) with citation counts being winsorized at the 99% level. Figures (c) and (d) show the event study estimates and the 90 percent confidence bands of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The outcome variable is the self-citation count. Figure (c) follows the specification without patent grant and SPC grant controls (Equation 2.1). Figure (d) follows the preferred specification (Equation 2.1). The unit of observation is the unique patent-drug level. Standard errors are clustered at the patent-drug level.
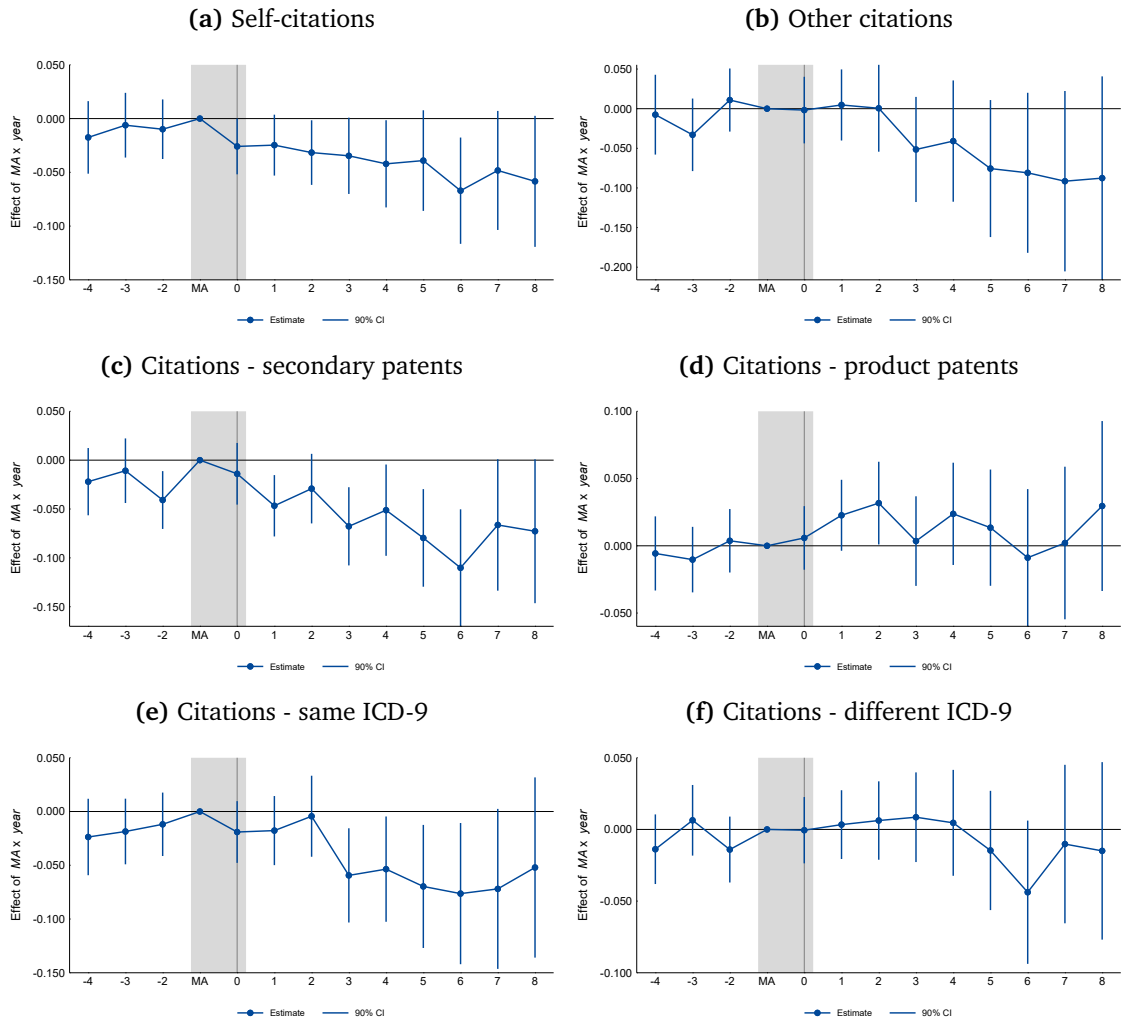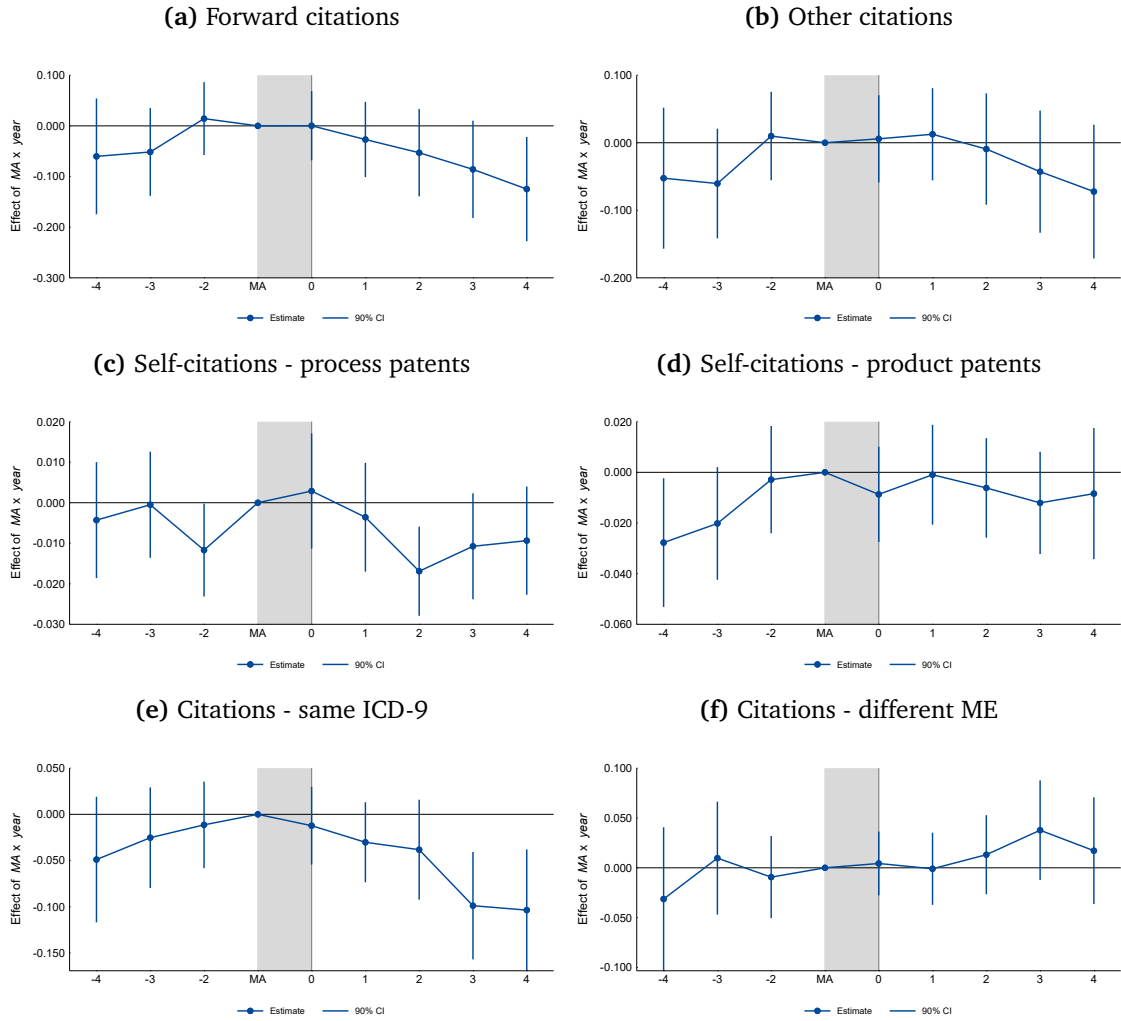
Figure B-7: Impact of marketing authorization on other parties' citations by type of patent

(a) Other citations

(b) Other citations - PPML

(c) Other citations - pharma patents

(d) Other citations - no pharma patents

(e) Other citations - same ICD-9

(f) Other citations - different ICD-9
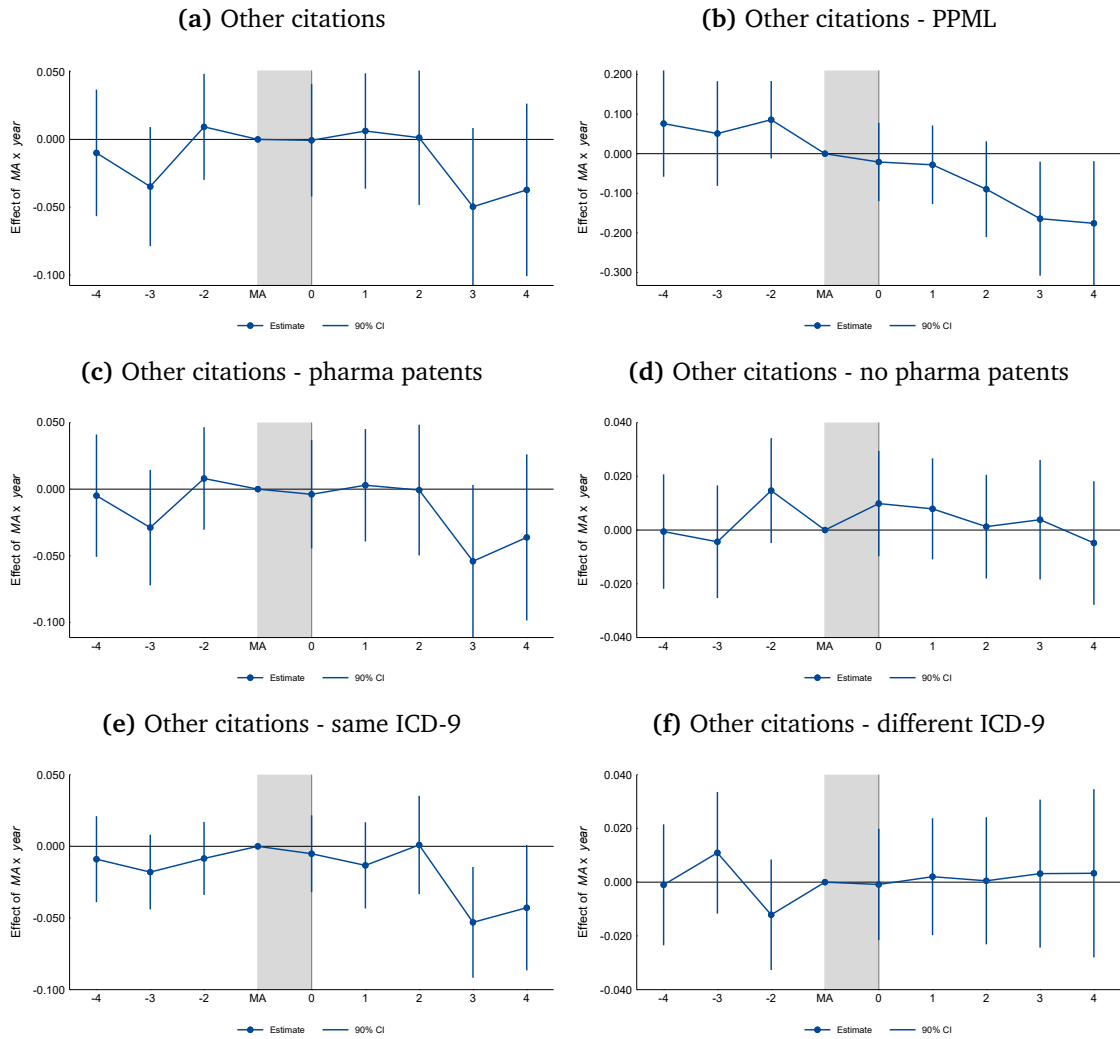
Notes: The Figures (a) & (c)–(f) show the event study estimates and the 90 percent confidence bands of linear regressions with high-dimensional fixed effects. The outcome variable is the log-transformed annual other parties' citation count generated by all patents in (a), the other pharmaceutical companies' citation count in (b), the other non-pharma companies' citation count in (c), the other parties' citation count generated by pharmaceutical patents in the same ICD-9 category in (e), and the other parties' citation count generated by pharmaceutical patents in a different ICD-9 category in (f). The Figures (b) shows the event study estimates and the 90 percent confidence bands of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The outcome variable is the annual other parties' citation. The unit of observation is the unique patent-drug level. Standard errors are clustered at the patent-drug level.
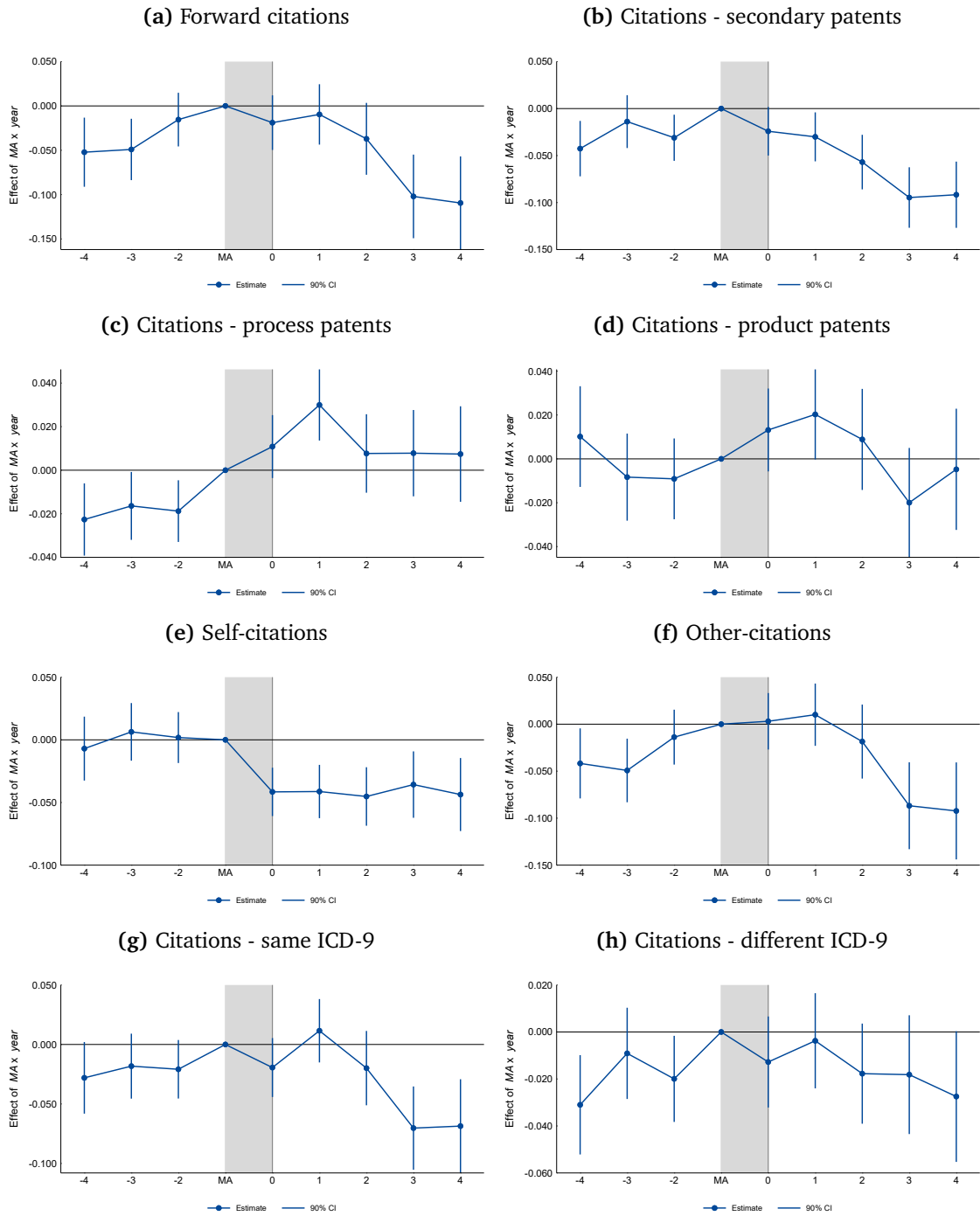
Figure B-8: Impact of marketing authorization on examiner/XY citations
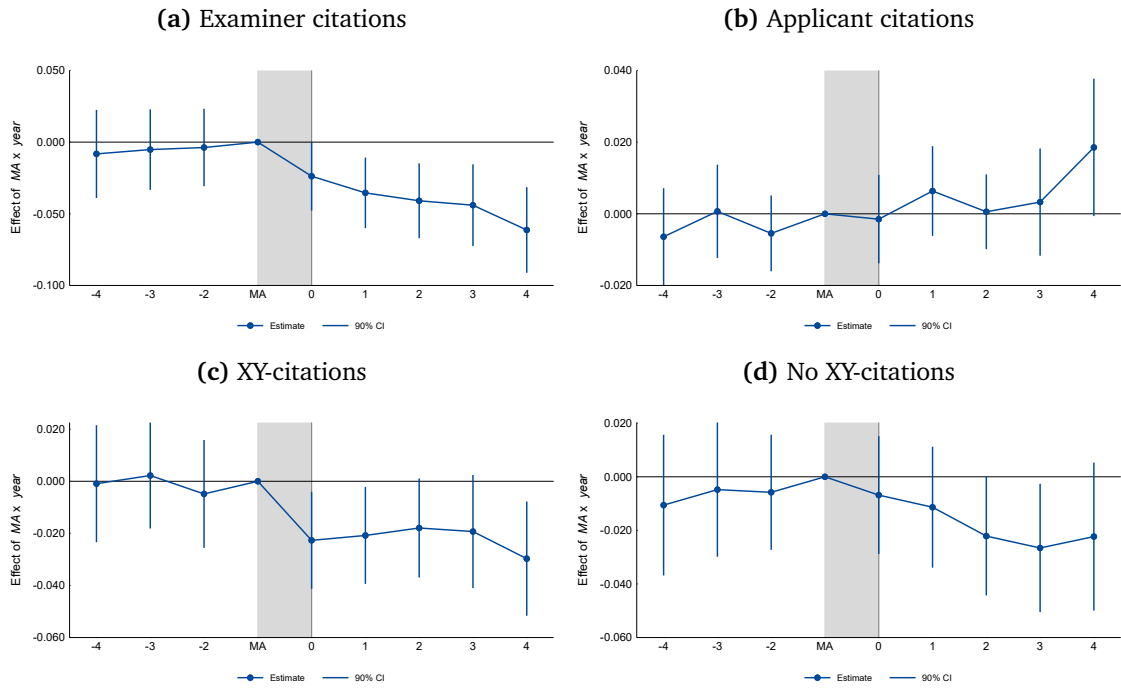
(a) Examiner citations

(b) Applicant citations

(c) XY-citations

(d) No XY-citations

Notes: The figures show the event study estimates and the 90 percent confidence bands of linear regressions with high-dimensional fixed effects. The outcome variable is the log-transformed annual self-citation count generated by (a) examiners, (b) applicants, (c) XY-references, and (d) no XY-references. The unit of observation is the unique patent-drug level. Standard errors are clustered at the patent-drug level.

Figure B-9: Impact of marketing authorization on citations – longer post period

(a) Self-citations

(b) Other citations

(c) Citations - secondary patents

(d) Citations - product patents

(e) Citations - same ICD-9

(f) Citations - different ICD-9

Notes: The figures show the event study estimates and the 90 percent confidence bands of linear regressions with high-dimensional fixed effects. It follows Equation 2.2 with longer effect windows. The treatment indicator MA_{it}^{j} is set to 1 at the endpoint $\bar{j} = +8$. The unit of observation is the unique patent-drug level. Standard errors are clustered at the patent-drug level.

Figure B-10: Impact of marketing authorization on citations – constant market exclusivity

(a) Forward citations

(b) Other citations

(c) Self-citations - process patents

(d) Self-citations - product patents

(e) Citations - same ICD-9

(f) Citations - different ME

Notes: The figures show the event study estimates and the 90 percent confidence bands of linear regressions with high-dimensional fixed effects. The sample comprises only those 288 patent-drugs links that are associated with a marketing authorization within 10 years. This early MA group has uniform market exclusivity period of 15 years. The outcome variable is the log-transformed annual forward citation count in (a), other parties' citation count in (b), self-citation count generated from process patents in (c), self-citation count generated from product patents in (d), the self-citation count generated by pharmaceutical patents in the same ICD-9 category in (e), and the self-citation count generated by pharmaceutical patents in a different ICD-9 category in (f). The unit of observation is the unique patent-drug level. Standard errors are clustered at the patent-drug level.

Figure B-11: Impact of marketing authorization on citations – full sample

(a) Forward citations

(b) Citations - secondary patents

(c) Citations - process patents

(d) Citations - product patents

(e) Self-citations

(f) Other-citations

(g) Citations - same ICD-9

(h) Citations - different ICD-9

Notes: The figures show the event study estimates and the 90 percent confidence bands of linear regressions with high-dimensional fixed effects using the full sample of 1,405 SPC filings. The unit of observation is the unique patent-drug level. Standard errors are clustered at the patent-drug level.

B.3 Tables

B.3.1 Tables – Summary Statistics

Table B-1: Summary statistics – ex-ante patent/drug characteristics

	N	Mean	Median	Std. Dev.	Min	Max
Time to approval	590	10.79	11.00	2.90	5	16
Time to patent grant	590	2.38	2.00	1.31	0	10
Patent priority year	590	1987.18	1987.00	5.48	1974	1996
Patent application year	590	1988.29	1988.50	5.53	1975	1997
First patent grant year	590	1989.51	1990.00	5.88	1975	2005
First MA year	590	1997.96	1998.00	6.07	1986	2012
SPC filing year	590	2000.00	1999.00	5.49	1993	2015
SPC grant year	490	2003.41	2004.00	5.66	1995	2016
Market exclusivity	590	13.92	15.00	1.61	9	15
Total forward cit.	590	16.41	7.00	29.07	0	263
Total self cit.	590	2.21	1.00	4.92	0	55
Total other cit.	590	14.16	6.00	26.69	0	253
Total pharma cit.	590	7.56	3.00	13.86	0	114
Total no pharma cit.	590	8.86	4.00	17.35	0	164
Total same ICD9 cit.	590	4.71	1.00	9.26	0	80
Total other ICD9 cit.	590	2.41	1.00	5.92	0	57
Total biotech patent cit.	590	0.23	0.00	1.08	0	13
Total secondary patent cit.	590	4.40	1.00	7.84	0	63
Total process patent cit.	590	1.17	0.00	2.60	0	22
Total product patent cit.	590	2.68	1.00	6.58	0	75
Total clinical trials cit.	590	1.90	0.00	5.26	0	59
Total clinical trials cit.	590	1.50	0.00	4.41	0	45
Pediatric drug	590	0.07	0.00	0.25	0	1
Drug combination	590	0.31	0.00	0.46	0	1
Salt of drug molecule	590	0.16	0.00	0.37	0	1
Size of patent family	590	25.78	24.00	15.71	1	115
Number of applicants	590	1.09	1.00	0.32	1	3
IPC in mainarea 1	590	0.01	0.00	0.08	0	1
IPC in mainarea 2	590	0.11	0.00	0.31	0	1
IPC in mainarea 3	590	1.00	1.00	0.04	0	1
IPC in mainarea 4	590	0.02	0.00	0.13	0	1
Transn. patent family	590	0.86	1.00	0.35	0	1
Triadic patent family	590	0.54	1.00	0.50	0	1
Tech area organic chem.	590	0.47	0.00	0.50	0	1
Tech area pharma.	590	0.42	0.00	0.49	0	1
Tech area biotech.	590	0.08	0.00	0.27	0	1
Tech area material chem.	590	0.01	0.00	0.09	0	1
Inventor country US	590	0.36	0.00	0.48	0	1
Inventor country Europe	590	0.40	0.00	0.49	0	1
Applicant country US	590	0.35	0.00	0.48	0	1
Applicant country Europe	590	0.45	0.00	0.50	0	1

Notes: This table displays the summary statistics of ex-ante patent & drug characteristics. *Initial* forward citation counts include all patent references within 12 months from the primary patents priority filing. The unit of observation is the unique patent-drug level.

Table B-2: Summary statistics – diseases

	N	Mean	Median	Std. Dev.	Min	Max
ICD9 code available	493	0.85	1.00	0.36	0	1
Number of ICD9	493	2.29	2.00	2.39	0	23
Infectious/parasitic diseases	493	0.24	0.00	0.43	0	1
Neoplasms	493	0.15	0.00	0.35	0	1
Endocrine/immun. disorders	493	0.15	0.00	0.36	0	1
Blood diseases	493	0.03	0.00	0.17	0	1
Mental disorders	493	0.07	0.00	0.25	0	1
Nervous system diseases	493	0.11	0.00	0.32	0	1
Circulatory system diseases	493	0.16	0.00	0.37	0	1
Respiratory system diseases	493	0.06	0.00	0.24	0	1
Digestive system diseases	493	0.05	0.00	0.22	0	1
Genitourinary diseases	493	0.10	0.00	0.30	0	1
Pregnancy/childbirth	493	0.00	0.00	0.06	0	1
Skin diseases diseases	493	0.04	0.00	0.21	0	1
Musculoskeletal diseases	493	0.08	0.00	0.27	0	1
Congenital anomalies	493	0.00	0.00	0.00	0	0
Conditions perinatal period	493	0.03	0.00	0.17	0	1
Ill-defined conditions	493	0.09	0.00	0.28	0	1
Ijury/poisoning	493	0.07	0.00	0.26	0	1

Notes: This table displays the summary statistics of disease characteristics. A primary patent can be associated with more than one ICD-9 category. The unit of observation is the unique patent-drug level.

B.3.2   Tables – Results from Cross Section

Table B-3: Impact of marketing authorization on citations – cross section

Log/Linear DV: Log(1+Citations)	(1) All	(2) Self	(3) Other	(4) Seondary	(5) Product	(6) = ICD9	(7) ≠ ICD9
Time to Approval (Priority)	0.027*	0.026**	0.015	0.020	0.018	0.027*	0.004
	(0.017)	(0.013)	(0.017)	(0.014)	(0.013)	(0.016)	(0.013)
Priority Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Patent Grant Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Patent Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Product Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Technology Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Inventor Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	584	584	584	584	584	584	584
Log-likelihood	−775	−636	−788	−669	−621	−726	−620

Notes: Columns (1) to (7) show the estimates of linear regressions using a cross-sectional data set. The dependent variable is the total number of log-transformed forward citations between year 5 and year 16 (first approval in data set and last approval in data set). Column (1) counts all citations, Column (2) self-citations, Column (3) other parties' citations, Column (4) citations generated from secondary patents, Column (5) citations generated from product patents, Column (6) citations within the same ICD-9 category, Column (7) citations in different ICD-9 categories. The unit of observation is the unique patent-drug level. Heteroskedasticity robust standard errors are shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table B-4: Impact of marketing authorization on citations by patent type – cross section

Log/Linear DV: Log(1+Citations)	(1) Secondary Patents	(2)	(3) Process Patents	(4)	(5) Product Patents	(6)	(7) Biotech Patents	(8)
	Self	Other	Self	Other	Self	Other	Self	Other
Time to Approval (Priority)	0.009	0.013	0.012**	0.012**	0.007	0.012	−0.001	0.010
	(0.008)	(0.013)	(0.005)	(0.005)	(0.006)	(0.013)	(0.001)	(0.006)
Priority Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Patent Grant Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Patent Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Product Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Technology Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Inventor Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	584	584	584	584	584	584	584	584
Log-likelihood	−404	−654	−44	−44	−232	−611	1146	−136

Notes: Columns (1) to (7) show the estimates of linear regressions using a cross-sectional data set. In Columns with uneven numbers, the dependent variable is the total number of log-transformed self-citations. In Columns with even numbers, the dependent variable is the total number of log-transformed other parties'-citations. In all cases, citations are counted between year 5 and year 16 (first approval in data set and last approval in data set). Citations are generated from secondary patents in Columns (1)/(2), process patents in Columns (3)/(4), product patents in Columns (5)/(6), biotech patents in Columns (7)/(8). The unit of observation is the unique patent-drug level. Heteroskedasticity robust standard errors are shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table B-5: Impact of marketing authorization on citations by originator – cross section

Log/Linear DV: Log(1+Citations)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Pharma		No Pharma		Same ICD9		Other ICD9	
	Self	Other	Self	Other	Self	Other	Self	Other
Time to Approval (Priority)	0.024**	0.010	0.004	0.016	0.023***	0.013	0.002	0.001
	(0.010)	(0.016)	(0.010)	(0.016)	(0.009)	(0.015)	(0.006)	(0.012)
Priority Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Patent Grant Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Patent Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Product Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Technology Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Inventor Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	584	584	584	584	584	584	584	584
Log-likelihood	−499	−734	−511	−776	−409	−714	−226	−596

Notes: Columns (1) to (7) show the estimates of linear regressions using a cross-sectional data set. In Columns with uneven numbers, the dependent variable is the total number of log-transformed self-citations. In Columns with even numbers, the dependent variable is the total number of log-transformed other parties'-citations. In all cases, citations are counted between year 5 and year 16 (first approval in data set and last approval in data set). Citations are generated from pharmaceutical patents in Columns (1)/(2), non-pharmaceutical patents in Columns (3)/(4), patents in the same ICD-9 category in Columns (5)/(6), patents in a different ICD-9 category in Columns (7)/(8). The unit of observation is the unique patent-drug level. Heteroskedasticity robust standard errors are shown in parentheses. Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table B-6: Impact of marketing authorization on self-citations – IV regression

Log/Linear DV: Log(1+Self Citations	(1)	(2)	(3)	(4)	(5)	(6)
	IV: Time to Phase I Trials)			IV: Time to Phase III Trials		
	OLS	Reduced Form	IV	OLS	Reduced Form	IV
Time to Approval (Priority)	0.049		0.290*	0.045		0.060
	(0.051)		(0.169)	(0.040)		(0.055)
IV: Time Phase I		0.082*				
		(0.044)				
IV: Time Phase III					0.047	
					(0.044)	
Priority Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Patent Grant Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Patent Controls	Yes	Yes	Yes	Yes	Yes	Yes
Underidentification test			9.18			36.99
Weak identification test			9.34			82.33
Endogeneity test			4.14			0.18
p-value			0.04			0.67
Observations	77	77	77	125	125	125

Notes: Columns (1), (2), (4), and (5) show the estimates of linear regressions using a cross-sectional data set. The dependent variable is the total number of log-transformed self-citations between year 5 and year 16 (first approval in data set and last approval in data set). Columns (3) and (6) show the estimates of 2SLS regressions. The "Time to Approval" variable is instrumented with the "Time to Phase I" in Column (3) and with the "Time to Phase III" in Column (6). The underidentification and weak identification tests are the heteroskedasticity-robust Kleibergen and Paap (2006) rk LM and Wald F statistics, respectively. I use the *ivreghdfe* Stata package as described in Correia (2018). The unit of observation is the unique patent-drug level. Heteroskedasticity robust standard errors are shown in parentheses. Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

B.3.3 Tables – Results from Panel

Table B-7: Impact of marketing authorization on self-citations – alternative specifications

Log/Linear	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				DV: Log(1+ Self Citations)			
	Baseline	Different Controls		Without θ FE		pre-1993	ICD9 avail.
n years before MA	0.006	0.004	−0.017	−0.019	−0.038	−0.006	−0.007
	(0.018)	(0.018)	(0.022)	(0.018)	(0.024)	(0.023)	(0.028)
4 years before MA	−0.015	−0.010	−0.025	−0.026	−0.039*	−0.012	−0.037
	(0.021)	(0.020)	(0.020)	(0.016)	(0.019)	(0.022)	(0.025)
3 years before MA	−0.005	−0.002	−0.011	−0.013	−0.017	−0.013	−0.008
	(0.015)	(0.014)	(0.018)	(0.017)	(0.020)	(0.020)	(0.024)
2 years before MA	−0.009	−0.007	−0.012	−0.013	−0.014	−0.023	−0.019
	(0.014)	(0.013)	(0.017)	(0.016)	(0.018)	(0.019)	(0.021)
1 year before MA	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Year of MA	−0.027*	−0.027*	−0.024	−0.022	−0.020	−0.042**	−0.033
	(0.015)	(0.014)	(0.016)	(0.017)	(0.021)	(0.017)	(0.020)
1 year after MA	−0.029*	−0.028*	−0.022	−0.019	−0.027	−0.039**	−0.049**
	(0.016)	(0.014)	(0.016)	(0.013)	(0.016)	(0.019)	(0.021)
2 years after MA	−0.036**	−0.034**	−0.026	−0.021	−0.030	−0.050**	−0.054**
	(0.016)	(0.014)	(0.017)	(0.016)	(0.019)	(0.020)	(0.022)
3 years after MA	−0.040**	−0.037*	−0.029	−0.020	−0.028	−0.056**	−0.065***
	(0.020)	(0.019)	(0.019)	(0.018)	(0.022)	(0.022)	(0.024)
4 years after MA	−0.046**	−0.044**	−0.035*	−0.021	−0.023	−0.068***	−0.067**
	(0.022)	(0.019)	(0.021)	(0.013)	(0.020)	(0.023)	(0.028)
n years after MA	−0.053**	−0.059***	−0.049**	−0.009	−0.014	−0.081***	−0.078**
	(0.024)	(0.021)	(0.025)	(0.021)	(0.027)	(0.029)	(0.033)
Patent Grant	Yes	No	Yes	Yes	Yes	Yes	Yes
SPC Grant	Yes	No	Yes	Yes	Yes	Yes	Yes
Age squared	No	No	Yes	No	No	No	No
Priority Year FE	No	No	No	Yes	Yes	No	No
ICD9 FE	No	No	No	No	Yes	No	No
Patent-Drug FE	Yes	Yes	Yes	No	No	Yes	Yes
Citation Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12390	12390	12390	12390	10353	9119	8756
Cluster	590	590	590	23	23	456	417
Log-likelihood	−1712	−1745	−1678	−3378	−3059	−885	−1790

Notes: Columns (1) to (7) show the estimates of linear regressions with high-dimensional fixed effects in various specifications. The dependent variable is the annual number of log-transformed self-citations. Column (1) shows the results from Equation 2.2. Column (2) is without controls for patent grant and SPC grant. Column (3) adds a control for squared age in order to account for potential nonlinearities not covered by fixed effects. Column (4) replaces patent-drug fixed effects by priority year FE and Column (5) by priority year FE and ICD-9 FE in order to account more explicitly for age-cohort links. Column (6) restricts the sample to pre-1993, in which patent term extension did not exist at the time of the patent filing. Lastly Column (7) restricts the sample to observations with available ICD-9 link. The unit of observation is the unique patent-drug level by year. Standard errors are clustered at the patent-drug level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table B-8: Impact of marketing authorization on citations by type of patent

Log/Linear DV: Log(1+Citations)	(1) Secondary	(2) Patent	(3) Process	(4) Patent	(5) Product	(6) Patent	(7) Biotech	(8) Patent
	Self	Other	Self	Other	Self	Other	Self	Other
n years before MA	−0.011	−0.008	0.001	−0.017	0.002	0.016	0.002	−0.005
	(0.012)	(0.024)	(0.005)	(0.013)	(0.009)	(0.019)	(0.003)	(0.006)
4 years before MA	−0.010	−0.009	−0.004	−0.018*	−0.012	0.016	0.001	−0.003
	(0.012)	(0.018)	(0.005)	(0.010)	(0.008)	(0.015)	(0.001)	(0.005)
3 years before MA	−0.007	0.000	0.004	−0.011	−0.007	0.002	0.001	−0.005
	(0.011)	(0.017)	(0.006)	(0.010)	(0.008)	(0.013)	(0.001)	(0.005)
2 years before MA	−0.015	−0.027*	−0.005	−0.013	0.001	0.004	0.002	−0.008
	(0.009)	(0.016)	(0.005)	(0.008)	(0.008)	(0.013)	(0.001)	(0.005)
1 year before MA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Year of MA	−0.020*	0.008	0.006	−0.003	−0.006	0.006	0.002	−0.001
	(0.010)	(0.018)	(0.006)	(0.010)	(0.007)	(0.013)	(0.002)	(0.005)
1 year after MA	−0.018*	−0.032*	−0.001	0.012	−0.002	0.018	0.000	−0.003
	(0.010)	(0.018)	(0.006)	(0.012)	(0.007)	(0.014)	(0.001)	(0.007)
2 years after MA	−0.022**	−0.013	−0.008*	−0.003	−0.002	0.024	−0.001	−0.006
	(0.011)	(0.019)	(0.005)	(0.013)	(0.007)	(0.017)	(0.001)	(0.005)
3 years after MA	−0.026**	−0.049**	−0.008*	−0.007	−0.002	−0.011	−0.001	−0.009
	(0.012)	(0.022)	(0.005)	(0.015)	(0.008)	(0.017)	(0.001)	(0.008)
4 years after MA	−0.028**	−0.030	−0.011*	−0.006	0.002	0.006	0.001	−0.005
	(0.013)	(0.025)	(0.006)	(0.017)	(0.009)	(0.020)	(0.001)	(0.008)
n years after MA	−0.037***	−0.054**	−0.010*	−0.009	0.001	−0.007	0.001	−0.006
	(0.013)	(0.028)	(0.006)	(0.018)	(0.010)	(0.023)	(0.001)	(0.009)
Patent Grant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SPC Grant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Patent-Drug FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Citation Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12390	12390	12390	12390	12390	12390	12390	12390
Cluster	590	590	590	590	590	590	590	590
Log-likelihood	5881	−3287	14458	2094	8917	−764	29205	13076

Notes: Columns (1) to (8) show the estimates of linear regressions with high-dimensional fixed effects in various specifications. In Columns with uneven numbers, the dependent variable is the total number of log-transformed self-citations. In Columns with even numbers, the dependent variable is the total number of log-transformed other parties'-citations. Citations are generated from secondary patents in Columns (1)/(2), process patents in Columns (3)/(4), product patents in Columns (5)/(6), biotech patents in Columns (7)/(8). The unit of observation is the unique patent-drug level by year. Standard errors are clustered at the patent-drug level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table B-9: Impact of marketing authorization on citations by disease type

Log/Linear DV: Log(1+Citations)	(1) Pharma	(2)	(3) No Pharma	(4)	(5) Same ICD9	(6)	(7) Other ICD9	(8)
	Self	Other	Self	Other	Self	Other	Self	Other
n years before MA	−0.007	−0.025	0.021	0.025	0.003	−0.025	−0.003	0.025
	(0.015)	(0.029)	(0.017)	(0.032)	(0.012)	(0.023)	(0.009)	(0.021)
4 years before MA	−0.016	−0.015	0.007	0.008	−0.008	−0.009	−0.007	−0.001
	(0.015)	(0.022)	(0.016)	(0.024)	(0.012)	(0.018)	(0.008)	(0.014)
3 years before MA	0.001	−0.018	−0.004	−0.012	0.003	−0.018	−0.002	0.011
	(0.013)	(0.020)	(0.015)	(0.022)	(0.011)	(0.016)	(0.008)	(0.014)
2 years before MA	−0.011	−0.030	0.002	0.039*	−0.007	−0.008	−0.001	−0.012
	(0.012)	(0.018)	(0.013)	(0.021)	(0.011)	(0.015)	(0.006)	(0.012)
1 year before MA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Year of MA	−0.019	0.000	−0.011	0.002	−0.022**	−0.005	0.002	−0.001
	(0.013)	(0.020)	(0.011)	(0.022)	(0.011)	(0.016)	(0.007)	(0.013)
1 year after MA	−0.018	−0.015	−0.014	0.019	−0.015	−0.013	−0.001	0.002
	(0.012)	(0.021)	(0.012)	(0.023)	(0.011)	(0.018)	(0.006)	(0.013)
2 years after MA	−0.021*	−0.002	−0.021*	−0.001	−0.020*	0.001	0.002	0.001
	(0.012)	(0.024)	(0.012)	(0.025)	(0.012)	(0.021)	(0.007)	(0.014)
3 years after MA	−0.025*	−0.062**	−0.016	0.003	−0.028**	−0.053**	0.004	0.003
	(0.014)	(0.028)	(0.013)	(0.030)	(0.012)	(0.023)	(0.008)	(0.017)
4 years after MA	−0.030*	−0.037	−0.017	−0.016	−0.033**	−0.043	−0.003	0.003
	(0.016)	(0.031)	(0.013)	(0.032)	(0.013)	(0.027)	(0.008)	(0.019)
n years after MA	−0.040**	−0.067*	−0.013	−0.018	−0.035**	−0.054*	−0.004	−0.017
	(0.017)	(0.035)	(0.017)	(0.041)	(0.014)	(0.030)	(0.009)	(0.022)
Patent Grant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SPC Grant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Patent-Drug FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Citation Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12390	12390	12390	12390	12390	12390	12390	12390
Cluster	590	590	590	590	590	590	590	590
Log-likelihood	2664	−5869	1351	−6796	5125	−3885	9751	−653

Notes: Columns (1) to (8) show the estimates of linear regressions with high-dimensional fixed effects in various specifications. In Columns with uneven numbers, the dependent variable is the total number of log-transformed self-citations. In Columns with even numbers, the dependent variable is the total number of log-transformed other parties'-citations. Citations are generated from pharmaceutical patents in Columns (1)/(2), non-pharmaceutical patents in Columns (3)/(4), patents in the same ICD-9 category in Columns (5)/(6), patents in a different ICD-9 category in Columns (7)/(8). The unit of observation is the unique patent-drug level by year. Standard errors are clustered at the patent-drug level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table B-10: Impact of marketing authorization on citations by citation type

Log/Linear DV: Log(1+Citations)	(1) Examiner	(2) 	(3) Applicant	(4) 	(5) XY	(6) 	(7) No XY	(8)
	Self	Other	Self	Other	Self	Other	Self	Other
n years before MA	0.012	0.005	−0.005	−0.040	0.003	0.021	0.004	−0.032
	(0.019)	(0.032)	(0.012)	(0.028)	(0.013)	(0.024)	(0.018)	(0.034)
4 years before MA	−0.008	0.012	−0.006	−0.034*	−0.001	0.012	−0.011	−0.033
	(0.019)	(0.026)	(0.008)	(0.019)	(0.014)	(0.021)	(0.016)	(0.025)
3 years before MA	−0.005	−0.028	0.001	−0.018	0.002	−0.002	−0.005	−0.040*
	(0.017)	(0.024)	(0.008)	(0.017)	(0.012)	(0.019)	(0.015)	(0.023)
2 years before MA	−0.004	0.018	−0.005	−0.014	−0.005	0.015	−0.006	0.000
	(0.016)	(0.023)	(0.006)	(0.014)	(0.013)	(0.019)	(0.013)	(0.021)
1 year before MA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Year of MA	−0.024	−0.008	−0.002	0.009	−0.023**	−0.006	−0.007	0.012
	(0.015)	(0.023)	(0.007)	(0.016)	(0.011)	(0.018)	(0.013)	(0.023)
1 year after MA	−0.035**	−0.007	0.006	0.023	−0.021*	0.023	−0.011	−0.013
	(0.015)	(0.024)	(0.008)	(0.018)	(0.011)	(0.020)	(0.014)	(0.023)
2 years after MA	−0.041**	−0.036	0.001	0.038*	−0.018	0.000	−0.022	0.002
	(0.016)	(0.027)	(0.006)	(0.022)	(0.012)	(0.021)	(0.013)	(0.027)
3 years after MA	−0.044**	−0.063**	0.003	0.014	−0.019	−0.015	−0.027*	−0.044
	(0.017)	(0.031)	(0.009)	(0.026)	(0.013)	(0.025)	(0.015)	(0.031)
4 years after MA	−0.061***	−0.098***	0.019	0.051*	−0.030**	−0.042	−0.022	−0.019
	(0.018)	(0.033)	(0.012)	(0.030)	(0.013)	(0.027)	(0.017)	(0.033)
n years after MA	−0.064***	−0.122***	0.011	0.028	−0.026*	−0.067**	−0.035*	−0.047
	(0.020)	(0.038)	(0.013)	(0.040)	(0.015)	(0.030)	(0.020)	(0.043)
Patent Grant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SPC Grant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Patent-Drug FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Citation Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12390	12390	12390	12390	12390	12390	12390	12390
Cluster	590	590	590	590	590	590	590	590
Log-likelihood	−435	−6962	7196	−5206	3342	−4127	981	−7508

Notes: Columns (1) to (8) show the estimates of linear regressions with high-dimensional fixed effects in various specifications. In Columns with uneven numbers, the dependent variable is the total number of log-transformed self-citations. In Columns with even numbers, the dependent variable is the total number of log-transformed other parties'-citations. Citations are generated by examiners in Columns (1)/(2), applicants in Columns (3)/(4), XY-references in Columns (5)/(6), no XY-references in Columns (7)/(8). The unit of observation is the unique patent-drug level by year. Standard errors are clustered at the patent-drug level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table B-11: Impact of marketing authorization on winsorized citations

Log/Linear	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			DV: Log(1+Forward Citations winsorized)					
	All	Self	Other	Secondary	Process	Product	= ICD9	≠ ICD9
n years before MA	−0.020	0.001	−0.025	−0.038	−0.015	0.000	−0.032	0.002
	(0.036)	(0.019)	(0.035)	(0.023)	(0.012)	(0.017)	(0.023)	(0.018)
4 years before MA	−0.030	−0.021	−0.016	−0.025	−0.020*	−0.003	−0.020	−0.017
	(0.031)	(0.018)	(0.028)	(0.020)	(0.010)	(0.016)	(0.020)	(0.014)
3 years before MA	−0.041	−0.008	−0.041	−0.013	−0.006	−0.007	−0.016	0.004
	(0.028)	(0.017)	(0.027)	(0.019)	(0.010)	(0.014)	(0.018)	(0.015)
2 years before MA	0.001	−0.009	0.007	−0.042**	−0.015*	0.005	−0.010	−0.015
	(0.026)	(0.016)	(0.024)	(0.018)	(0.009)	(0.014)	(0.018)	(0.014)
1 year before MA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Year of MA	−0.021	−0.025*	0.001	−0.015	0.004	0.006	−0.021	0.000
	(0.026)	(0.015)	(0.025)	(0.019)	(0.011)	(0.014)	(0.017)	(0.014)
1 year after MA	−0.014	−0.025	0.009	−0.045**	0.014	0.016	−0.024	0.005
	(0.028)	(0.016)	(0.026)	(0.019)	(0.012)	(0.015)	(0.019)	(0.014)
2 years after MA	−0.020	−0.031*	0.005	−0.030	−0.008	0.021	−0.015	0.008
	(0.032)	(0.016)	(0.030)	(0.021)	(0.013)	(0.017)	(0.021)	(0.016)
3 years after MA	−0.079**	−0.037**	−0.046	−0.069***	−0.016	−0.011	−0.071***	0.007
	(0.037)	(0.018)	(0.035)	(0.023)	(0.014)	(0.017)	(0.024)	(0.017)
4 years after MA	−0.070*	−0.041**	−0.034	−0.053**	−0.014	0.005	−0.066**	0.002
	(0.040)	(0.020)	(0.039)	(0.025)	(0.016)	(0.019)	(0.026)	(0.019)
n years after MA	−0.120**	−0.051**	−0.076	−0.088***	−0.017	−0.011	−0.079***	−0.028
	(0.048)	(0.023)	(0.046)	(0.028)	(0.017)	(0.022)	(0.031)	(0.022)
Patent Grant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SPC Grant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Patent-Drug FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Citation Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12390	12390	12390	12390	12390	12390	12390	12390
Cluster	590	590	590	590	590	590	590	590
Log-likelihood	−9090	−935	−8606	−3521	2628	−551	−4007	−428

Notes: Columns (1) to (8) show the estimates of linear regressions with high-dimensional fixed effects. The dependent variable is the annual number of log-transformed forward citations in Column (1), self-citations in Column (2), other parties' citations in Column (3), forward citations generated by secondary patents in Column (4), by process patents in Column (5), by product patents in Column (6), within the same ICD-9 category in Column (7), and in a different ICD-9 disease category in Column (8). All outcome variables are winsorized at the 99% level. The unit of observation is the unique patent-drug level by year. Standard errors are clustered at the patent-drug level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table B-12: Impact of marketing authorization on citations – PPML regression

Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				DV: Forward Citations				
	All	Self	Other	Secondary	Process	Product	= ICD9	≠ ICD9
n years before MA	0.291**	0.230	0.279**	0.196	−0.662**	0.456**	0.009	0.501**
	(0.125)	(0.254)	(0.125)	(0.176)	(0.311)	(0.217)	(0.201)	(0.216)
4 years before MA	0.090	0.088	0.076	0.031	−0.982***	0.160	−0.023	−0.070
	(0.083)	(0.195)	(0.082)	(0.130)	(0.375)	(0.179)	(0.151)	(0.188)
3 years before MA	0.071	0.076	0.051	0.101	−0.324	−0.054	−0.028	0.135
	(0.075)	(0.173)	(0.080)	(0.117)	(0.285)	(0.157)	(0.115)	(0.180)
2 years before MA	0.067	−0.023	0.086	−0.178*	−0.598**	0.072	−0.056	−0.120
	(0.057)	(0.146)	(0.059)	(0.101)	(0.236)	(0.135)	(0.104)	(0.163)
1 year before MA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Year of MA	−0.060	−0.216	−0.021	−0.064	0.056	−0.041	−0.108	0.028
	(0.057)	(0.143)	(0.060)	(0.104)	(0.216)	(0.134)	(0.096)	(0.140)
1 year after MA	−0.074	−0.302*	−0.028	−0.238**	0.225	0.119	−0.094	−0.019
	(0.058)	(0.159)	(0.060)	(0.098)	(0.222)	(0.140)	(0.104)	(0.132)
2 years after MA	−0.143**	−0.397**	−0.090	−0.244**	−0.054	0.045	−0.049	−0.093
	(0.069)	(0.170)	(0.074)	(0.095)	(0.253)	(0.154)	(0.111)	(0.137)
3 years after MA	−0.209**	−0.353*	−0.164*	−0.409***	0.076	−0.141	−0.287**	−0.020
	(0.082)	(0.212)	(0.088)	(0.107)	(0.267)	(0.144)	(0.131)	(0.137)
4 years after MA	−0.232***	−0.432*	−0.176*	−0.311**	0.065	−0.029	−0.204	−0.068
	(0.089)	(0.235)	(0.095)	(0.121)	(0.248)	(0.157)	(0.139)	(0.158)
n years after MA	−0.265***	−0.354	−0.217**	−0.441***	0.115	−0.109	−0.258*	−0.169
	(0.097)	(0.280)	(0.103)	(0.115)	(0.264)	(0.169)	(0.134)	(0.154)
Patent Grant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SPC Grant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Patent-Drug FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Citation Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12182	8853	12035	9687	6161	8331	8512	7727
Cluster	581	423	574	465	301	397	407	371
Log-likelihood	−17258	−5113	−15440	−7157	−2905	−4955	−7388	−4477

Notes: Columns (1) to (8) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The dependent variable is the annual number of forward citations in Column (1), self-citations in Column (2), other parties' citations in Column (3), forward citations generated by secondary patents in Column (4), by process patents in Column (5), by product patents in Column (6), within the same ICD-9 category in Column (7), and in a different ICD-9 disease category in Column (8). The unit of observation is the unique patent-drug level by year. Standard errors are clustered at the patent-drug level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table B-13: Impact of end on phase II/beginning of phase III clinical trials on citations

Log/Linear	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				DV: Log(1+Forward Citations)				
	All	Self	Other	Secondary	Process	Product	= ICD9	$\neq$ ICD9
n years before Phase III	−0.238**	−0.065	−0.201*	−0.064	−0.140***	−0.112**	−0.273***	0.024
	(0.115)	(0.057)	(0.115)	(0.086)	(0.049)	(0.056)	(0.084)	(0.064)
4 years before Phase III	−0.157*	−0.058	−0.118	−0.096*	−0.075**	−0.156***	−0.230***	−0.025
	(0.081)	(0.051)	(0.077)	(0.057)	(0.035)	(0.043)	(0.062)	(0.041)
3 years before Phase III	−0.152**	−0.076*	−0.102	−0.022	−0.042	−0.066	−0.096*	0.013
	(0.074)	(0.043)	(0.070)	(0.048)	(0.029)	(0.042)	(0.058)	(0.036)
2 years before Phase III	−0.140**	0.010	−0.148**	−0.082**	−0.026	−0.076**	−0.075	−0.032
	(0.066)	(0.044)	(0.062)	(0.037)	(0.028)	(0.037)	(0.049)	(0.028)
1 year before Phase III	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Year of Phase III	0.066	0.057	0.036	0.056	0.034	0.002	0.019	0.071*
	(0.064)	(0.042)	(0.059)	(0.048)	(0.032)	(0.040)	(0.051)	(0.038)
1 year after Phase III	0.111	0.044	0.088	0.106*	0.024	0.027	0.102*	0.077*
	(0.074)	(0.043)	(0.068)	(0.056)	(0.033)	(0.043)	(0.058)	(0.043)
2 years after Phase III	0.144*	0.006	0.152**	0.143**	0.070*	0.057	0.137**	0.063
	(0.074)	(0.043)	(0.068)	(0.059)	(0.038)	(0.044)	(0.061)	(0.040)
3 years after Phase III	0.187**	0.092*	0.142*	0.142**	0.100**	0.106**	0.167**	0.119**
	(0.088)	(0.052)	(0.084)	(0.064)	(0.049)	(0.051)	(0.066)	(0.051)
4 years after Phase III	0.162*	0.043	0.170*	0.088	0.105**	0.128**	0.089	0.129**
	(0.094)	(0.053)	(0.094)	(0.074)	(0.047)	(0.057)	(0.072)	(0.054)
n years after Phase III	0.144	0.027	0.159	0.043	0.110**	0.141**	0.065	0.122**
	(0.111)	(0.059)	(0.113)	(0.077)	(0.054)	(0.060)	(0.086)	(0.059)
Patent Grant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SPC Grant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Patent-Drug FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Citation Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2750	2750	2750	2750	2750	2750	2750	2750
Cluster	131	131	131	131	131	131	131	131
Log-likelihood	−2192	−566	−2110	−1327	−400	−809	−1509	−860

Notes: Columns (1) to (8) show the estimates of linear regressions with high-dimensional fixed effects. The event is the end of phase2/beginning of phase III clinical trials. The dependent variable is the annual number of log-transformed forward citations in Column (1), self-citations in Column (2), other parties' citations in Column (3), forward citations generated by secondary patents in Column (4), by process patents in Column (5), by product patents in Column (6), within the same ICD-9 category in Column (7), and in a different ICD-9 disease category in Column (8). The unit of observation is the unique patent-drug level by year. Standard errors are clustered at the patent-drug level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

C

Appendix to Chapter 3

Market Size and Research

C.1   Data Construction

Control Variables

This data appendix describes the construction of our control variables: projected market size, NIH funding, and research opportunities.

Projected Market Size

We build a measure of the exogenous components of U.S. market size (Acemoglu and Linn, 2004; Blume-Kohout and Sood, 2013). Each disease group has a different age profile and, hence, is differently affected by both domestic and global demographic trends. Therefore, we use demographic (projection) data from the UN World Population Prospects[1] for the United States between 1997 and 2040 in order to calculate how the potential future market size would develop if only population growth mattered.

To this end, we keep the age profile of each disease constant and calculate the average expenditure share of drugs associated with each five-year age bin for each ICD-9 group in the pre-MMA period. Drug expenditures are measured in real-terms (base year of 2003) based on the MEPS data. We then attribute the US population growth to each age bin until 2040. Hence, our measure for projected US market size M_{it} displays the annual expected drug expenditures in each ICD-9 group i in year t. In concordance with Blume-Kohout and Sood (2013), we accumulate the projected market size M_{it} over a period of 12 lead years as of year t. This reflects the average market exclusivity term of new drugs (Adams and Brantner, 2006).

Since pharmaceutical markets are typically considered as global (Acemoglu and Linn, 2004), we build the same measure for the potential market size in all OECD countries.

NIH Funding

We control for previous years' public research funding related to each disease category. Many scholars have shown the importance of public research funding, *e.g.*, from the NIH, for progress in biomedical research[2] and pharmaceutical innovation.[3] Since Congress doubled the NIH budget in the five years preceding the MMA from $13.6 billion in 1998 to $27.1 billion in 2003, it becomes an especially important determinant of R&D in any analysis of Part D (Smith, 2006). The NIH consists of twenty-seven Institutes and Centers, where each receives its own Congressional appropriation (Azoulay et al., 2019). However, the historical doubling of research funding was distributed unequally between these Institutes. In order to control for any

---

[1]The data can be found here: https://population.un.org/wpp/Download/Standard/Population/.

[2]See Jacob and Lefgren (2011) on scientific productivity, Myers (2020) on the direction of science, and Packalen and Bhattacharya (2018) on novelty.

[3]See Azoulay et al. (2019) on patenting and Blume-Kohout and Sood (2013) on NMEs entering clinical trials, and Toole (2012) on new drugs approved.

disproportionate increase in funding correlated with the MMA, we calculate for each of our 129 ICD-9 groups the exposure to the NIH budget over time.[4]

To this end, we assign each ICD-9 group to one of these Institutes (*e.g.*, ICD-9 162 *malignant neoplasm of trachea, bronchus, and lung* to the NCI *National Cancer Institute*). Since research grants are distributed within Institutes primarily by scientific merit (see discussion on NIH funding rules by Azoulay et al. (2019)) and not by allocation to narrower disease categories, we attribute the full annual Institute's budget to each ICD-9 group.[5] In an alternative approach we attribute budgets based on the share of all publications in a disease category that acknowledge a specific Institute. According to aforementioned studies, the effect of funding on research typically materializes within the first years from the grant. Therefore, we accumulate the NIH_{it} funding over a period of 12 lagged years until year t.[6]

Research Opportunities

Scientists may switch research projects to take advantage of greater research opportunities. We therefore account for the *availability of research opportunities* in similar fashion to Bhattacharya and Packalen (2011).[7]

We develop a direct measure of new research opportunities taking advantage of the development of the MeSH hierarchy over time.[8] The MeSH vocabulary in its current form was introduced in 1963 (Rogers, 1963) and was intended as a dynamic list that incorporates new concepts in the medical field.[9] The NLM introduces annually hundred new MeSH terms based on the need to appropriately describe concepts being discussed in the literature. This includes new diseases, a more detailed definition of existing diseases, as well as additional terminology to reflect topical areas that are not well represented in MeSH.[10] We interpret the introduction of a new MeSH term as an emerging research opportunity since NLM employees collect new terms that begin to appear in the scientific literature, for example in emerging areas of

---

[4]We retrieve NIH spending data (Mechanism Detail by IC, FY 1983-2019) from https://officeofbudget.od.nih.gov/ spending_hist.html [downloaded on February 17, 2020].

[5]In rare cases we assign more then one Institute or Center to an ICD-9 group. In these cases we attribute both budgets to the disease category.

[6]We use the Biomedical Research and Development Price Index in order to calculate real values with the base year of 2003. The data can be found here: https://officeofbudget.od.nih.gov/gbipriceindexes.html.

[7]Bhattacharya and Packalen (2011) construct a measures of research opportunities based on the content of research inputs and the first appearance of the idea in a scientific publication. Using the set of approved active ingredients as an input factor for future scientific research, they estimate structural productivity parameters, which takes into account diffusion and exhaustion of knowledge, in order to infer the quality of associated opportunities. The disadvantage of the approach is that it relies on a very narrow set of research inputs that relate primarily to drug-related medical research but not basic science.

[8]MeSH terms are organised into a hierarchy called the MeSH tree. Disease groups are first defined very broadly, but become more narrow with every sub-type of a disorder. The bulk data can be found here: https://www. nlm.nih.gov/databases/download/mesh.html.

[9]See https://www.nlm.nih.gov/mesh/intro_preface.html#pref_rem.

[10]The list of new MeSH Headings for 2020 published by the NLM is available here: https://www.nlm.nih.gov/ mesh/2020/download/2020NewMeSHheadingsSingleColumn.pdf.

research.[11] In 2003, the NLM added, for instance, the following MeSH terms to the vocabulary: *Retinoschisis* reflects a more detailed conceptualisation of an existing disease (Retinal Degeneration), *Severe Acute Respiratory Syndrome* describes a newly occurring disease related to the 2000s outbreaks of the SARS-Coronavirus. All together, the introduction of each term approximates the beginning of a new research field.

We measure new research opportunities K_{it} that are associated with an ICD-9 group by calculating the number of *new* MeSH terms that occur below the hierarchy level of our ICD-9-MeSH crosswalk, introduced in a given year t.[12] Since new research opportunities likely become obsolete over time, we add a discount factor of 0.8.[13] This approach is novel to the literature, which typically uses MeSH terms statically as keywords to understand shifts in the direction of science, but not the dynamic development of opportunities.

---

[11]See https://www.nlm.nih.gov/pubs/factsheets/mesh.html.

[12]We use the date of establishment since this is not sensitive to the transformation of the analogue MeSH vocabulary to the digital vocabulary in 1999. For more details on the variables, see https://www.nlm.nih.gov/mesh/xml_data_elements.html.

[13]Estimated depreciation rates of knowledge capital vary in the literature. Common values lie between 15% (Griliches, 1981; Hall and Mairesse, 1995) and 25% (Pakes and Schankerman, 1984). Our results are robust to applying different depreciation rates.

Evolution Alternative Controls

Figure C-1: OECD market size growth
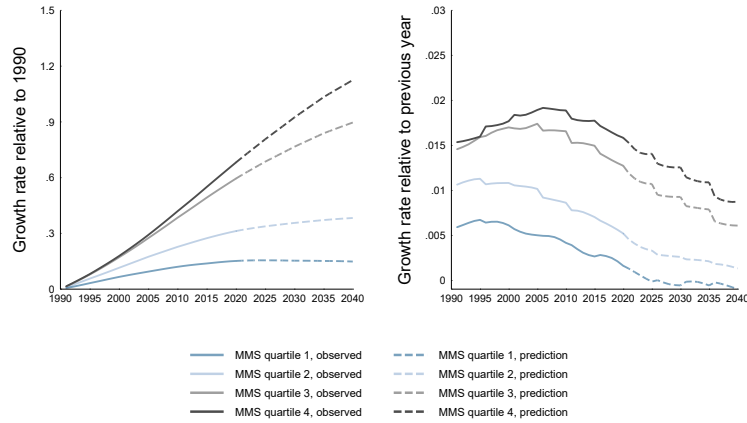
Notes: The left figure presents the annual OECD population-growth driven market size (in 2003 values) of each ICD-9 group, aggregated by MMS quartiles, and normalized in 1990. The right figure shows the annual change in OECD market size relative to the prior year.

Figure C-2: NIH funding trends (proportionally by the share of all publications)
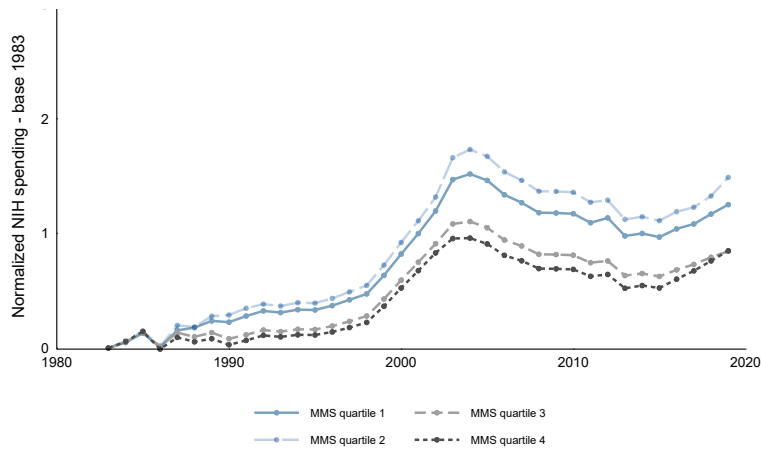
Notes: The figure presents the annual NIH spending (in 2003 values) attributed to each ICD-9 group, averaged by MMS quartiles, and normalized in 1990. We attribute NIH budgets based on the share of all publications in an ICD-9 group that acknowledge a specific Institute/Center.

C.2 Figures

C.2.1 Figures – Medicare Market Shares

Figure C-3: Distribution of Medicare market shares in 1997-2003

(a) Based on number of Medicare patients
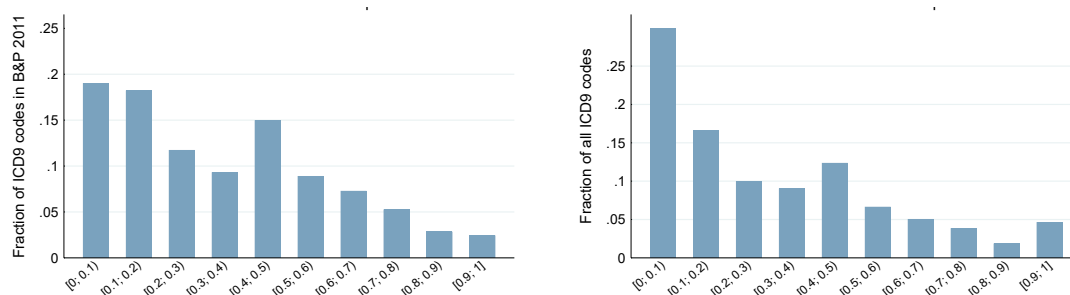
(b) Based on number of Medicare prescriptions
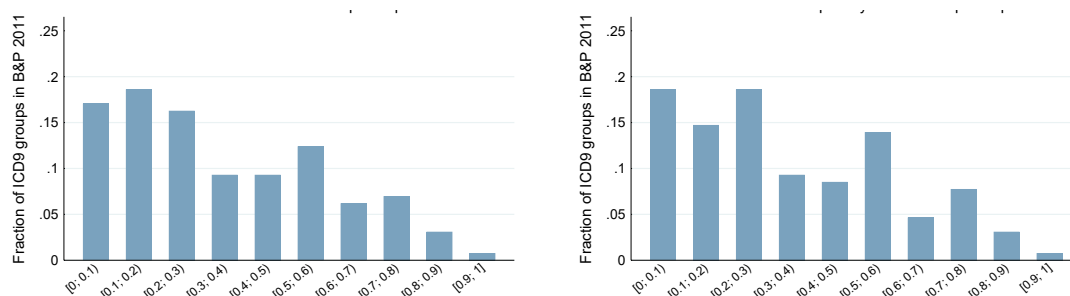
Notes: The figures in the top present the distribution of MMS scores among ICD-9 three-digit codes. Figure (a) shows all 272 ICD-9 three-digit codes, which are included in the MeSH-ICD-9 crosswalk (B&P2011 sample). Figure (b) shows all 752 ICD-9 three-digit codes in the 1997-2003 MEPS. We use the patient-weighted average of each year between 1997-2003. The annual MMS are calculated using the total number of patients in Medicare relative to all patients for each ICD-9 three-digit code. The figures in the bottom present the distribution of MMS scores among the 129 ICD-9 groups, which are included in the MeSH-ICD-9 crosswalk (B&P2011 sample). In Figure (c) we use the prescription count-weighted average of each year between 1997-2003. The annual MMS are calculated using the total number of prescriptions financed by Medicare relative to all prescriptions for each ICD-9 group. In Figure (d) we use the prescription quantity-weighted average of each year between 1997-2003. The annual MMS are calculated using the quantity-weighted prescriptions financed by Medicare relative to all quantity-weighted prescriptions for each ICD-9 group.

Figure C-4: Evolution of Medicare drug prescriptions by MMS quartiles
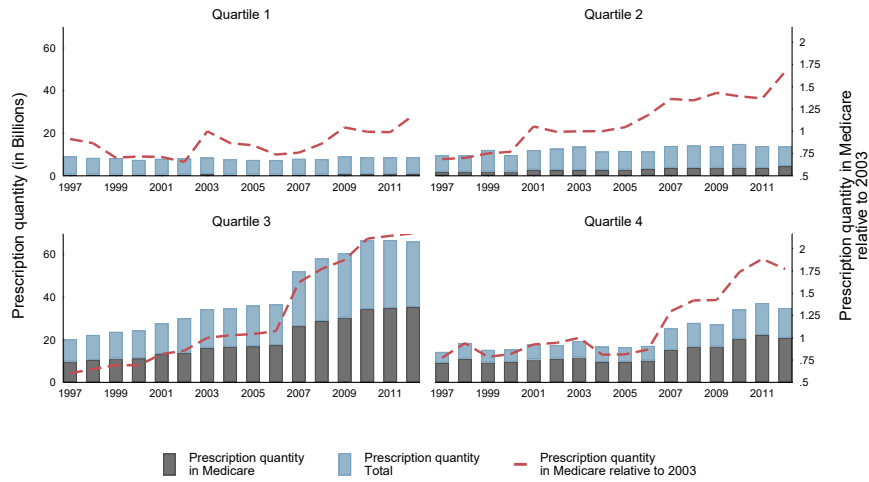
Notes: The figure shows the evolution of prescription quantities for each ICD-9 group aggregated by MMS quartiles. The grey bars display the quantity of prescriptions financed by Medicare, the blue bars display the non-Medicare prescription quantities. Prescriptions are counted multiple times if they appear in more than one ICD-9 group. The red line represents the relative increase in the quantity of Medicare prescriptions with respect to the baseline year 2003. MMS quartiles are based on the pre-2004 weighted average of patient-based MMS in order to be consistent with other outcome variables.

Figure C-5: Evolution of drug prices and total revenue

(a) Evolution of drug price indices
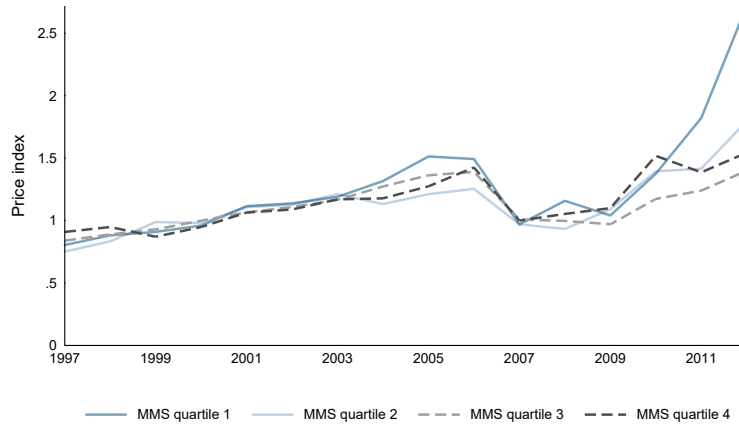
(b) Composition of total revenue
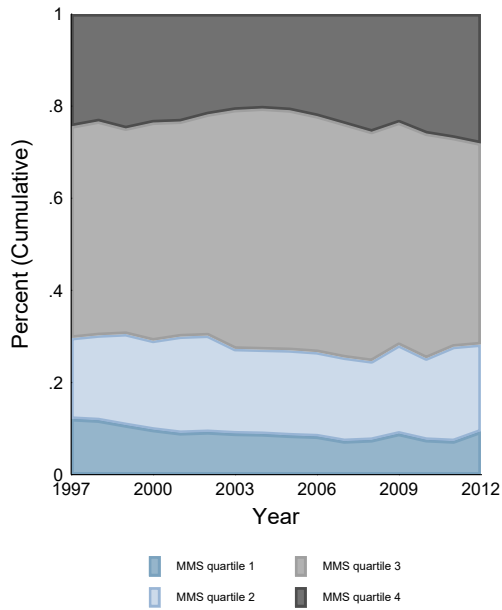
(c) Evolution of total revenue
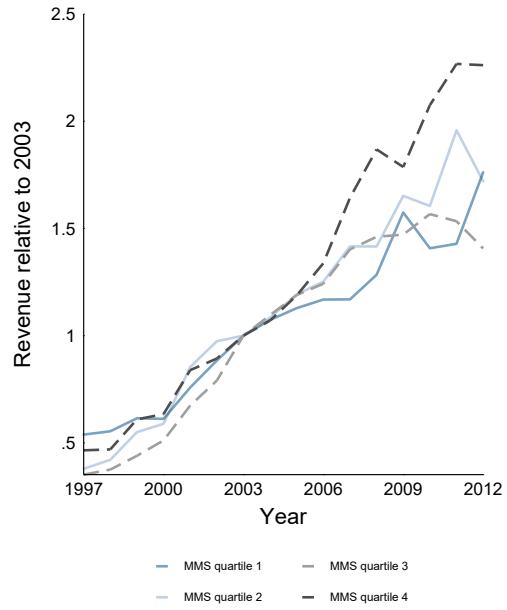
Notes: The top figure presents the evolution of a price index by MMS quartile. It is calculated based on a prescription quantity-weighted basket of drugs (1000 most sold drugs in total) following the procedure outlined in Duggan and Scott Morton (2010). Drug-level prices are inferred from the MEPS, winsorized, and imputed/extrapolated if missing. We drop a drug-form-disease combination if it does not appear in at least 2/3 of the sample periods, in 3 consecutive years, and exhibits price growth in the top 1% of the distribution. The index is set to one by using a 1997-2003 divisor. The bottom left figure presents the composition of total drug revenues by MMS quartile over time. Revenues are calculated based on the total payment of all drugs which are prescribed for a certain ICD-9 group. The bottom right figure shows the normalized evolution of total revenues. 2003 serves as the baseline year.

C.2.2 Figures – Descriptive Analysis

Figure C-6: Distribution of pre-MMA dependent/independent variables
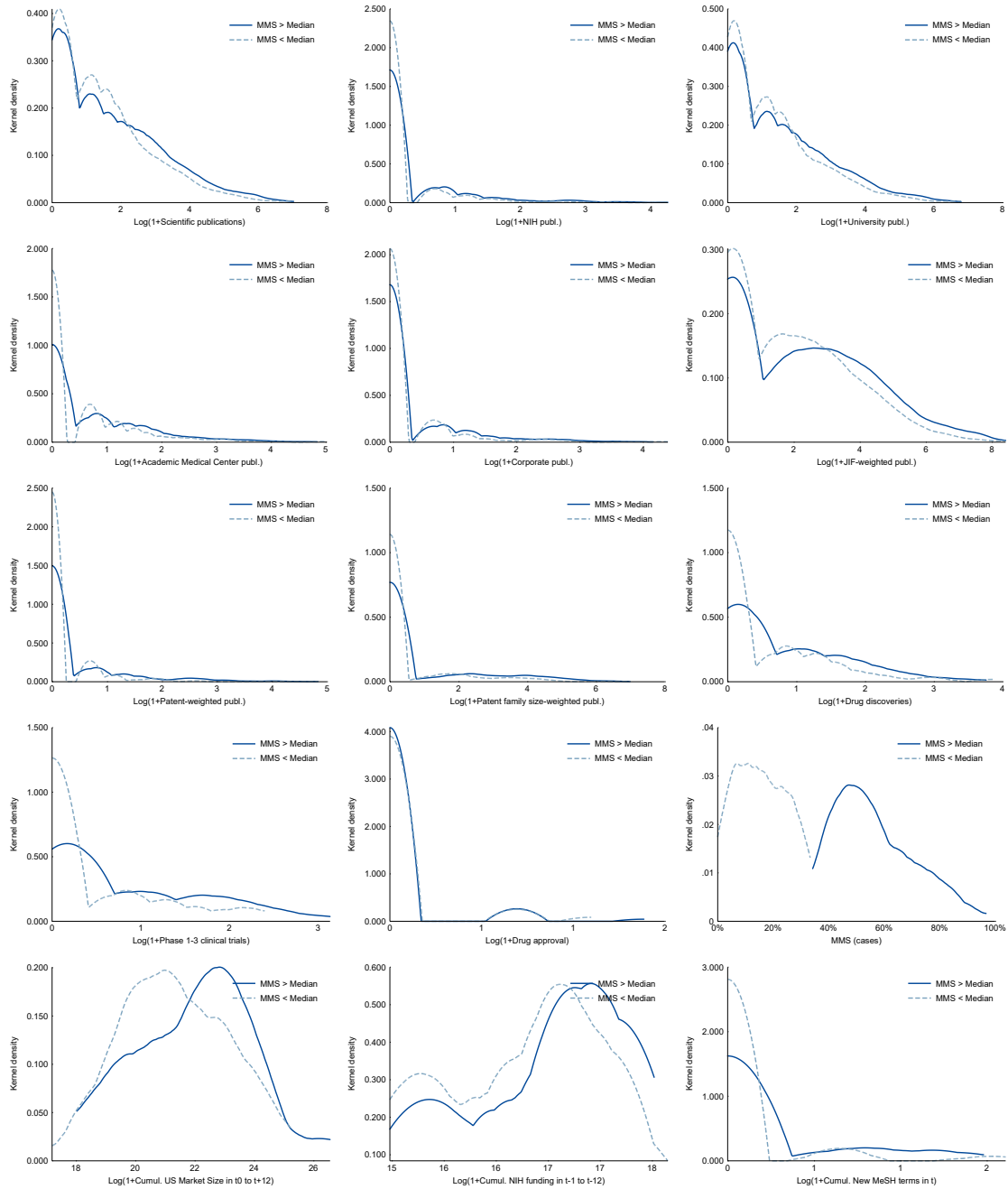
Notes: The figures compare the Kernel density of dependent and independent variables split at the Median MMS in the year 2003. The unit of observation is the MeSH level for publications related variables, the ICD-9 three-digit code level for drug related variables, and the ICD-9 group level for MMS/control variables.

Figure C-7: Trends in scientific publications and drug development by MMS

(a) NIH Publications

(b) University Publications

(c) Academic Medical Center Publications

(d) Patent-weighted Publications

(e) Phase I-III

(f) All Development Stages
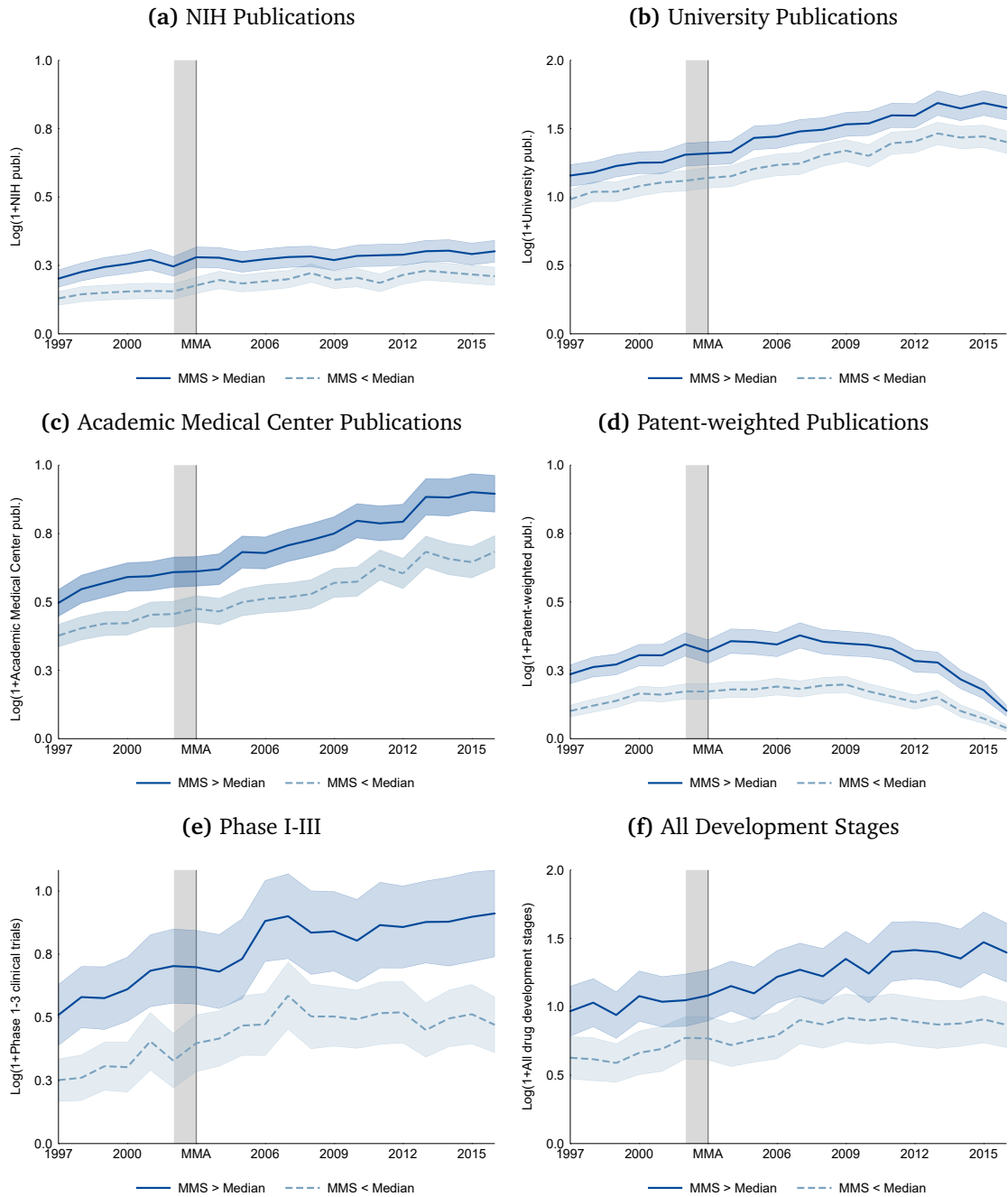
Notes: Figures (a)-(c) present the log-transformed average number of annual publication counts (a) with NIH participation, (b) with university participation, and (c) with academic medical center participation. Figure (d) presents the log-transformed average number of patent-weighted publications from all affiliations. Figures (e)-(f) present the log-transformed average number of annual NME (e) in phase I-III clinical trials and (f) in all drug development stages (preclinical, phase I-III clinial trials, registration, approval). In all graphs, the unit of observation is the unique ICD-9 group level.

C.2.3 Figures – Multivariate Analysis Clinical Drug Development

Figure C-8: Event study – drug development (alternative outcomes)

(a) Phase I-III clinical trials

(b) All development stages
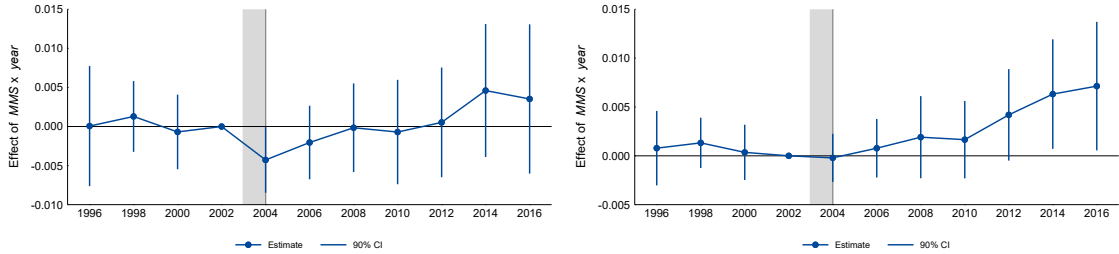
Notes: The figures show the event study estimates and the 90 percent confidence bands of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects following Equation 3.1. The unit of observation is the ICD-9 three-digit code level, with MMS being calculated based on patient counts at the ICD-9 group level. Standard errors are clustered at the level of treatment (ICD-9 group level).

Figure C-9: Event study – drug development (alternative MMS)

MMS based on prescription counts

(a) Drug discoveries
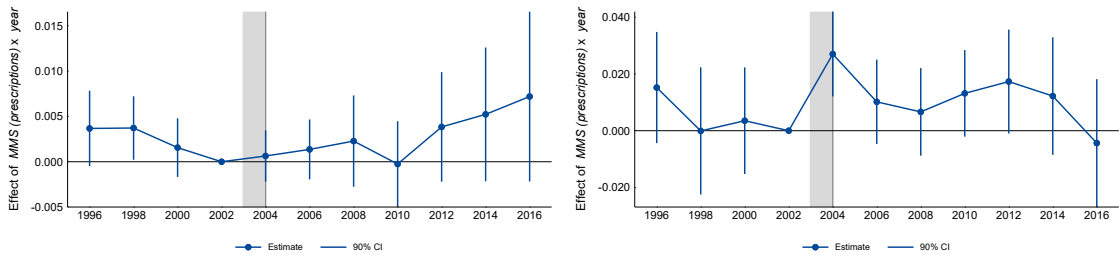
(b) Drug approvals

MMS based on prescription quantity

(c) Drug discoveries
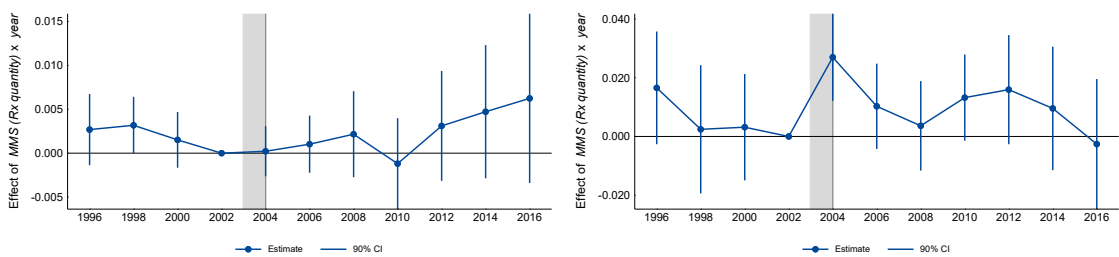
(d) Drug approvals

Notes: The figures show the event study estimates and the 90 percent confidence bands of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects following Equation 3.1. The unit of observation is the *ICD-9 three-digit code* level. In the top figures, the MMS is calculated based on prescription counts, in the bottom figures the MMS is calculated based on prescription quantity, both at the ICD-9 group level. Standard errors are clustered at the level of treatment (ICD-9 group level).

C.2.4 Figures – Multivariate Analysis Biomedical Science

Figure C-10: Event study – scientific publications (alternative outcomes)

(a) Publications (fractional)
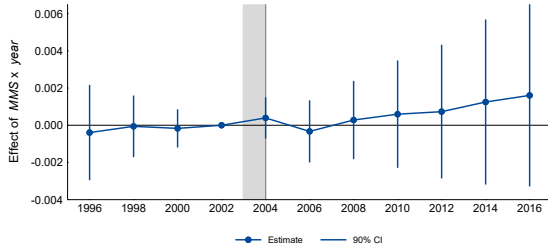
(b) Publications (winsorized)
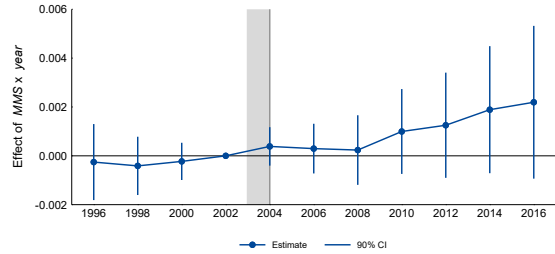
(c) University publications (fractional)
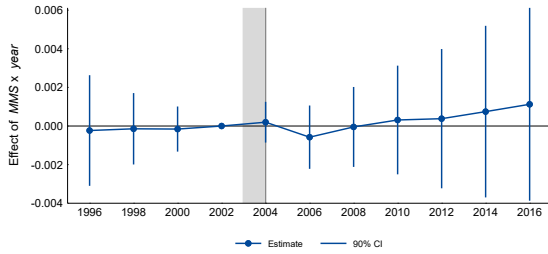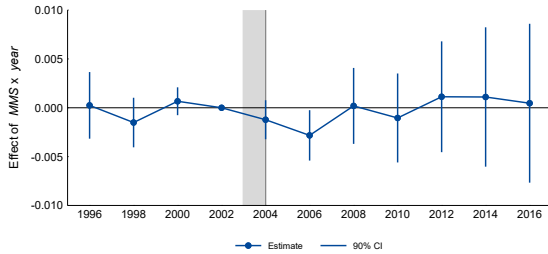
(d) University publications (winsorized)
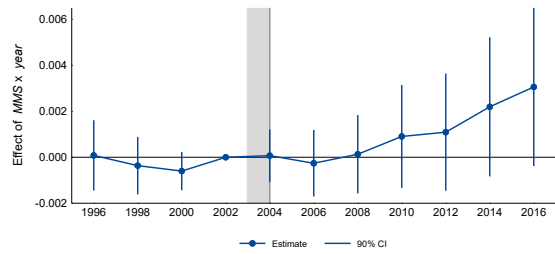
(e) Corporate publications (fractional)
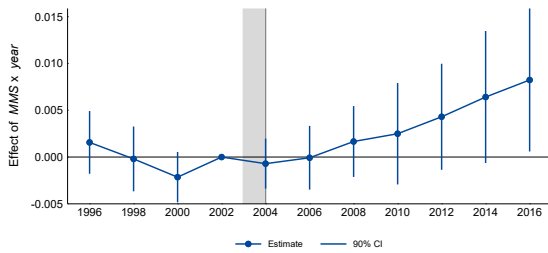
(f) Corporate publications (winsorized)
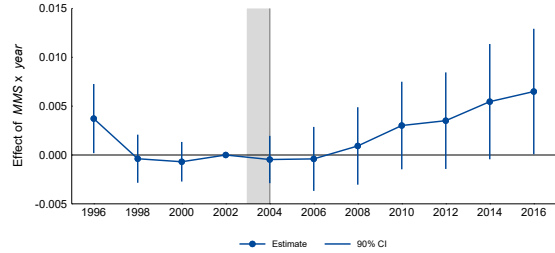
Notes: The figures show the event study estimates and the 90 percent confidence bands of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects following Equation 3.1. The unit of observation is the MeSH term level. In the left figures, the dependent variable is the annual number of scientific publications weighted by the number of distinct ICD-9 per publications (thus, counted fractional). In the right figures, the number of scientific publications is winsozired at the annual 99th percentile. Standard errors are clustered at the level of treatment (ICD-9 group level).

Figure C-11: Event study – scientific publications (alternative MMS)

MMS based on patient counts (2003 only)

(a) University publications
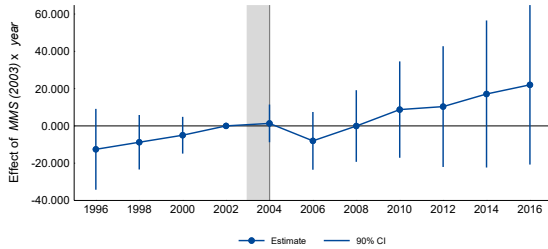
(b) Corporate publications
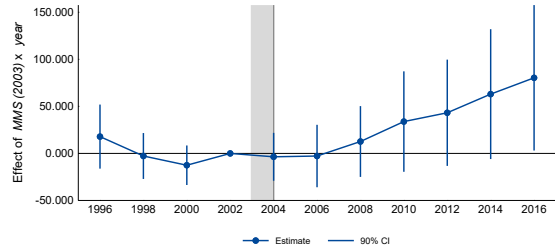
MMS based on prescription counts

(c) University publications
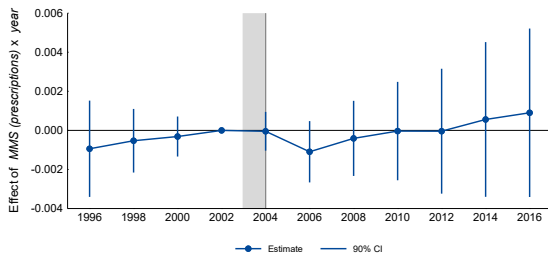
(d) Corporate publications
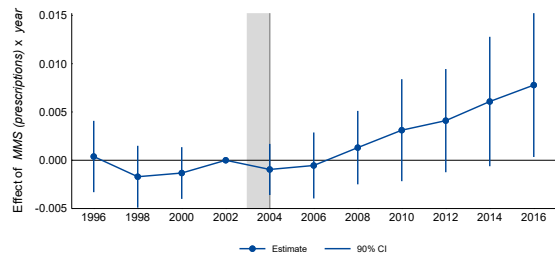
MMS based on prescription quantity

(e) University publications
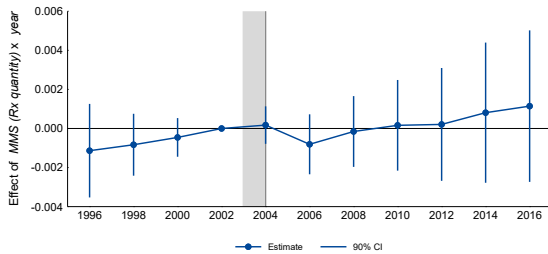
(f) Corporate publications
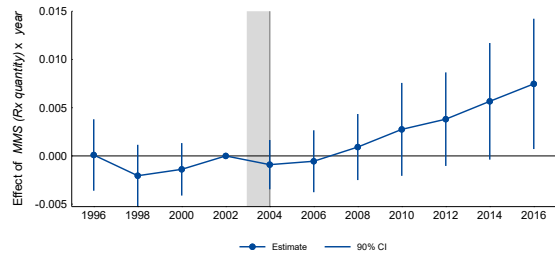
MMS based on patient counts (above median)

(g) University publications
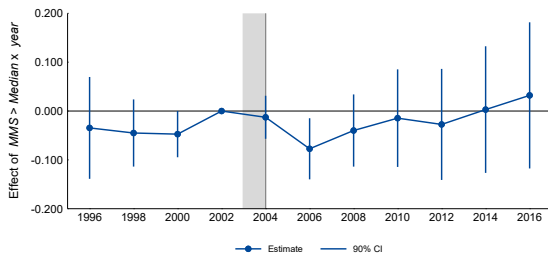
(h) Corporate publications
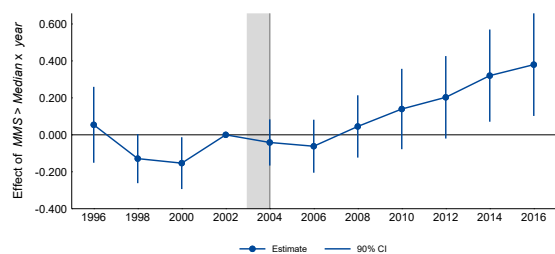
Notes: The figures show the event study estimates of Poisson pseudo-maximum likelihood regressions. The dependent variable is the annual number of scientific publications. In the top figures, the MMS is calculated based on 2003 patients counts. In the middle figures, the MMS is calculated based on prescription counts/quantity. In the bottom figures, the treatment variable is binary based on patients counts. Standard errors are clustered at the level of treatment (ICD-9 group level).

Figure C-12: Event study – scientific publications (protected classes)

Unprotected diseases
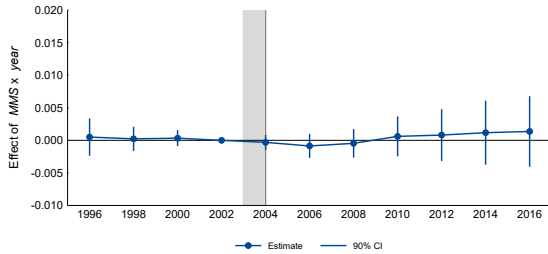
(a) University publications · · · · · · · · · · · **(b)** Corporate publications
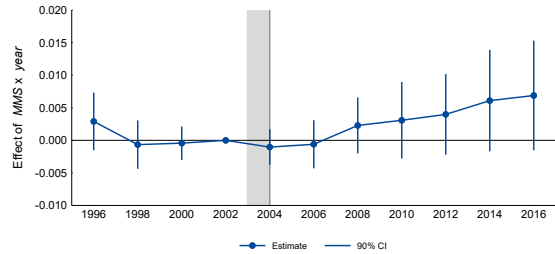
Protected diseases

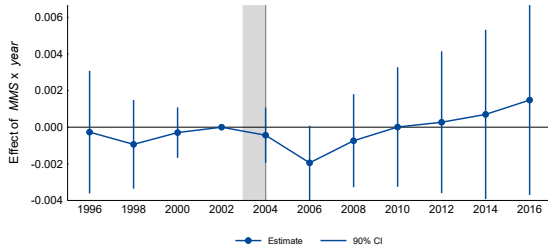**(c)** University publications · · · · · · · · · · · **(d)** Corporate publications
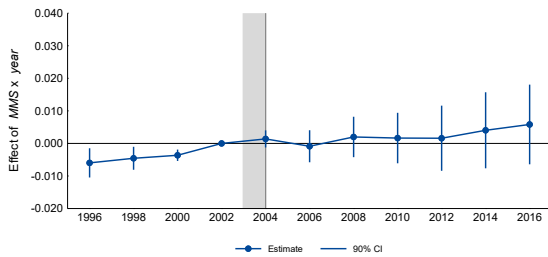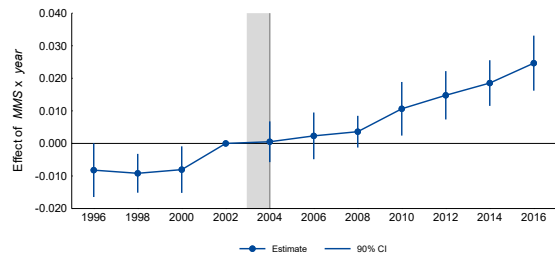
Notes: The figures show the event study estimates and the 90 percent confidence bands of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects following Equation 3.1. The unit of observation is the MeSH term level. The dependent variable is the annual number of scientific publications. In the top figures, we *exclude* all ICD-9 groups related to protected drug classes. In the bottom figures, we *include* only ICD-9 groups related to protected drug classes. Standard errors are clustered at the level of treatment (ICD-9 group level).

Figure C-13: Event study – scientific publications by type of research

Clinical trial publications

(a) University publications   **(b)** Corporate publications
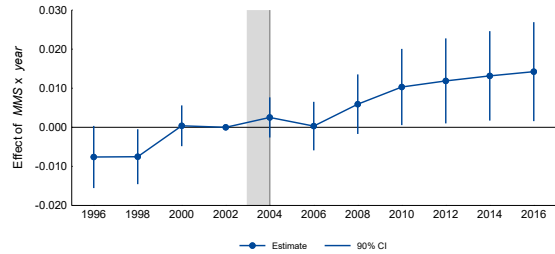
Pharmaceutical publications
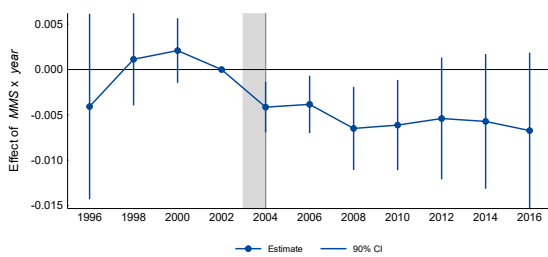
(c) University publications   **(d)** Corporate publications
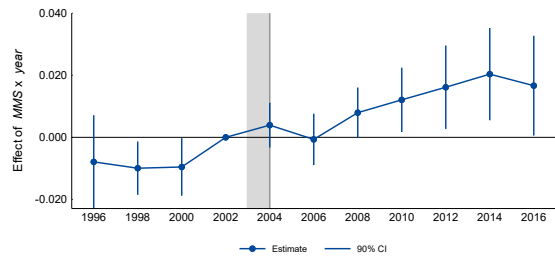
Basic science publications

(e) University publications   **(f)** Corporate publications
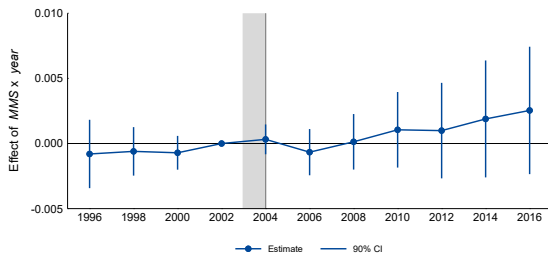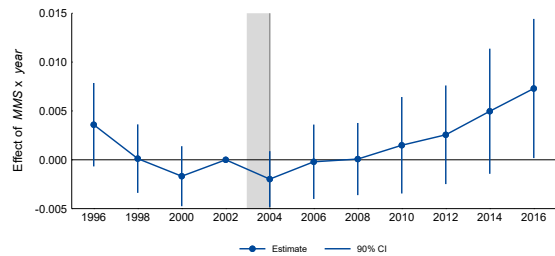
Notes: The figures show the event study estimates and the 90 percent confidence bands of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects following Equation 3.1. The unit of observation is the MeSH term level. In the top figures, the dependent variable is the annual number of scientific publications, which are associated with a MeSH term related to "clinical trials". In the middle figure, the dependent variable is the annual number of scientific publications, which are associated with a MeSH term related to "pharmaceutical preparations". In the bottom figures, the dependent variables is the residual, thus scientific publications neither related to clinical trials nor pharmaceutical products. Standard errors are clustered at the level of treatment (ICD-9 group level).
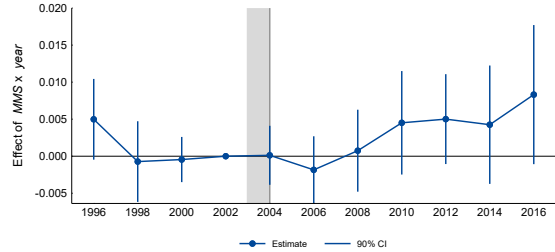
Figure C-14: Event study – weighted scientific publications

JIF-weighted counts
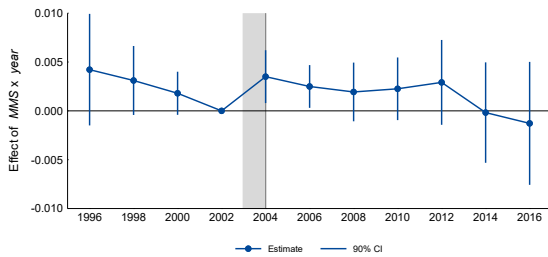
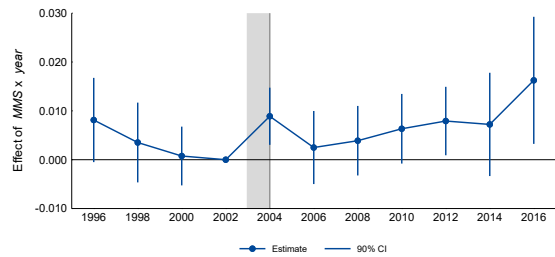**(a)** University publications   **(b)** Corporate publications

Patent-weighted counts

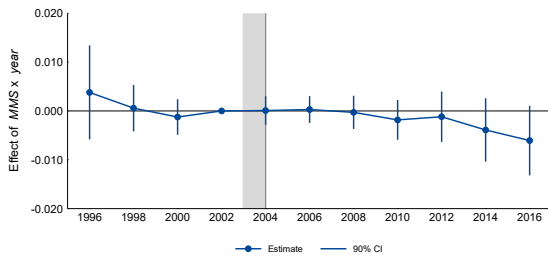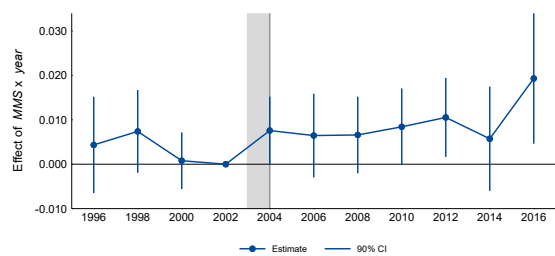**(c)** University publications   **(d)** Corporate publications

Family size-weighted counts

(e) University publications   **(f)** Corporate publications

Notes: The figures show the event study estimates and the 90 percent confidence bands of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects following Equation 3.1. The unit of observation is the MeSH term level. In the top figures, the dependent variable is the journal impact factor-weighted number of university/corporate scientific publications. In the middle figure, the dependent variable is the number of university/corporate scientific publications that are associated with at least one patent application (patent-weighted). In the bottom figures, we weight the number scientific publications by the size of the average patent family associated with the publication. A patent/family size-weight is calculated based on the patent family's first application being filed within 5 years from the scientific publication. Standard errors are clustered at the level of treatment (ICD-9 group level).

C.3 Tables

C.3.1 Tables – Data Overview

Table C-1: Updated ICD-9/MeSH crosswalk (based on B&P2011)

ICD group	ICD code	ICD entry/entries	MeSH entry/entries	MeSH ID
Infectious and parasitic diseases				
011	011	Pulmonary tuberculosis	Tuberculosis, Pulmonary	D014397
034	034	Streptococcal sore throat/scarlet fever	Scarlet Fever	D012541
042	042	Human immunodeficiency virus disease	HIV Infections	D015658
052	052	Chickenpox	Chickenpox	D002644
053	053	Herpes zoster	Herpes Zoster	D006562
054	054	Herpes simplex	Herpes Simplex	D006561
070	070	Viral hepatitis	Hepatitis	D006505
075	075	Infectious mononucleosis	Infectious Mononucleosis	D007244
110	110	Dermatophytosis	Tinea	D014005
	111	Dermatomycosis (unspecified)	Tinea Versicolor	D014010
112	112	Candidiasis	Candidiasis	D002177
132	132	Pediculosis/phthirus infestation	Lice Infestations	D010373
133	133	Acariasis	Mite Infestations	D008924
No Match	038	Septicemia		
	074	Specific diseases due to Coxsackie virus		
Neoplasms				
150	150-159	Malignant neoplasm digestive organs	Digestive System Neoplasms	D004067
	211	Benign neoplasm digestive system	Abdominal Neoplasms	D000008
			Anal Gland Neoplasms	D000694
162	162	Malignant neoplasm bronchus/lung	Respiratory Tract Neoplasms	D012142
	163	Malignant neoplasm pleura		
171	171	Malignant melanoma skin	Soft Tissue Neoplasms	D012983
	214	Lipoma		
	215	Benign neoplasm connective tissue		
172	172	Malignant melanoma skin	Skin Neoplasms	D012878
	173	Malignant neoplasm skin		
	216	Benign neoplasm skin		
174	174	Malignant neoplasm female breast	Breast Neoplasms	D001943
	175	Malignant neoplasm male breast		
	217	Benign neoplasm breast		
179	179	Malignant neoplasm uterus	Genital Neoplasms, Female	D005833
	180	Malignant neoplasm cervix uteri	Genital Neoplasms, Male	D005834
	181	Malignant neoplasm placenta	Urologic Neoplasms	D014571
	182	Malignant neoplasm body of uterus		
	183	Malignant neoplasm ovary		
	184	Malignant neoplasm female genitals		
	218	Uterine leiomyoma		
	219	Benign neoplasm uterus		
	220	Benign neoplasm ovary		
	221	Benign neoplasm female genitals		

ICD group	ICD code	ICD entry/entries	MeSH entry/entries	MeSH ID
	185	Malignant neoplasm prostate		
	186	Malignant neoplasm testis		
	187	Malignant neoplasm penis/male genitals		
	222	Benign neoplasm male genital organs		
	188	Malignant neoplasm bladder		
	189	Malignant neoplasm kidney		
	223	Benign neoplasm kidney		
200	200-208	Malignant neoplasm lymphatic tissue	Leukemia	D007938
			Lymphoma	D008223
230	230-234	Carcinoma in situ	Carcinoma in Situ	D002278
Endocrine, nutritional and metabolic diseases, and immunity disorders				
240	240	Simple goiter	Goiter	D006042
	241	Nontoxic nodular goiter		
242	242	Thyrotoxicosis with/without goiter	Hyperthyroidism	D006980
243	243	Congenital hypothyroidism	Hypothyroidism	D007037
	244	Acquired hypothyroidism		
250	250	Diabetes mellitus	Diabetes Mellitus	D003920
265	265	Thiamine/niacin deficiency states	Vitamin B Deficiency	D014804
	266	Deficiency B-complex components		
272	272	Disorders lipoid metabolism	Lipid Metabolism Disorders	D052439
274	274	Gout	Gout	D006073
275	275	Disorders mineral metabolism	Hemochromatosis	D006432
			Hepatolenticular Degeneration	D006527
			Hypophosphatemia, Familial	D007015
			Hypercalcemia	D006934
			Hypocalcemia	D006996
276	276	Disorders acid-base balance	Hypokalemia	D007008
			Hypernatremia	D006955
			Acidosis	D000138
			Alkalosis	D000471
279	279	Disorders immune mechanism	Agammaglobulinemia	D000361
			DiGeorge Syndrome	D004062
			Dysgammaglobulinemia	D004406
			Wiskott-Aldrich Syndrome	D014923
No Match	256	Ovarian dysfunction		
Diseases of blood and blood-forming organs				
280	280	Iron deficiency anemias	Anemia	D000740
	281	Deficiency anemias		
	282	Hereditary hemolytic anemias		
	283	Acquired hemolytic anemias		
	284	Aplastic anemia		
	285	Anemias		
288	288	Diseases white blood cells	Agranulocytosis	D000380
			Granulomatous Disease, Chronic	D006105
			Eosinophilia	D004802
			Leukocytosis	D007964
Mental disorders				
295	295	Schizophrenic psychoses	Schizophrenia	D012559

ICD group	ICD code	ICD entry/entries	MeSH entry/entries	MeSH ID
296	296	Affective psychoses	Mood Disorders	D019964
	309	Adjustment reaction	Adjustment Disorders	D000275
299	299	Psychoses with origin in childhood	Child Development Disorders	D002659
300	300	Neurotic disorders	Anxiety Disorders	D001008
			Dissociative Disorders	D004213
			Feeding/Eating Disorders	D001068
			Somatoform Disorders	D013001
301	301	Personality disorders	Personality Disorders	D010554
302	302	Sexual deviations/disorders	Sexual and Gender Disorders	D019968
303	303	Alcohol dependence syndrome	Substance-Related Disorders	D019966
	304	Drug dependence		
	305	Nondependent drug abuse		
314	314	Hyperkinetic syndrome (childhood)	Attention Deficit Disorder	D001289
315	315	Specific delays in development	Developmental Disabilities	D002658
			Communication Disorders	D003147
No Match	308	Acute reaction to stress		
	306	Physiological malfunction		
Diseases of the nervous system and sense organs				
320	320	Bacterial meningitis	Meningitis	D008581
	321	Meningitis (other organisms)	Central Nervous System - Viral Diseases	D020805
	322	Meningitis	Myelitis	D009187
	323	Encephalitis/myelitis/encephalomyelitis		
331	331	Cerebral degenerations (Alzheimer's disease)	Alzheimer Disease	D000544
332	332	Parkinson's disease	Parkinsonian Disorders	D020734
340	340	Multiple sclerosis	Multiple Sclerosis	D009103
343	343	Infantile cerebral palsy	Cerebral Palsy	D002547
345	345	Epilepsy	Epilepsy	D004827
346	346	Migraine	Migraine Disorders	D008881
350	350-359	Disorders peripheral nervous system	Peripheral Nervous - System Diseases	D010523
		Disorders peripheral nervous system		
361	361	Retinal detachments/defects	Retinal Diseases	D012164
	362	Retinal disorders		
363	360	Disorders of the globe	Uveal Diseases	D014603
363	363	Chorioretinal inflammations/scars		
363	364	Disorders iris/ciliary body		
365	365	Glaucoma	Glaucoma	D005901
366	366	Cataract	Cataract	D002386
367	367	Disorders of refraction	Refractive Errors	D012030
368	368	Visual disturbances	Vision Disorders	D014786
	369	Blindness/low vision		
371	371	Corneal opacity/disorders of cornea	Corneal Diseases	D003316
372	372	Disorders conjunctiva	Conjunctival Diseases	D003229
373	373	Inflammation eyelids	Eyelid Diseases	D005141

ICD group	ICD code	ICD entry/entries	MeSH entry/entries	MeSH ID
	374	Disorders eyelids		
375	375	Disorders lacrimal system	Lacrimal Apparatus Diseases	D007766
380	380	Disorders external ear	Otitis Externa	D010032
381	381	Nonsuppurative otitis media	Otitis Media	D010033
	382	Suppurative otitis media		
	383	Mastoiditis/related conditions		
386	386	Vertiginous synd. vestibular system	Labyrinth Diseases	D007759
389	389	Hearing loss	Hearing Loss	D034381
Diseases of the circulatory system				
401	401-405	Hypertensive disease	Hypertension	D006973
410	410	Acute myocardial infarction	Myocardial Infarction	D009203
	412	Old myocardial infarction		
413	413	Angina pectoris	Angina Pectoris	D000787
414	414	Chronic ischemic heart disease	Arteriosclerosis	D001161
	440	Atherosclerosis	Aneurysm	D000783
	441	Aortic aneurysm/dissection		
	442	Aneurysm		
426	426	Conduction disorders	Arrhythmias, Cardiac	D001145
	427	Cardiac dysrhythmias		
428	428	Heart failure	Heart Failure	D006333
430	430-438	Cerebrovascular disease	Cerebrovascular Disorders	D002561
444	444	Arterial embolism/thrombosis	Embolism and Thrombosis	D016769
	451	Phlebitis/thrombophlebitis	Phlebitis	D010689
	452	Portal vein thrombosis		
	453	Venous embolism/thrombosis		
454	454	Varicose veins lower extremities	Varicose Veins	D014648
	456	Varicose veins other sites		
455	455	Hemorrhoids	Hemorrhoids	D006484
458	458	Hypotension	Hypotension	D007022
Diseases of the respiratory system				
460	460	Acute nasopharyngitis	Nasopharyngitis	D009304
	462	Acute pharyngitis	Pharyngitis	D010612
	472	Chronic pharyngitis/nasopharyngitis		
461	461	Acute sinusitis	Sinusitis	D012852
	473	Chronic sinusitis		
463	463	Acute tonsillitis	Tonsillitis	D014069
	474	Chronic disease tonsils/adenoids		
464	464	Acute laryngitis/tracheitis	Laryngitis	D007827
	476	Chronic laryngitis/laryngotracheitis	Tracheitis	D014136
			Epiglottitis	D004826
			Croup	D003440
466	466	Acute bronchitis/bronchiolitis	Bronchitis	D001991
	490	Bronchitis		
	491	Chronic bronchitis		
477	477	Allergic rhinitis	Rhinitis	D012220

ICD group	ICD code	ICD entry/entries	MeSH entry/entries	MeSH ID
480	480	Viral pneumonia	Pneumonia	D011014
	481	Pneumococcal pneumonia		
	482	Bacterial pneumonia		
	483	Pneumonia (other specified organism)		
	484	Pneumonia in infectious diseases		
	485	Bronchopneumonia		
	486	Pneumonia		
	514	Pulmonary congestion/hypostasis		
487	487	Influenza	Influenza, Human	D007251
492	492	Emphysema	Emphysema	D004646
493	493	Asthma	Asthma	D001249
511	511	Pleurisy	Pleurisy	D010998
No Match	470	Deviated nasal septum		
Diseases of the digestive system				
520	520	Disorders tooth development	Tooth Abnormalities	D014071
	521	Diseases hard tissues of teeth	Tooth Erosion	D014077
	524	Dentofacial anomalies	Tooth Abrasion	D014072
			Malocclusion	D008310
522	522	Diseases pulp/periapical tissues	Periapical Diseases	D010483
	523	Gingival/periodontal diseases	Dental Pulp Diseases	D003788
			Periodontitis	D010518
			Gingival Diseases	D005882
526	526	Diseases jaws	Jaw Cysts	D007570
			Granuloma, Giant Cell	D006101
527	527	Diseases salivary glands	Salivary Gland Diseases	D012466
528	528	Diseases oral soft tissues	Stomatitis	D013280
			Noma	D009625
530	530	Diseases esophagus	Esophageal Diseases	D004935
531	531	Gastric ulcer	Peptic Ulcer	D010437
	532	Duodenal ulcer	Peptic Ulcer Hemorrhage	D010438
	533	Peptic ulcer	Gastrointestinal Hemorr.	D006471
	534	Gastrojejunal ulcer		
	578	Gastrointestinal hemorrhage		
535	535	Gastritis/duodenitis	Gastritis	D005756
	555-558	Noninfective enteritis/colitis	Duodenitis	D004382
			Enteritis	D004751
			Colitis	D003092
536	536	Disorders function of stomach	Achlorhydria	D000126
			Gastric Dilatation	D013271
			Dyspepsia	D004415
540	540	Acute appendicitis	Appendicitis	D001064
	541	Appendicitis, unqualified		
	542	Appendicitis		
550	550-553	Hernia of abdominal cavity	Hernia	D006547
560	560	Intestinal obstruction	Intestinal Obstruction	D007415
562	562	Diverticula of intestine	Diverticulum, Colon	D004241
			Diverticulum, Stomach	D013273

ICD group	ICD code	ICD entry/entries	MeSH entry/entries	MeSH ID
574	574	Cholelithiasis	Cholelithiasis	D002769
577	577	Diseases pancreas	Pancreatitis Pancreatic Cyst	D010195 D010181
No Match	571	Chronic liver disease/cirrhosis		
Diseases of the genitourinary system				
590	590	Infections kidney	Nephritis	D009393
592	592	Calculus kidney/ureter	Nephrolithiasis Ureterolithiasis	D053040 D053039
595	595	Cystitis	Cystitis	D003556
600	600 601 602	Hyperplasia prostate Inflammatory diseases prostate Disorders prostate	Prostatic Diseases	D011469
607	607	Disorders penis	Penile Diseases	D010409
610	610 611	Benign mammary dysplasias Disorders breast	Breast Diseases	D001941
614	614 620	Inflammatory disease ovary Noninflammatory disorders ovary	Adnexal Diseases	D000291
615	615 616 618 621 622 623 624 625	Inflammatory diseases uterus Inflammatory disease cervix/vagina/vulva Genital prolapse Disorders uterus Noninflammatory disorders cervix Noninflammatory disorders vagina Noninflammatory disorders vulva/perineum Pain associated with female genital organs	Uterine Diseases Vaginal Diseases Vulvar Diseases	D014591 D014623 D014845
617	617	Endometriosis	Endometriosis	D004715
628	628	Infertility, female	Infertility, Female	D007247
No Match	627	Menopausal/postmenopausal disorders		
Diseases of the skin and subcutaneous tissue				
680	680	Carbuncle/furuncle	Furunculosis	D005667
681	681 682	Cellulitis/abscess finger/toe Cellulitis/abscess	Cellulitis	D002481
684	684	Impetigo	Impetigo	D007169
690	690 706	Erythematosquamous dermatosis Diseases sebaceous glands	Dermatitis, Seborrheic Acne Vulgaris	D012628 D000152
691	691 692	Atopic dermatitis/related conditions Contact dermatitis/eczema	Dermatitis, Atopic Dermatitis, Contact	D003876 D003877
696	696	Psoriasis/similar disorders	Psoriasis Pityriasis Parapsoriasis	D011565 D010915 D010267
698	698	Pruritus/related conditions	Pruritus Prurigo Neurodermatitis	D011537 D011536 D009450
700	700	Corns/callosities	Callosities	D002145
703	703	Diseases nail	Nail Diseases	D009260

ICD group	ICD code	ICD entry/entries	MeSH entry/entries	MeSH ID
704	704	Diseases hair/hair follicles	Hair Diseases	D006201
705	705	Disorders sweat glands	Sweat Gland Diseases	D013543
708	708	Urticaria	Urticaria	D014581
No Match	707	Chronic ulcer skin		
	695	Erythematous conditions		
	693	Dermatitis (substances taken internally)		
Diseases of the musculoskeletal system and connective tissue				
710	710	Diffuse diseases connective tissue	Sjogrens Syndrome	D012859
	728	Disorders muscle/ligament/fascia	Scleroderma, Systemic	D012595
			Scleroderma, Localized	D012594
			Dermatomyositis	D003882
			Myositis	D009220
715	715	Osteoarthrosis/allied disorders	Osteoarthritis	D010003
	721	Spondylosis/allied disorders		
722	722	Intervertebral disc disorders	Intervertebral Disk Displac.	D007405
726	726	Peripheral enthesopathies	Bursitis	D002062
734	734	Flat foot	Flatfoot	D005413
735	735	Acquired deformities toe	Hallux Valgus	D006215
			Hallux Varus	D050488
737	737	Curvature spine	Spinal Curvatures	D013121
No Match	717	Internal derangement knee		

Table C-2: Summary statistics by high/low MMS

High MMS vs Low MMS	High MMS (N = 778)			Low MMS (N = 785)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Mean	Median	Std. Error	Mean	Median	Std. Error	Diff.	p-value
MMS (cases)	56.22	54.73	14.27	18.29	18.38	9.22	−37.94	0.0000***
MMS (prescriptions)	58.41	61.08	16.01	20.40	18.28	10.60	−38.02	0.0000***
MMS (Rx quantity)	59.23	58.15	17.49	21.60	19.48	12.26	−37.64	0.0000***
Scientific publications	19.16	2.00	69.47	14.59	2.00	69.17	−4.57	0.1922
Publications - fractional	14.85	1.33	60.16	11.38	1.00	57.72	−3.47	0.2454
University publications	15.09	2.00	55.31	11.48	1.00	54.41	−3.61	0.1940
Corporate publications	1.04	0.00	4.25	0.79	0.00	4.56	−0.25	0.2624
Academic Medical Center publications	2.63	0.00	8.63	1.97	0.00	8.63	−0.66	0.1299
NIH publications	1.01	0.00	4.32	0.63	0.00	3.94	−0.38	0.0678*
Academic journal publications	0.17	0.00	0.72	0.26	0.00	1.04	0.09	0.0564*
Clinical journal publications	7.41	1.00	26.88	6.36	1.00	26.68	−1.06	0.4359
Clinical-practice journal publications	4.08	0.00	13.16	2.34	0.00	9.90	−1.74	0.0032***
Industry-Clinical journal publications	4.96	0.00	21.49	3.35	0.00	22.34	−1.61	0.1471
Industrial journal publications	0.44	0.00	2.45	0.23	0.00	1.93	−0.20	0.0697*
Industry-practice journal publications	0.17	0.00	0.83	0.09	0.00	0.66	−0.08	0.0387**
NIH funded publications	7.31	0.00	30.65	6.03	0.00	34.94	−1.28	0.4405
Clinical trial university publications	1.25	0.00	5.20	1.28	0.00	6.25	0.02	0.9399
Clinical trial corporate publications	0.21	0.00	0.93	0.21	0.00	1.35	−0.01	0.9222
Pharmaceutical university publications	0.31	0.00	1.50	0.35	0.00	1.95	0.04	0.6346
Pharmaceutical corporate publications	0.07	0.00	0.39	0.07	0.00	0.52	0.00	0.8903
Citation-weighted publications	429.17	19.50	1777.20	278.17	15.00	1503.38	−151.01	0.0698*
JIF-weighted publications	76.29	5.33	285.15	45.61	3.34	245.43	−30.68	0.0227**
Patent-weighted publications	1.47	0.00	7.05	0.63	0.00	5.32	−0.83	0.0085***
Patent-weighted university publications	1.15	0.00	5.57	0.51	0.00	4.33	−0.64	0.0118**
Patent-weighted corporate publications	0.17	0.00	0.82	0.09	0.00	0.68	−0.09	0.0224**
Patent family size-weighted publications	12.79	0.00	59.95	5.59	0.00	45.87	−7.20	0.0077***
Cumul. US Market Size$_{t \text{ to } t+12}$	28 700.49	6105.04	64 374.47	10 778.88	3040.69	20 202.34	−17 921.61	0.0000***
Cumul. NIH funding$_{t-1 \text{ to } -12}$	22.57	21.46	12.50	12.75	11.98	8.21	−9.82	0.0000***
Cumul. New MeSH terms$_t$	0.83	0.00	1.01	0.30	0.00	0.66	−0.53	0.0000***

Notes: This table compares observations split at the MMS Median with t-tests. The unit of observation is the MeSH terms in 2003 year. Significance levels:
* p<0.1, ** p<0.05, *** p<0.01.

C.3.2 Tables – Multivariate Analysis Clinical Drug Development

Table C-3: Drug development – parallel trends

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)
		Drug Development		
	Early	All Stages	Phase 1-3	Approval
MMS × 2000-02	−0.0014	−0.0008	−0.0017	−0.0035
	(0.002)	(0.002)	(0.003)	(0.009)
ICD9 group FE	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes
Observations	1002	1044	918	648
ICD9-codes	167	174	153	108
ICD9-groups	93	100	83	52
Log-likelihood	−1633	−2457	−1368	−270

Notes: Columns (1) to (4) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The time period of these regressions is the pre-MMA period 1997-2002. We employ a placebo test (2000-2002 x MMS with 1997-1999 as the baseline period) in order to test the parallel trends assumption. The dependent variable is the annual number of newly discovered NME in Column (1), NME in all development stages in Column (2), NME in phase I-III clinical trials in Column (3), and the annual number of approved drugs in Column (4). The unit of observation is the ICD-9 three-digit level by year. Standard errors are clustered at the level of treatment (ICD-9 group level) and shown in parentheses. Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table C-4: Drug development – all ICD-9 three-digit codes

All ICD9 codes Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
			Drug Development			
	Early Development		All Development Stages		Approval	
MMS (ICD9 level) × 2004-05	−0.0005	−0.0002	0.0005	0.0006	0.0195***	0.0246***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.006)	(0.008)
MMS (ICD9 level) × 2006-08	0.0007	0.0019	0.0001	0.0009	0.0052	0.0041
	(0.002)	(0.002)	(0.002)	(0.002)	(0.005)	(0.007)
MMS (ICD9 level) × 2009-11	0.0015	0.0036	0.0006	0.0019	0.0112**	0.0060
	(0.002)	(0.002)	(0.002)	(0.002)	(0.005)	(0.008)
MMS (ICD9 level) × 2012-14	0.0032	0.0076***	0.0018	0.0049*	0.0066	0.0104
	(0.002)	(0.003)	(0.002)	(0.003)	(0.005)	(0.009)
MMS (ICD9 level) × 2015-16	0.0051*	0.0119***	0.0037	0.0085**	0.0104**	0.0074
	(0.003)	(0.004)	(0.003)	(0.004)	(0.005)	(0.010)
Cumul. US Market Size$_{t \text{ to } t+12}$	No	Yes	No	Yes	No	Yes
Cumul. NIH funding$_{t-1 \text{ to } -12}$	No	Yes	No	Yes	No	Yes
Cumul. New MeSH ids$_t$	No	Yes	No	Yes	No	Yes
ICD9 code FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6980	3340	7740	3740	4260	2380
ICD9-codes	349	167	387	187	213	119
Log-likelihood	−9333	−4785	−12764	−6620	−2067	−1183

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. Column (1), (3), and (5) include all ICD-9 three-digit codes available in Cortellis. Column (2), (4), and (6) include all ICD-9 three-digit codes available in the ICD-9-MeSH crosswalk by B&P2011. In both cases, the MMS is calculated at the ICD-9 three-digit code level. The unit of observation of the dependent variable is the ICD-9 three-digit code by year. It is the annual number of newly discovered NME in Columns (1) and (2), NME in all development stages (preclinicals, clinical trials, registrations, approvals) in Columns (3) and (4), and approved NME in Column (5) and (6). The control variables are log-transformed. Standard errors are clustered at the level of treatment (in this table, ICD-9 three-digit code level) and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table C-5: Drug development – alternative controls

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
			Drug Development			
	Early Development		All Develop. Stages		Approval	
MMS × 2004-05	−0.0004	−0.0006	−0.0010	−0.0007	0.0275***	0.0241***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.009)	(0.008)
MMS × 2006-08	0.0019	0.0011	0.0004	0.0004	0.0098	0.0046
	(0.003)	(0.003)	(0.003)	(0.002)	(0.008)	(0.007)
MMS × 2009-11	0.0029	0.0012	0.0008	0.0004	0.0179*	0.0096
	(0.004)	(0.002)	(0.004)	(0.003)	(0.010)	(0.007)
MMS × 2012-14	0.0063	0.0039	0.0035	0.0027	0.0253**	0.0140
	(0.004)	(0.003)	(0.004)	(0.003)	(0.012)	(0.009)
MMS × 2015-16	0.0099*	0.0066*	0.0060	0.0046	0.0261*	0.0117
	(0.005)	(0.003)	(0.005)	(0.004)	(0.014)	(0.010)
Cumul. OECD Market Size$_{t\ to\ t+12}$	Yes	No	Yes	No	Yes	No
Cumul. NIH funding (Share)$_{t-1\ to\ -12}$	Yes	No	Yes	No	Yes	No
Cumul. New MeSH ids$_t$	Yes	No	Yes	No	Yes	No
US Market Size$_t$	No	Yes	No	Yes	No	Yes
NIH funding$_{t-1}$	No	Yes	No	Yes	No	Yes
New MeSH terms$_t$	No	Yes	No	Yes	No	Yes
ICD9 group FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3780	3800	3860	3880	3200	3220
ICD9-codes	189	190	193	194	160	161
ICD9-groups	109	110	113	114	86	87
Log-likelihood	−6681	−6717	−10378	−10447	−1308	−1315

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. Column (1), (3), and (5) include alternative control variables such as OECD market size and NIH funding calculated as a share of all publications in a disease category that acknowledge a specific Institute/Center. Column (2), (4), and (6) include control variables that do not accumulate future/past periods but only consider year t. In both cases, control variables are log-transformed and the MMS is calculated at the ICD-9 group level. The unit of observation of the dependent variable is the ICD-9 three-digit code by year. It is the annual number of newly discovered NME in Columns (1) and (2), NME in all development stages (preclinicals, clinical trials, registrations, approvals) in Columns (3) and (4), and approved NME in Column (5) and (6). Standard errors are clustered at the level of treatment (ICD-9 group level) and shown in parentheses. Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table C-6: Drug development – pre-MMS period until 2005

All ICD9 Groups	(1)	(2)	(3)	(4)
Count/PPML		Drug Development		
	Early	All Stages	Phase 1-3	Approval
MMS × 2006-08	0.0015	0.0009	−0.0003	−0.0025
	(0.002)	(0.002)	(0.003)	(0.006)
MMS × 2009-11	0.0022	0.0014	0.0012	0.0028
	(0.002)	(0.002)	(0.003)	(0.008)
MMS × 2012-14	0.0056**	0.0043*	0.0034	0.0073
	(0.003)	(0.003)	(0.004)	(0.009)
MMS × 2015-16	0.0090**	0.0070**	0.0063	0.0054
	(0.004)	(0.003)	(0.005)	(0.010)
Cumul. US Market Size$_{t \text{ to } t+12}$	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t-1 \text{ to } -12}$	Yes	Yes	Yes	Yes
Cumul. New MeSH ids$_t$	Yes	Yes	Yes	Yes
ICD9 group FE	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes
Observations	3800	3880	3760	3220
ICD9-codes	190	194	188	161
ICD9-groups	110	114	110	87
Log-likelihood	−6723	−10449	−6139	−1320

Notes: Columns (1) to (4) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The pre-MMA period includes the years 1997-2005 (until the implementation of the MMA). The dependent variable is the annual number of newly discovered NME in Column (1), NME in all development stages in Column (2), NME in phase I-III clinical trials in Column (3), and the annual number of approved drugs in Column (4). The unit of observation is the ICD-9 three-digit level by year. The control variables are log-transformed. Standard errors are clustered at the level of treatment (ICD-9 group level) and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

C.3.3 Tables – Multivariate Analysis Biomedical Science

Table C-7: Scientific publications – parallel trends

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
			Scientific Publications			
	All	All	Fractions	University	Corporate	Patent-Weighted
MMS × 2000-02	0.0002		−0.0002	0.0002	−0.0007	−0.0008
	(0.001)		(0.001)	(0.001)	(0.001)	(0.002)
MMS > Median × 2000-02		0.0239				
		(0.026)				
ICD9 group FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9366	9366	9366	9366	9030	9030
MeSH IDs	1561	1561	1561	1561	1505	1505
ICD-group	127	127	127	127	108	107
Log-likelihood	−180453	−180451	−148840	−142788	−12798	−15509

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The time period of these regressions is the pre-MMA period 1997-2002. We employ a placebo test (2000-2002 x MMS with 1997-1999 as the baseline period) in order to test the parallel trends assumption. The dependent variable is the annual number of scientific publications from all affiliations (1)-(2), disease weighted (fractional) number of scientific publications from all affiliations (3), number of scientific publications from at least one author affiliated with an university (4), affiliated with a corporation (5), and patent-weighted number of scientific publications (6). A patent-weight is calculated based on the patent family's first application being filed within 5 years from the scientific publication. The unit of observation is the MeSH term by year. Standard errors are clustered at the level of treatment (ICD-9 group level) and shown in parentheses. Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table C-8: Scientific publications – fractional counts/winsorized

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
	Scientific Publications		Scientific Publications - fractions			
	winsorized	winsorized	count	count	winsorized	winsorized
MMS × post 2003	0.0004		0.0001		0.0004	
	(0.001)		(0.001)		(0.001)	
MMS × 2004-05		0.0006		0.0005		0.0006
		(0.001)		(0.001)		(0.001)
MMS × 2006-08		0.0004		−0.0001		0.0004
		(0.001)		(0.001)		(0.001)
MMS × 2009-11		0.0010		0.0007		0.0009
		(0.001)		(0.002)		(0.001)
MMS × 2012-14		0.0015		0.0009		0.0013
		(0.001)		(0.003)		(0.002)
MMS × 2015-16		0.0023		0.0018		0.0018
		(0.002)		(0.003)		(0.002)
Cumul. US Market Size$_{t \text{ to } t+12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t-1 \text{ to } -12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. New MeSH ids$_t$	Yes	Yes	Yes	Yes	Yes	Yes
ICD9 group FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	31260	31260	31260	31260	31260	31260
MeSH IDs	1563	1563	1563	1563	1563	1563
ICD-group	129	129	129	129	129	129
Log-likelihood	−224810	−224798	−739655	−739630	−174355	−174351

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. The dependent variable is the annual number of scientific publications winsorized at the 99 percentile in Columns (1) and (2), the number of scientific publications weighted by the inverse number of linked diseases (fractional counts) in Columns (3) and (4), and the winsorized number of scientific publications weighted by the inverse number of linked diseases in Columns (5) and (6). The control variables are log-transformed. Standard errors are clustered at the level of treatment (ICD-9 group level) and shown in parentheses. Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table C-9: Scientific publications – alternative controls/different models

All ICD9 Groups	(1)	(2)	(3)	(4)
DV: Scientific Publications		Count/PPML		Count/Linear
	MeSH FE	Different Controls		OLS
MMS × 2004-05	0.0006	0.0014	0.0005	0.0908**
	(0.001)	(0.001)	(0.001)	(0.043)
MMS × 2006-08	−0.0001	0.0010	−0.0004	0.1174*
	(0.001)	(0.001)	(0.001)	(0.068)
MMS × 2009-11	0.0013	0.0030	0.0009	0.2071**
	(0.002)	(0.002)	(0.001)	(0.099)
MMS × 2012-14	0.0017	0.0039	0.0012	0.2695**
	(0.003)	(0.003)	(0.002)	(0.125)
MMS × 2015-16	0.0030	0.0058*	0.0024	0.3414**
	(0.003)	(0.003)	(0.003)	(0.149)
Cumul. US Market Size$_{t \text{ to } t+12}$	Yes	No	No	Yes
Cumul. NIH funding$_{t-1 \text{ to } -12}$	Yes	No	No	Yes
Cumul. New MeSH terms$_t$	Yes	Yes	No	Yes
Cumul. OECD Market Size$_{t \text{ to } t+12}$	No	Yes	No	No
Cumul. NIH funding (Share)$_{t-1 \text{ to } -12}$	No	Yes	No	No
US Market Size$_t$	No	No	Yes	No
NIH funding$_{t-1}$	No	No	Yes	No
New MeSH terms$_t$	No	No	Yes	No
ICD9 group FE	No	Yes	Yes	Yes
MeSH FE	Yes	No	No	No
Calendar Year FE	Yes	Yes	Yes	Yes
Observations	30680	31180	31260	31260
MeSH terms	1534	1559	1563	1563
ICD-group	129	126	129	129
Log-likelihood	−89923	−894188	−894359	−183443

Notes: Columns (1) to (3) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. The dependent variable is the annual number of scientific publications. In Column (1), we use MeSH term fixed effects instead of ICD-9 group level fixed effects and cluster standard errors at the MeSH term level instead of ICD-9 group level. Column (3) includes alternative control variables such as OECD market size and NIH funding calculated as a share of all publications in a disease category that acknowledge a specific Institute/Center. Column (4) includes control variables that do not accumulate future/past periods but only consider year t. In Column (5), we show the estimates of linear regressions with high-dimensional fixed effects. The control variables are always log-transformed. Standard errors are shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table C-10: Scientific publications by affiliation type – US only

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)
	Scientific Publications - US only				
	NIH	University	No University	AMC	Corporate
MMS × 2004-05	−0.0011	0.0003	0.0006	0.0003	−0.0002
	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)
MMS × 2006-08	−0.0042	−0.0006	−0.0004	−0.0002	0.0013
	(0.003)	(0.001)	(0.001)	(0.002)	(0.002)
MMS × 2009-11	−0.0026	0.0008	0.0009	0.0011	0.0037
	(0.004)	(0.002)	(0.002)	(0.003)	(0.003)
MMS × 2012-14	−0.0032	0.0007	0.0014	0.0026	0.0064**
	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)
MMS × 2015-16	−0.0027	0.0014	0.0022	0.0030	0.0113**
	(0.006)	(0.003)	(0.003)	(0.004)	(0.005)
Cumul. US Market Size$_{t\ to\ t+12}$	Yes	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t-1\ to\ -12}$	Yes	Yes	Yes	Yes	Yes
Cumul. New MeSH ids$_t$	Yes	Yes	Yes	Yes	Yes
ICD9 group FE	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes
Observations	30400	31260	31220	31140	30840
MeSH IDs	1520	1563	1561	1557	1542
ICD-group	108	129	128	125	121
Log-likelihood	−38133	−532245	−403089	−95553	−46363

Notes: Columns (1) to (5) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. The dependent variable is the annual number of scientific publications split by at least one author's affiliation. All authors have US affiliations. In Column (1) at least one author is affiliated with the NIH, in Column (2) with a university, in Column (4) with an academic medical center (AMC), and in Column (5) with a firm. Column (3) includes publications that have at least one author not affiliated with a US university. The control variables are log-transformed. Standard errors are clustered at the level of treatment (ICD-9 group level) and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table C-11: Scientific publications by affiliation type – all authors from same type

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)
			Scientific Publications		
	NIH only	Uni only	No Uni only	AMC only	Corp only
MMS × 2004-05	−0.0007	0.0003	0.0011	0.0027	−0.0022
	(0.003)	(0.001)	(0.001)	(0.002)	(0.005)
MMS × 2006-08	−0.0056	−0.0003	0.0003	0.0028	0.0028
	(0.004)	(0.001)	(0.001)	(0.003)	(0.004)
MMS × 2009-11	−0.0029	0.0007	0.0012	0.0036	0.0009
	(0.004)	(0.002)	(0.002)	(0.004)	(0.006)
MMS × 2012-14	−0.0035	0.0008	0.0018	0.0042	0.0058
	(0.006)	(0.002)	(0.003)	(0.005)	(0.005)
MMS × 2015-16	−0.0047	0.0022	0.0030	0.0063	0.0081
	(0.008)	(0.003)	(0.003)	(0.006)	(0.007)
Cumul. US Market Size$_{t\ to\ t+12}$	Yes	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t-1\ to\ -12}$	Yes	Yes	Yes	Yes	Yes
Cumul. New MeSH terms$_t$	Yes	Yes	Yes	Yes	Yes
ICD9 group FE	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes
Observations	28080	31260	31200	30300	30100
MeSH terms	1404	1563	1560	1515	1505
ICD-group	91	129	127	120	111
Log-likelihood	−13385	−337848	−177541	−20659	−10336

Notes: Columns (1) to (5) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. The dependent variable is the annual number of scientific publications split by all author's affiliations. In Column (1) all authors are affiliated with the NIH, in Column (2) with a university, in Column (4) with an academic medical center, and in Column (5) with a firm. Column (3) includes publications that have all authors not affiliated with a university. The control variables are log-transformed. Standard errors are clustered at the level of treatment (ICD-9 group level) and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table C-12: Scientific publications by affiliation type – fractional count/winsorized

All ICD9 Groups Count/PPML	(1) NIH	(2) University	(3) AMC	(4) Corporate	(5) University	(6) Corporate
	\multicolumn{4}{Scientific Publications - fractional}			winsorized	
MMS × 2004-05	−0.0008	0.0003	−0.0003	−0.0003	0.0003	−0.0005
	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)
MMS × 2006-08	−0.0016	−0.0004	−0.0003	0.0012	−0.0001	−0.0002
	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)
MMS × 2009-11	−0.0017	0.0005	0.0003	0.0024	0.0010	0.0026
	(0.003)	(0.002)	(0.002)	(0.003)	(0.001)	(0.003)
MMS × 2012-14	−0.0023	0.0005	0.0009	0.0048	0.0015	0.0037
	(0.004)	(0.003)	(0.003)	(0.004)	(0.002)	(0.003)
MMS × 2015-16	−0.0013	0.0013	0.0016	0.0081*	0.0029	0.0068*
	(0.005)	(0.003)	(0.004)	(0.005)	(0.002)	(0.004)
Cumul. US Market Size$_{t\ to\ t+12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t-1\ to\ -12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. New MeSH terms$_t$	Yes	Yes	Yes	Yes	Yes	Yes
ICD9 group FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	30480	31260	31240	31120	31260	31120
MeSH terms	1524	1563	1562	1556	1563	1556
ICD-group	111	129	128	125	129	125
Log-likelihood	−45219	−614764	−126599	−56819	−540867	−52867

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. In Columns (1) to (4), the dependent variable is the annual number of scientific publications weighted by the inverse number of linked diseases (fractional counts) split by at least one author's affiliation. In Columns (5) to (6), the dependent variable is the annual number of scientific publications winsorized at the 99 percentile. The control variables are log-transformed. Standard errors are clustered at the level of treatment (ICD-9 group level) and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table C-13: University publications – alternative controls/different models

All ICD9 Groups	(1)	(2)	(3)	(4)
DV: University Publications		Count/PPML		Count/Linear
	MeSH FE	Different Controls		OLS
MMS × 2004-05	0.0004	0.0012	0.0004	0.0808**
	(0.001)	(0.001)	(0.001)	(0.040)
MMS × 2006-08	−0.0005	0.0009	−0.0006	0.1059*
	(0.002)	(0.002)	(0.001)	(0.062)
MMS × 2009-11	0.0010	0.0029	0.0007	0.1885**
	(0.002)	(0.002)	(0.002)	(0.092)
MMS × 2012-14	0.0012	0.0038	0.0009	0.2456**
	(0.003)	(0.003)	(0.002)	(0.117)
MMS × 2015-16	0.0025	0.0057*	0.0021	0.3103**
	(0.003)	(0.003)	(0.003)	(0.138)
Cumul. US Market Size$_{t \text{ to } t+12}$	Yes	No	No	Yes
Cumul. NIH funding$_{t-1 \text{ to } -12}$	Yes	No	No	Yes
Cumul. New MeSH terms$_t$	Yes	Yes	No	Yes
Cumul. OECD Market Size$_{t \text{ to } t+12}$	No	Yes	No	No
Cumul. NIH funding (Share)$_{t-1 \text{ to } -12}$	No	Yes	No	No
US Market Size$_t$	No	No	Yes	No
NIH funding$_{t-1}$	No	No	Yes	No
New MeSH terms$_t$	No	No	Yes	No
ICD9 group FE	No	Yes	Yes	Yes
MeSH FE	Yes	No	No	No
Calendar Year FE	Yes	Yes	Yes	Yes
Observations	30560	31180	31260	31260
MeSH terms	1528	1559	1563	1563
ICD-group	129	126	129	129
Log-likelihood	−79980	−742452	−742594	−177688

Notes: Columns (1) to (3) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. The dependent variable is the annual number of scientific publications with at least one author affiliated with an university. In Column (1), we use MeSH term fixed effects instead of ICD-9 group level fixed effects and cluster standard errors at the MeSH term level instead of ICD-9 group level. Column (2) includes alternative control variables such as OECD market size and NIH funding calculated as a share of all publications in a disease category that acknowledge a specific Institute/Center. Column (3) includes control variables that do not accumulate future/past periods but only consider year t. In Column (4), we show the estimates of linear regressions with high-dimensional fixed effects. The control variables are always log-transformed. Standard errors are shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table C-14: Corporate publications – alternative controls/different models

All ICD9 Groups	(1)	(2)	(3)	(4)
DV: Corporate Publications		Count/PPML		Count/Linear
	MeSH FE	Different Controls		OLS
MMS × 2004-05	−0.0001	0.0012	−0.0012	0.0047
	(0.002)	(0.002)	(0.002)	(0.003)
MMS × 2006-08	0.0010	0.0031	−0.0005	0.0088
	(0.002)	(0.002)	(0.002)	(0.006)
MMS × 2009-11	0.0036	0.0065**	0.0014	0.0151*
	(0.003)	(0.003)	(0.003)	(0.008)
MMS × 2012-14	0.0059	0.0096***	0.0031	0.0206**
	(0.004)	(0.003)	(0.003)	(0.010)
MMS × 2015-16	0.0098**	0.0142***	0.0064	0.0295**
	(0.004)	(0.004)	(0.004)	(0.012)
Cumul. US Market Size$_{t \text{ to } t+12}$	Yes	No	No	Yes
Cumul. NIH funding$_{t-1 \text{ to } -12}$	Yes	No	No	Yes
Cumul. New MeSH terms$_t$	Yes	Yes	No	Yes
Cumul. OECD Market Size$_{t \text{ to } t+12}$	No	Yes	No	No
Cumul. NIH funding (Share)$_{t-1 \text{ to } -12}$	No	Yes	No	No
US Market Size$_t$	No	No	Yes	No
NIH funding$_{t-1}$	No	No	Yes	No
New MeSH terms$_t$	No	No	Yes	No
ICD9 group FE	No	Yes	Yes	Yes
MeSH FE	Yes	No	No	No
Calendar Year FE	Yes	Yes	Yes	Yes
Observations	19880	31100	31120	31260
MeSH terms	994	1555	1556	1563
ICD-group	125	124	125	129
Log-likelihood	−18220	−68061	−68103	−96677

Notes: Columns (1) to (3) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. The dependent variable is the annual number of scientific publications with at least one author affiliated with a corporation. In Column (1), we use MeSH term fixed effects instead of ICD-9 group level fixed effects and cluster standard errors at the MeSH term level instead of ICD-9 group level. Column (2) includes alternative control variables such as OECD market size and NIH funding calculated as a share of all publications in a disease category that acknowledge a specific Institute/Center. Column (3) includes control variables that do not accumulate future/past periods but only consider year t. In Column (4), we show the estimates of linear regressions with high-dimensional fixed effects. The control variables are always log-transformed. Standard errors are shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table C-15: University publications by journal type

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
			Scientific Publications			
	Academic	Clinical	Clin. Pract	Ind./Clin.	Industrial	Ind. Pract
MMS × 2004-05	0.0012	−0.0011	0.0017	0.0020	0.0006	0.0050
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)
MMS × 2006-08	−0.0052	−0.0023	0.0011	0.0002	0.0011	0.0015
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
MMS × 2009-11	−0.0038	−0.0001	0.0017	0.0008	0.0026	−0.0012
	(0.004)	(0.002)	(0.002)	(0.003)	(0.002)	(0.004)
MMS × 2012-14	−0.0027	−0.0004	0.0027	0.0011	−0.0012	0.0020
	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)	(0.007)
MMS × 2015-16	−0.0037	−0.0002	0.0049	0.0018	−0.0025	0.0030
	(0.006)	(0.004)	(0.003)	(0.004)	(0.004)	(0.006)
Cumul. US Market Size$_{t\ to\ t+12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t-1\ to\ -12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. New MeSH terms$_t$	Yes	Yes	Yes	Yes	Yes	Yes
ICD9 group FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	30440	31180	31240	31120	29840	26800
MeSH terms	1522	1559	1562	1556	1492	1340
ICD-group	111	127	128	124	107	86
Log-likelihood	−16852	−286422	−124736	−188377	−22302	−9915

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. The dependent variable is the number of scientific publications with at least one author affiliated with an university in academic journals in Column (1), in clinically relevant journals in Column (2), in clinical practice journals in Column (3), in industry-clinical journals in Column (4), in industrial journals in Column (5), and in industry practice journals in Column (6). Journal classification is based on the proportion of published research coming from general hospitals and industry using the publicly available data set provided by Tijssen (2010). The control variables are log-transformed. Standard errors are clustered at the level of treatment (ICD-9 group level) and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table C-16: Corporate publications by journal type

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
			Scientific	Publications		
	Academic	Clinical	Clin. Pract	Ind./Clin.	Industrial	Ind. Pract
MMS × 2004-05	0.0015	−0.0037	0.0029	0.0017	−0.0003	−0.0046
	(0.014)	(0.003)	(0.004)	(0.002)	(0.006)	(0.006)
MMS × 2006-08	−0.0048	−0.0011	0.0069	0.0003	−0.0054	−0.0109*
	(0.014)	(0.004)	(0.005)	(0.003)	(0.004)	(0.006)
MMS × 2009-11	−0.0129	0.0031	0.0038	0.0021	0.0056	−0.0127**
	(0.015)	(0.005)	(0.006)	(0.005)	(0.007)	(0.006)
MMS × 2012-14	−0.0113	0.0024	0.0085	0.0062	−0.0080	−0.0046
	(0.017)	(0.007)	(0.007)	(0.006)	(0.006)	(0.007)
MMS × 2015-16	−0.0205	0.0060	0.0133	0.0101	−0.0131	0.0035
	(0.024)	(0.008)	(0.009)	(0.008)	(0.009)	(0.006)
Cumul. US Market Size$_{t \text{ to } t+12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t-1 \text{ to } -12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. New MeSH terms$_t$	Yes	Yes	Yes	Yes	Yes	Yes
ICD9 group FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21800	30300	29580	30040	26320	24020
MeSH terms	1090	1515	1479	1502	1316	1201
ICD-group	53	109	111	111	76	66
Log-likelihood	−990	−22655	−10788	−26692	−2656	−3119

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. The dependent variable is the number of scientific publications with at least one author affiliated with a corporation in academic journals in Column (1), in clinically relevant journals in Column (2), in clinical practice journals in Column (3), in industry-clinical journals in Column (4), in industrial journals in Column (5), and in industry practice journals in Column (6). Journal classification is based on the proportion of published research coming from general hospitals and industry using the publicly available data set provided by Tijssen (2010). The control variables are log-transformed. Standard errors are clustered at the level of treatment (ICD-9 group level) and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table C-17: Scientific publications – alternative impact/patent weights

All ICD9 Groups Count/PPML	(1) Citation-Weighted	(2) Citation-Weighted	(3) Drug Patent (Cortellis)	(4) Drug Patent (Cortellis)	(5) Drug Patent same ICD	(6) Drug Patent same ICD
	Uni	Corporate	Uni	Corporate	Uni	Corporate
MMS × 2004-05	−0.0006	0.0010	−0.0008	0.0068	−0.0008	0.0104*
	(0.001)	(0.004)	(0.003)	(0.006)	(0.004)	(0.006)
MMS × 2006-08	−0.0018	−0.0010	−0.0031	−0.0060	−0.0025	−0.0020
	(0.002)	(0.004)	(0.003)	(0.005)	(0.003)	(0.005)
MMS × 2009-11	−0.0007	0.0029	−0.0042	0.0040	−0.0051	0.0044
	(0.002)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)
MMS × 2012-14	0.0000	0.0012	−0.0083	0.0021	−0.0092	0.0022
	(0.003)	(0.006)	(0.005)	(0.006)	(0.006)	(0.006)
MMS × 2015-16	0.0033	0.0040	−0.0058	−0.0017	−0.0054	0.0006
	(0.004)	(0.008)	(0.007)	(0.007)	(0.008)	(0.008)
Cumul. US Market Size$_{t \text{ to } t+12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t-1 \text{ to } -12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. New MeSH terms$_t$	Yes	Yes	Yes	Yes	Yes	Yes
ICD9 group FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	31260	31120	29580	26240	29100	26020
MeSH terms	1563	1556	1479	1312	1455	1301
ICD-group	129	125	104	81	100	79
Log-likelihood	−14628648	−1677386	−21419	−5591	−17709	−4857

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. In Columns (1)-(2), the dependent variable is the forward citation-weighted number of university/corporate scientific publications. In Columns (3)-(4), the dependent variable is the number of university/corporate scientific publications that are associated with at least one patent family (patent-weighted) that appears in the Cortellis drug database. In Columns (5)-(6), the patent family has to appear in the Cortellis drug database in the same ICD-9 group as the scientific publications. A patent-weight is calculated based on the patent family's first application being filed within 5 years from the scientific publication. The match between patent families with SNPLs and Cortellis was conducted at the INPADOC family level. The control variables are log-transformed. Standard errors are clustered at the level of treatment (ICD-9 group level) and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table C-18: Academic Medical Center publications – type of research/impact

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
		MeSH-Weighted		JIF-Weighted	Patent-Weighted	
Academic Medical Center	Basic	CT	Pharma	JIF	Patent	Family Size
MMS × 2004-05	0.0005	−0.0015	−0.0029	−0.0012	−0.0011	−0.0053
	(0.001)	(0.002)	(0.004)	(0.002)	(0.004)	(0.004)
MMS × 2006-08	−0.0001	−0.0003	−0.0037	−0.0015	0.0028	0.0042
	(0.002)	(0.003)	(0.005)	(0.003)	(0.004)	(0.005)
MMS × 2009-11	0.0014	0.0005	−0.0016	−0.0005	0.0023	0.0030
	(0.002)	(0.004)	(0.006)	(0.004)	(0.005)	(0.007)
MMS × 2012-14	0.0019	0.0029	0.0025	0.0010	0.0049	0.0061
	(0.003)	(0.005)	(0.007)	(0.004)	(0.005)	(0.007)
MMS × 2015-16	0.0029	0.0038	0.0015	0.0021	0.0019	−0.0003
	(0.003)	(0.006)	(0.007)	(0.005)	(0.008)	(0.008)
Cumul. US Market Size$_{t\ to\ t+12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t-1\ to\ -12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. New MeSH terms$_t$	Yes	Yes	Yes	Yes	Yes	Yes
ICD9 group FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	31240	30740	29600	31240	29480	29480
MeSH terms	1562	1537	1480	1562	1474	1474
ICD-group	128	117	105	128	104	104
Log-likelihood	−134852	−29085	−7536	−695655	−11485	−74365

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. The dependent variable is the number of scientific publications from at least one author affiliated with an academic medical center that are associated with MeSH terms related to clinical trials in Column (2) and pharmaceutical preparations in Column (3). In Column (1), we include the residual academic medical center publications (non CT/non pharmaceutical preparations). In Columns (4), the dependent variable is the journal impact factor-weighted number of scientific publications from at least one author affiliated with an academic medical center. In Column (5), the dependent variable is the number of academic medical center publications that are associated with at least one patent application (patent-weighted). In Column (6), we weight the number of academic medical center publications by the size of the average patent family associated with the publication. A patent/family size-weight is calculated based on the patent family's first application being filed within 5 years from the scientific publication. The control variables are log-transformed. Standard errors are clustered at the level of treatment (ICD-9 group level) and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table C-19: Scientific publications – pre-MMS period until 2005

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
		Scientific Publications			Patent-Weighted	
	All	Fractions	University	Corporate	University	Corporate
MMS × 2006-08	−0.0004	−0.0003	−0.0006	0.0011	0.0001	−0.0009
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
MMS × 2009-11	0.0010	0.0005	0.0008	0.0037	0.0002	−0.0003
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
MMS × 2012-14	0.0014	0.0006	0.0010	0.0060*	−0.0002	0.0025
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
MMS × 2015-16	0.0027	0.0015	0.0023	0.0098**	−0.0009	0.0026
	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.005)
Cumul. US Market Size$_{t\ to\ t+12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t-1\ to\ -12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. New MeSH terms$_t$	Yes	Yes	Yes	Yes	Yes	Yes
ICD9 group FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	31260	31260	31260	31120	30700	27980
MeSH terms	1563	1563	1563	1556	1535	1399
ICD-group	129	129	129	125	114	92
Log-likelihood	−894324	−739633	−742563	−68090	−50084	−10339

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. The pre-period lasts until 2005, which is the year before the implementation of the MMA. The dependent variable is the annual number of scientific publications from all affiliation types in Column (1), weighted by the inverse number of linked diseases in Column (2), with at least one author being affiliated with an university in Column (3), and with a corporation in Column (4). In Columns (5) and (6), the dependent variable is the number of university/corporate scientific publications that are associated with at least one patent application (patent-weighted). A patent-weight is calculated based on the patent family's first application being filed within 5 years from the scientific publication. The control variables are log-transformed. Standard errors are clustered at the level of treatment (ICD-9 group level) and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table C-20: Scientific publications – all PMID (incl. confounded)

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
	Scientific Publications				Patent-Weighted	
	All	Fractions	University	Corporate	University	Corporate
MMS × 2004-05	0.0014**	0.0009	0.0013**	0.0011	0.0034***	0.0044**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
MMS × 2006-08	0.0007	0.0003	0.0005	0.0006	0.0024**	0.0028
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
MMS × 2009-11	0.0014	0.0010	0.0014	0.0018	0.0019	0.0022
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)
MMS × 2012-14	0.0016	0.0011	0.0014	0.0035	0.0016	0.0063*
	(0.001)	(0.002)	(0.002)	(0.003)	(0.002)	(0.004)
MMS × 2015-16	0.0026	0.0019	0.0025	0.0065*	0.0033	0.0098**
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.004)
Cumul. US Market Size$_{t \text{ to } t+12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t-1 \text{ to } -12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. New MeSH terms$_t$	Yes	Yes	Yes	Yes	Yes	Yes
ICD9 group FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	31460	31460	31460	31440	31280	30160
MeSH terms	1573	1573	1573	1572	1564	1508
ICD-group	129	129	129	128	122	107
Log-likelihood	−1765706	−1037220	−1447553	−120655	−93633	−18148

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. The dependent variable is the annual number of scientific publications including PMID, which are associated with a disease MeSH term with unknown MMS (confounded). It includes publications from all affiliation types in Column (1), weighted by the inverse number of linked diseases in Column (2), with at least one author being affiliated with an university in Column (3), and with a corporation in Column (4). In Columns (5) and (6), the dependent variable is the number of university/corporate scientific publications (including confounded PMID) that are associated with at least one patent application (patent-weighted). A patent-weight is calculated based on the patent family's first application being filed within 5 years from the scientific publication. The control variables are log-transformed. Standard errors are clustered at the level of treatment (ICD-9 group level) and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table C-21: Scientific publications – ICD-9 group level

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
		Scientific	Publications		Patent-Weighted	
	All	Fractions	University	Corporate	University	Corporate
MMS × 2004-05	0.0005	0.0004	0.0003	−0.0002	0.0013	0.0056*
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)
MMS × 2006-08	−0.0003	−0.0002	−0.0006	0.0010	0.0002	0.0018
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)
MMS × 2009-11	0.0007	0.0008	0.0005	0.0030	0.0003	0.0013
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
MMS × 2012-14	0.0010	0.0011	0.0006	0.0052	−0.0003	0.0038
	(0.002)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)
MMS × 2015-16	0.0021	0.0022	0.0016	0.0082*	−0.0011	0.0031
	(0.003)	(0.003)	(0.003)	(0.005)	(0.004)	(0.005)
Cumul. US Market Size$_{t \text{ to } t+12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t-1 \text{ to } -12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. New MeSH terms$_t$	Yes	Yes	Yes	Yes	Yes	Yes
ICD9 group FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2580	2580	2580	2500	2280	1840
ICD-group	129	129	129	125	114	92
Log-likelihood	−13270	−12351	−12072	−5187	−3917	−1931

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the ICD-9 group level (MeSH terms aggregated to the ICD-9 group in the B&P2011 ICD-9-MeSH crosswalk presented in Table C-1). The dependent variable is the annual number of scientific publications from all affiliation types in Column (1), weighted by the inverse number of linked diseases in Column (2), with at least one author being affiliated with an university in Column (3), and with a corporation in Column (4). In Columns (5) and (6), the dependent variable is the number of university/corporate scientific publications that are associated with at least one patent application (patent-weighted). A patent-weight is calculated based on the patent family's first application being filed within 5 years from the scientific publication. The control variables are log-transformed. Standard errors are clustered at the level of treatment (ICD-9 group level) and shown in parentheses. Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

List of Figures

LIST OF FIGURES

List of Tables

Bibliography

Abud, M. J., B. Hall, and C. Helmers (2015). An Empirical Analysis of Primary and Secondary Pharmaceutical Patents in Chile. *PLoS One 10*(4), e0124257.

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

Adams, C. P. and V. V. Brantner (2006). Estimating the Cost of New Drug Development: Is it Really $802 million? *Health Affairs 25*(2), 420–428.

Agarwal, R. and P. Gaule (2021). What Drives Innovation? Lessons from COVID-19 R&D. IZA Discussion Paper No. 14079.

Aghion, P., A. Bergeaud, M. Lequien, and M. J. Melitz (2018). The Heterogeneous Impact of Market Size on Innovation: Evidence from French Firm-Level Exports. NBER Working Paper No. 24600.

Aghion, P., M. Dewatripont, and J. C. Stein (2008). Academic Freedom, Private-Sector Focus, and the Process of Innovation. *The RAND Journal of Economics 39*(3), 617–635.

Ahmadpoor, M. and B. F. Jones (2017). The Dual Frontier: Patented Inventions and Prior Scientific Advance. *Science 357*(6351), 583–587.

Amin, T. and A. S. Kesselheim (2012). Secondary Patenting of Branded Pharmaceuticals: A Case Study of How Patents on Two HIV Drugs Could Be Extended for Decades. *Health Affairs 31*(10), 2286–2294.

Arcidiacono, P., P. B. Ellickson, P. Landry, and D. B. Ridley (2013). Pharmaceutical Followers. *International Journal of Industrial Organization 31*(5), 538–553.

Arora, A., S. Belenzon, and A. Patacconi (2018). The Decline of Science in Corporate R&D. *Strategic Management Journal 39*(1), 3–32.

Arrow, K. (1962). Economic Welfare and the Allocation of Resources for Invention. In *The Rate and Direction of Inventive Activity: Economic and Social Factors*, pp. 609–626. Princeton University Press.

Azoulay, P., J. S. Graff Zivin, D. Li, and B. N. Sampat (2018). Public R&D Investments and Private-Sector Patenting: Evidence from NIH Funding Rules. *The Review of Economic Studies 86*(1), 117–152.

Azoulay, P., J. S. Graff Zivin, D. Li, and B. N. Sampat (2019). Public R&D Investments and Private-Sector Patenting: Evidence from NIH Funding Rules. *The Review of Economic Studies 86*(1), 117–152.

Azoulay, P., J. S. Graff Zivin, and J. Wang (2010). Superstar Extinction. *The Quarterly Journal of Economics 125*(2), 549–589.

Baruffaldi, S. H. and F. Gaessler (2021). The Returns to Physical Capital in Knowledge Production: Evidence from Lab Disasters. Max Planck Institute for Innovation & Competition Research Paper No. 21-19.

Batista, P. H. D., D. Byrski, M. Lamping, and R. Romandini (2019). IP-based Incentives against Antimicrobial Crisis: a European Perspective. *IIC-International Review of Intellectual Property and Competition Law 50*(1), 30–76.

Bhattacharya, J. and M. Packalen (2011). Opportunities and Benefits as Determinants of the Direction of Scientific Research. *Journal of Health Economics 30*(4), 603–615.

Blake, J. A., J. E. Richardson, C. J. Bult, J. A. Kadin, and J. T. Eppig (2003). MGD: The Mouse Genome Database. *Nucleic Acids Research 31*(1), 193–195.

Bloom, D. E., D. Canning, R. Kotschy, K. Prettner, and J. J. Schünemann (2019). Health and Economic Growth: Reconciling the Micro and Macro Evidence. NBER Working Paper No. 26003.

Blume-Kohout, M. E. (2012). Does Targeted, Disease-Specific Public Research Funding Influence Pharmaceutical Innovation? *Journal of Policy Analysis and Management 31*(3), 641–660.

Blume-Kohout, M. E. and N. Sood (2013). Market Size and Innovation: Effects of Medicare Part D on Pharmaceutical Research and Development. *Journal of Public Economics 97*, 327–336.

Bokhari, F. A. and G. M. Fournier (2013). Entry in the ADHD Drugs Market: Welfare Impact of Generics and Me-too's. *The Journal of Industrial Economics 61*(2), 339–392.

Branstetter, L., C. Chatterjee, and M. J. Higgins (2016). Regulation and Welfare: Evidence from Paragraph IV Generic Entry in the Pharmaceutical Industry. *The RAND Journal of Economics 47*(4), 857–890.

Breckenridge, A. and R. Jacob (2019). Overcoming the Legal and Regulatory Barriers to Drug Repurposing. *Nature Reviews Drug Discovery 18*(1), 1–2.

Bryant, C. D. (2011). The Blessings and Curses of C57BL/6 Substrains in Mouse Genetic Studies. *Annals of the New York Academy of Sciences 1245*, 31–33.

Budish, E., B. N. Roin, and H. Williams (2015). Do Firms Underinvest in Long-Term Research? Evidence from Cancer Clinical Trials. *The American Economic Review 105*(7), 2044–2085.

Bush, V. (1945). *Science, The Endless Frontier: A Report to the President*. Washington: United States Government Printing Office.

Byrski, D., F. Gaessler, and M. J. Higgins (2021). Market Size and Research: Evidence from the Pharmaceutical Industry. NBER Working Paper No. 28858.

Clemens, J. P. and M. Olsen (2021). Medicare and the Rise of American Medical Patenting: The Economics of User-Driven Innovation. CESifo Working Paper No. 9008.

Cockburn, I. M. and R. M. Henderson (2000). Publicly Funded Science and the Productivity of the Pharmaceutical Industry. *Innovation Policy and the Economy 1*, 1–34.

Cohen, W. M., R. R. Nelson, and J. P. Walsh (2000). Protecting their Intellectual Assets: Appropriability Conditions and Why US Manufacturing Firms Patent (or Not). NBER Working Paper No. 7552.

Cohen, W. M., H. Sauermann, and P. Stephan (2020). Not in the Job Description: The Commercial Activities of Academic Scientists and Engineers. *Management Science 66*(9), 4108–4117.

Correia, S. (2016). A Feasible Estimator for Linear Models with Multi-Way Fixed Effects. Working Paper.

Correia, S. (2018). ivreghdfe: Stata Module for Extended Instrumental Variable Regressions with Multiple Levels of Fixed Effects. mimeo.

Correia, S., P. Guimarães, and T. Zylkin (2020). Fast Poisson Estimation with High-Dimensional Fixed Effects. *The Stata Journal 20*(1), 95–115.

Cuénot, L. (1902). La loi de Mendel et l'hérédité de la pigmentation chez les souris. *Archives de Zoologie Expérimentale et Générale 3 Series*.

Cunningham, C., F. Ederer, and S. Ma (2021). Killer Acquisitions. *Journal of Political Economy 129*(3), 649–702.

Cutler, D. M., G. Long, E. R. Berndt, J. Royer, A.-A. Fournier, A. Sasser, and P. Cremieux (2007). The Value of Antihypertensive Drugs: A Perspective on Medical Innovation. *Health Affairs 26*(1), 97–110.

David, P. A. (1985). Clio and the Economics of QWERTY. *The American Economic Review 75*(2), 332–337.

de Rassenfosse, G. (2018). Notice Failure Revisited: Evidence on the Use of Virtual Patent Marking. NBER Working Paper No. 24288.

Della Malva, A. and K. Hussinger (2012). Corporate Science in the Patent System: An Analysis of the Semiconductor Technology. *Journal of Economic Behavior & Organization 84*(1), 118–135.

DiMasi, J. A., H. G. Grabowski, and R. W. Hansen (2016). Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs. *Journal of Health Economics 47*, 20–33.

Dosi, G. (1982). Technological Paradigms and Technological Trajectories. A Suggested Interpretation of the Determinants and Directions of Technical Change. *Research Policy 11*(3), 147–162.

Dranove, D., C. Garthwaite, and M. I. Hermosilla (2020). Expected Profits and The Scientific Novelty of Innovation. NBER Working Paper No. 27093.

Dranove, D. and D. Meltzer (1994). Do Important Drugs Reach the Market Sooner? *The RAND Journal of Economics*, 402–423.

Dubois, P., O. De Mouzon, F. Scott-Morton, and P. Seabright (2015). Market Size and Pharmaceutical Innovation. *The RAND Journal of Economics 46*(4), 844–871.

Duggan, M. and F. Scott Morton (2010). The Effect of Medicare Part D on Pharmaceutical Prices and Utilization. *The American Economic Review 100*(1), 590–607.

Duggan, M. G. and F. Scott Morton (2011). The Medium-Term Impact of Medicare Part D on Pharmaceutical Prices. *The American Economic Review 101*(3), 387–92.

European Commission (2009). Pharmaceutical Sector Inquiry Final Report.

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

Foray, D. and F. Lissoni (2010). University Research and Public - Private Interaction. In B. H. Hall and N. Rosenberg (Eds.), *Handbook of the Economics of Innovation*, Volume 1, pp. 275–314. Elsevier.

Frakes, M. D. and M. F. Wasserman (2020). Investing in Ex Ante Regulation: Evidence from Pharmaceutical Patent Examination. NBER Working Paper No. 27579.

Franzoni, C. (2009). Do Scientists Get Fundamental Research Ideas by Solving Practical Problems? *Industrial and Corporate Change 18*(4), 671–699.

Furman, J. L. and S. Stern (2011). Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research. *The American Economic Review 101*(5), 1933–1963.

Furman, J. L. and F. Teodoridis (2020). Automation, Research Technology, and Researchers' Trajectories: Evidence from Computer Science and Electrical Engineering. *Organization Science 31*(2), 330–354.

Gaessler, F. and D. Byrski (2018). Overall Use of the SPC System in the EU. In R. Romandini, R. Hilty, and A. Kur (Eds.), *Study on the Legal Aspects of Supplementary Protection Certificates in the EU - Final Report*, pp. 91–105. European Commission.

Gaessler, F., D. Harhoff, and S. Sorg (2019). Bargaining Failure and Freedom to Operate: Re-Evaluating the Effect of Patents on Cumulative Innovation. Max Planck Institute for Innovation & Competition Research Paper No. 19-11.

Gaessler, F. and S. Wagner (2020). Patents, Data Exclusivity, and the Development of New Drugs. *The Review of Economics and Statistics*, 1–49.

Galasso, A. and M. Schankerman (2015). Patents and Cumulative Innovation: Causal Evidence from the Courts. *The Quarterly Journal of Economics 130*(1), 317–369.

Gans, J. S., D. H. Hsu, and S. Stern (2008). The Impact of Uncertain Intellectual Property Rights on the Market for Ideas: Evidence from Patent Grant Delays. *Management Science 54*(5), 982–997.

Garthwaite, C., R. Sachs, and A. D. Stern (2020). Which Markets (Don't) Drive Pharmaceutical Innovation? Evidence from U.S. Medicaid Expansions. mimeo.

Gilchrist, D. S. (2016). Patents as a Spur to Subsequent Innovation? Evidence from Pharmaceuticals. *American Economic Journal: Applied Economics 8*(4), 189–221.

Greve, H. R. and M. D. L. Seidel (2015, apr). The Thin Red Line between Success and Failure: Path Dependence in the Diffusion of Innovative Production Technologies. *Strategic Management Journal 36*(4), 475–496.

Griliches, Z. (1957). Hybrid Corn: An Exploration in the Economics of Technological Change. *Econometrica, Journal of the Econometric Society*, 501–522.

Griliches, Z. (1981). Market Value, R&D, and Patents. *Economics Letters 7*(2), 183–187.

Gupta, C. (2020). One Product, Many Patents: Imperfect Intellectual Property Rights in the Pharmaceutical Industry. *Available at SSRN 3748158*.

Gurgula, O. (2020). Strategic Patenting by Pharmaceutical Companies - Should Competition Law Intervene? *IIC-International Review of Intellectual Property and Competition Law 51*(9), 1062–1085.

Hall, B. H. and D. Harhoff (2012). Recent Research on the Economics of Patents. *Annual Review of Economics 4*(1), 541–565.

Hall, B. H. and J. Mairesse (1995). Exploring the Relationship between R&D and Productivity in French Manufacturing Firms. *Journal of Econometrics 65*(1), 263–293.

Hall, B. H., J. Mairesse, and P. Mohnen (2010). Measuring the Returns to R&D. In B. H. Hall and N. Rosenberg (Eds.), *Handbook of the Economics of Innovation*, Volume 2, pp. 1033–1082. Elsevier.

Harhoff, D., F. Narin, F. M. Scherer, and K. Vopel (1999). Citation Frequency and the Value of Patented Inventions. *The Review of Economics and Statistics 81*(3), 511–515.

Harhoff, D., F. M. Scherer, and K. Vopel (2003). Citations, Family Size, Opposition and the Value of Patent Rights. *Research Policy 32*(8), 1343–1363.

Harhoff, D. and S. Wagner (2009). The Duration of Patent Examination at the European Patent Office. *Management Science 55*(12), 1969–1984.

Hausman, J. A., B. H. Hall, and Z. Griliches (1984). Econometric Models for Count Data with an Application to the Patents-R&D Relationship. *Econometrica 52*(4), 909–938.

Hemphill, C. S. and B. N. Sampat (2012). Evergreening, Patent Challenges, and Effective Market Life in Pharmaceuticals. *Journal of Health Economics 31*(2), 327–339.

Henderson, R. and I. Cockburn (1996). Scale, Scope, and Spillovers: The Determinants of Research Productivity in Drug Discovery. *The RAND Journal of Economics*, 32–59.

Hermosilla, M. and Y. Wu (2018). Market Size and Innovation: The Intermediary Role of Technology Licensing. *Research Policy 47*(5), 980–991.

Higgins, M. J., M. J. Kronlund, J. M. Park, and J. Pollet (2020). The Role of Assets in Place: Loss of Market Exclusivity and Investment. NBER Working Paper No. 27588.

Higgins, M. J. and D. Rodriguez (2006). The Outsourcing of R&D through Acquisitions in the Pharmaceutical Industry. *Journal of Financial Economics 80*(2), 351–383.

Hill, R. (2019). Searching for Superstars: Research Risk and Talent Discovery in Astronomy. mimeo.

Howard, D. H., M. E. Chernew, T. Abdelgawad, G. L. Smith, J. Sollano, and D. C. Grabowski (2016). New Anticancer Drugs Associated with Large Increases in Costs and Life Expectancy. *Health Affairs 35*(9), 1581–1587.

Howard, L. (2007). Use of Patents in Drug Lifecycle Management. *Journal of Generic Medicines 4*(3), 230–236.

Huckman, R. S. (2003). The Utilization of Competing Technologies Within the Firm: Evidence from Cardiac Procedures. *Management Science 49*(5), 599–617.

Hvide, H. K. and B. F. Jones (2018). University Innovation and the Professor's Privilege. *The American Economic Review 108*(7), 1860–98.

Iaria, A., C. Schwarz, and F. Waldinger (2018). Frontier Knowledge and Scientific Production: Evidence from the Collapse of International Science. *The Quarterly Journal of Economics 133*(2), 927–991.

International Committee on Standardized Genetic Nomenclature for Mice (2016). Guidelines for Nomenclature of Mouse and Rat Strains. Technical report.

Izhak, O., T. Saxell, and T. Takalo (2020). Optimal Patent Policy for Pharmaceutical Industry. VATT Institute for Economic Research Working Papers No. 131.

Jacob, B. A. and L. Lefgren (2011). The Impact of Research Grant Funding on Scientific Productivity. *Journal of Public Economics 95*(9), 1168–1177.

Jaravel, X., N. Petkova, and A. Bell (2018). Team-Specific Capital and Innovation. *The American Economic Review 108*(4-5), 1034–73.

Jayachandran, S., A. Lleras-Muney, and K. V. Smith (2010). Modern Medicine and the Twentieth Century Decline in Mortality: Evidence on the Impact of Sulfa Drugs. *American Economic Journal: Applied Economics 2*(2), 118–46.

Kallenbach, L. and M. K. Vallazza (2018). Are the New Clinical Trial Transparency Rules Incompatible with the Patentability Requirements in Europe? *Nature Biotechnology 36*(10), 928–930.

Klemperer, P. (1987). Markets with Consumer Switching Costs. *The Quarterly Journal of Economics 102*(2), 375–394.

Kolympiris, C., K. Drivas, M. Helsby, and A. Chalmers (2021). How Scientists Search. mimeo.

Kremer, M. and H. Williams (2010). Incentivizing Innovation: Adding to the Tool Kit. *Innovation Policy and the Economy 10*(1), 1–17.

Krieger, J. L., D. Li, and D. Papanikolaou (2018). Missing Novelty in Drug Development. NBER Working Paper No. 24595.

Kyle, M. (2017). Economic Analysis of Supplementary Protection Certificates in Europe. European Commission/MINES ParisTech (CERNA) Working Paper.

Kyle, M. K. (2020). The Alignment of Innovation Policy and Social Welfare: Evidence from Pharmaceuticals. *Innovation Policy and the Economy 20*(1), 95–123.

Kyle, M. K. and A. M. McGahan (2012). Investments in Pharmaceuticals Before and After TRIPS. *The Review of Economics and Statistics 94*(4), 1157–1172.

Lach, S. and M. Schankerman (2008). Incentives and Invention in Universities. *The RAND Journal of Economics 39*(2), 403–433.

Lakdawalla, D. N. (2018). Economics of the Pharmaceutical Industry. *Journal of Economic Literature 56*(2), 397–449.

Lander, B. (2013). Sectoral Collaboration in Biomedical Research and Development. *Scientometrics 94*(1), 343–357.

Lander, B. and J. Atkinson-Grosjean (2011). Translational Science and the Hidden Research System in Universities and Academic Hospitals: A Case Study. *Social Science & Medicine 72*(4), 537–544.

Levin, S. G. and P. E. Stephan (1991). Research Productivity over the Life Cycle: Evidence for Academic Scientists. *The American Economic Review*, 114–132.

Lichtenberg, F. R. (2018). The Impact of New Drug Launches on Life-Years Lost in 2015 from 19 Types of Cancer in 36 Countries. *Journal of Demographic Economics 84*(3), 309–354.

Lichtenberg, F. R. (2019). How Many Life-Years Have New Drugs Saved? A Three-Way Fixed-Effects Analysis of 66 Diseases in 27 Countries, 2000-2013. *International Health 11*(5), 403–416.

Lichtenberg, F. R. and J. Waldfogel (2009). Does Misery Love Company? Evidence from Pharmaceutical Markets Before and After the Orphan Drug Act. *Michigan Telecommunications and Technology Law Review 46*(4), 844–871.

Lutz, C. M., C. C. Linder, and M. T. Davisson (2012). *Strains, Stocks and Mutant Mice in The Laboratory Mouse*. Academic Press.

Lyon, M. F. and A. G. Searle (1989). *Genetic Variants and Strains of the Laboratory Mouse*. Oxford University Press.

Malakoff, D. (2000). The Rise of the Mouse, Biomedicine's Model Mammal. *Science 288*(5464), 248–253.

Mansfield, E. (1986). Patents and Innovation: An Empirical Study. *Management Science 32*(2), 173–181.

Martinez, C. (2010). Insight into Different Types of Patent Families. OECD Directorate for Science, Technology and Innovation Working Paper.

BIBLIOGRAPHY

Marx, M. and A. Fuegi (2020). Reliance on Science: Worldwide Front-Page Patent Citations to Scientific Articles. *Strategic Management Journal*.

Medicare Trustees Report (2006). Annual Report on the Board of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds. Technical report.

Mejer, M. (2017). 25 Years of SPC Protection for Medical Products in Europe: Insights and Challenges. *European Commission Working Paper*.

Mello, M., J. Francer, M. Wilenzick, P. Teden, B. Bierer, and M. Barnes (2013). Preparing for Responsible Sharing of Clinical Trial Data. *The New England Journal of Medicine 369*(17), 1651–1658.

Merton, R. K. (1973). *The Sociology of Science: Theoretical and Empirical Investigations*. University of Chicago Press.

Mokyr, J. (2002). *The Gifts of Athena: Historical Origins of the Knowledge Economy*. Princeton University Press.

Morton, F. S. and M. Kyle (2011). Markets for Pharmaceutical Products. In M. V. Pauly, T. G. McGuire, and P. P. Barros (Eds.), *Handbook of Health Economics*, Volume 2, pp. 763–823. Elsevier.

Mouse Genome Sequencing Consortium and others (2002). Initial Sequencing and Comparative Analysis of the Mouse Genome. *Nature 420*(6915), 520.

Mowery, D. and N. Rosenberg (1979). The Influence of Market Demand upon Innovation: A Critical Review of some Recent Empirical Studies. *Research Policy 8*(2), 102–153.

Murray, F., P. Aghion, M. Dewatripont, J. Kolev, and S. Stern (2016). Of Mice and Academics: Examining the Effect of Openness on Innovation. *American Economic Journal: Economic Policy 8*(1), 212–252.

Myers, K. (2020). The Elasticity of Science. *American Economic Journal: Applied Economics 12*(4), 103–34.

Nagler, M. and S. Sorg (2020). The Disciplinary Effect of Post-Grant Review - Causal Evidence from European Patent Opposition. *Research Policy 49*(3), 103915.

Nelson, R. R. (1959). The Simple Economics of Basic Scientific Research. *Journal of Political Economy 67*(3), 297–306.

Nevalainen, T. (2014, 12). Animal Husbandry and Experimental Design. *ILAR Journal 55*(3), 392–398.

Nordhaus, W. D. (1969). *Invention, Growth, and Welfare: A Theoretical Treatment of Technological Change*. Cambridge, MA: MIT Press.

Oettl, A. (2012). Reconceptualizing Stars: Scientist Helpfulness and Peer Performance. *Management Science 58*(6), 1122–1140.

Owen-Smith, J. and W. W. Powell (2001). To Patent or Not: Faculty Decisions and Institutional Success at Technology Transfer. *The Journal of Technology Transfer 26*(1), 99–114.

Packalen, M. and J. Bhattacharya (2018). Does the NIH Fund Edge Science? NBER Working Paper No. 24860.

Paigen, K. (1995). A Miracle Enough: the Power of Mice. *Nature Medicine 1*(3), 215.

Pakes, A. and M. Schankerman (1984). The Rate of Obsolescence of Patents, Research Gestation Lags, and the Private Rate of Return to Research Resources. In Z. Griliches (Ed.), *R&D, Patents, and Productivity*, pp. 73–88. University of Chicago Press.

Poege, F., D. Harhoff, F. Gaessler, and S. Baruffaldi (2019). Science Quality and the Value of Inventions. *Science Advances 5*(12). eaay7323.

Prinz, D., M. Chernew, D. Cutler, and A. Frakt (2018). Health and Economic Activity over the Lifecycle: Literature Review. NBER Working Paper No. 24865.

Roach, M. and H. Sauermann (2010). A Taste for Science? PhD Scientists' Academic Orientation and Self-Selection into Research Careers in Industry. *Research policy 39*(3), 422–434.

Robinson, W. T., G. Kalyanaram, and G. L. Urban (1994). First-Mover Advantages from Pioneering New Markets: A Survey of Empirical Evidence. *Review of Industrial Organization 9*(1), 1–23.

Rogers, F. B. (1963). Medical Subject Headings - Communication to the Editor. *Bulletin of the Medical Library Association 51*(1), 114–116.

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

Rose, M. E. and J. R. Kitchin (2019). Pybliometrics: Scriptable Bibliometrics Using a Python Interface to Scopus. *SoftwareX 10*, 100263.

Rosenberg, N. (1992). Scientific Instrumentation and University Research. *Research Policy 21*(4), 381–390.

Sampat, B. and H. L. Williams (2019). How Do Patents Affect Follow-on Innovation? Evidence from the Human Genome. *The American Economic Review 109*(1), 203–236.

Scherer, F. M. (2010). Pharmaceutical Innovation. *Handbook of the Economics of Innovation 1*, 539–574.

Schmidheiny, K. and S. Siegloch (2019). On Event Study Designs and Distributed-Lag Models: Equivalence, Generalization and Practical Implications. CEPR Discussion Paper No. DP13477.

Schmookler, J. (1966). *Invention and Economic Growth*. Harvard University Press.

Schrepel, T. (2018). Predatory Innovation: The Definite Need for Legal Recognition. *SMU Science & Technology Law Review 21*, 19.

Schumpeter, J. A. (1939). *Business Cycles*. McGraw-Hill.

Seong, E., T. L. Saunders, C. L. Stewart, and M. Burmeister (2004). To Knockout in 129 or in C57BL/6: That is the Question. *Trends in Genetics 20*(2), 59–62.

Smith, P. W. (2006). The National Institutes of Health (NIH): Organization, Funding, and Congressional Issues. CRS Report for Congress RL33695. Library of Congress, Washington, DC.

Stephan, P. E. (1996). The Economics of Science. *Journal of Economic literature 34*(3), 1199–1235.

Stephan, P. E. (2012). *How Economics Shapes Science*. Harvard University Press.

Stern, S. (2004). Do Scientists Pay to be Scientists? *Management Science 50*(6), 835–853.

Sternitzke, C. (2010). Knowledge Sources, Patent Protection, and Commercialization of Pharmaceutical Innovations. *Research Policy 39*(6), 810–821.

Sternitzke, C. (2013). An Exploratory Analysis of Patent Fencing in Pharmaceuticals: The Case of PDE5 Inhibitors. *Research Policy 42*(2), 542–551.

Stock, J. H., M. Yogo, et al. (2005). Testing for Weak Instruments in Linear IV Regression. *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg 80*(4.2), 1.

Teodoridis, F. (2017). Understanding Team Knowledge Production: The Interrelated Roles of Technology and Expertise. *Management Science 64*(8), 3625–3648.

The Jackson Laboratory (2009a). *Breeding Strategies for Maintaining Colonies of Laboratory Mice - A Jackson Laboratory Resource Manual*.

The Jackson Laboratory (2009b). *The Jackson Laboratory Handbook on Genetically Standardized Mice*.

BIBLIOGRAPHY

The Jackson Laboratory Archives (2012a). *JAX Archives Box Inventory List*, Box Inventory. 3. The Jackson Laboratory, The Mouseion at the JAXlibrary.

The Jackson Laboratory Archives (2012b). *JAX Archives Complete Inventory List*, Complete Inventory. 1. The Jackson Laboratory, The Mouseion at the JAXlibrary.

Thursby, J. G., R. Jensen, and M. C. Thursby (2001). Objectives, Characteristics and Outcomes of University Licensing: A Survey of Major US Universities. *The Journal of Technology Transfer 26*(1), 59–72.

Thursby, J. G. and M. C. Thursby (2011). Has the Bayh-Dole Act Compromised Basic Research? *Research Policy 40*(8), 1077–1083.

Tijssen, R. J. (2010). Discarding the 'Basic Science/Applied Science'Dichotomy: A Knowledge Utilization Triangle Classification System of Research Journals. *Journal of the American Society for Information Science and Technology 61*(9), 1842–1852.

Toole, A. A. (2012). The Impact of Public Basic Research on Industrial Innovation: Evidence from the Pharmaceutical Industry. *Research Policy 41*(1), 1–12.

Von Graevenitz, G., B. H. Hall, D. Harhoff, S. Wagner, A. Gambardella, P. Giuri, and K. Hoisl (2007). The Strategic Use of Patents and its Implications for Enterprise and Competition Policies. Final Report to the European Commission.

Wagner, S. and S. Wakeman (2016). What do Patent-Based Measures Tell us about Product Commercialization? Evidence from the Pharmaceutical Industry. *Research Policy 45*(5), 1091–1102.

Waldinger, F. (2016). Bombs, Brains, and Science: The Role of Human and Physical Capital for the Creation of Scientific Knowledge. *The Review of Economics and Statistics 98*(5), 811–831.

Walsh, J. P., W. M. Cohen, and C. Cho (2007). Where Excludability Matters: Material versus Intellectual Property in Academic Biomedical Research. *Research Policy 36*(8), 1184–1203.

Ward, M. R. and D. Dranove (1995). The Vertical Chain of Research and Development in the Pharmaceutical Industry. *Economic Inquiry 33*(1), 70–87.

Watzinger, M. and M. Schnitzer (2019). Standing on the Shoulders of Science. CEPR Discussion Paper No. DP13766.

Weil, D. N. (2014). Health and Economic Growth. In P. Aghion and S. N. Durlauf (Eds.), *Handbook of Economic Growth*, Volume 2, pp. 623–682. Elsevier.

Whitney, C. G., F. Zhou, J. Singleton, and A. Schuchat (2014). Benefits from Immunization During the Vaccines for Children Program Era—United States, 1994-2013. *MMWR. Morbidity and Mortality Weekly Report 63*(16), 352.

Williams, H. L. (2013). Intellectual Property Rights and Innovation: Evidence from the Human Genome. *Journal of Political Economy 121*(1), 1–27.

Yin, N. (2017). Pharmaceuticals, Incremental Innovation and Market Exclusivity. mimeo.

Zyontz, S. (2019). Running with (CRISPR) Scissors: Tool Adoption and Team Assembly. mimeo.