Essays in the Economics of Health and Innovation

Essays in the Economics of Health and Innovation

A dissertation presented

by

Jennifer Kao

to

The Department of Public Policy

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Public Policy

Harvard University

Cambridge, Massachusetts

May 2019

Dissertation Advisors:
Professor David Cutler
Professor Pierre Azoulay

Author:
Jennifer Kao

Essays in the Economics of Health and Innovation

Abstract

This dissertation explores three questions at the intersection of health economics and the economics of innovation. The first chapter of this dissertation explores how publicly available scientific information shapes the quantity and profitability of private-sector research. I examine the impact of large-scale cancer genome mapping studies, which systematically map the genetic abnormalities in cancer, on research productivity in the pharmaceutical industry. Using a newly-constructed dataset from cancer genome mapping studies and clinical trials, I find that mapping information increases private-sector investment in clinical trials by nearly 50 percent. Considering the types of private-sector research investments, I find that cancer mapping significantly increases trials evaluating drugs previously approved or tested for one disease in an additional disease. Using trial results reported in abstracts submitted to a major cancer conference, I also find that cancer mapping information increases the profitability of firms' research decisions: when genetic information is known, firms are more likely terminate drug investments that are unlikely to be successful in the long run and to continue investment projects that are most likely to generate promising clinical results. This evidence suggests that publicly available, detailed scientific maps can increase and improve private research efforts.

The second chapter considers firms' incentives to disclose information about novel medical products. Using a dataset of 16,000 clinical trials over a 17-year period, I examine the relationship between competition and pharmaceutical firms' incentives to disclose clinical trial information regarding drug safety and efficacy in the public database, ClinicalTrials.gov. I find

evidence suggesting that greater competition spurs firms to positively influence consumers' perceptions of drug quality through investments in well-designed and more specialized trials. Further, I find that firms experiencing more competition are 24 percent more likely to disclose trial results, even after controlling for drug and firm characteristics. The results suggest that investment in perceived product quality and the subsequent disclosure of quality information is an important firm strategy for market entry.

The final chapter examines what happens when an institution that supports the speed and diversity of medical research is strained and is based on joint work with Pierre Azoulay and Misty Heggeness. Academic Medical Centers (AMCs)—comprising medical schools, teaching hospitals, and their affiliated physicians, residents, and students play an important role in the American system of biomedical research and innovation. We consider how changes in the level of health care financing affect research productivity within academic medical centers (AMCs). We examine the role of the Balanced Budget Act of 1997, which changed the formula used to reimburse Medicare inpatient claims and teaching hospital subsidies, on research outcomes within AMCs. We compare AMCs' relative exposure to the reform and how these differences affect their researchers' ability to attract scientific grant funding and produce scientific publications. We find that in response to the BBA, research activity falls by 4 percent among the average teaching hospital and nearly 7 percent among major teaching hospitals. Further, we analyze how changes in financing shift the quality and direction of research by examining heterogeneity across publication types. We find little evidence of concurrent changes in clinical outcomes. Our estimates offer insight into how changes in reimbursements to health care providers can shape the rate and direction of scientific progress within biomedical research.

Contents

Abstract . . . iii
Acknowledgments . . . xii

1 Charted Territory: Evidence from Mapping the Cancer Genome and R&D Decisions in the Pharmaceutical Industry 1
1.1 Introduction . . . 1
1.2 Case Study and Conceptual Framework . . . 5
1.2.1 A Large-Scale Ovarian Cancer Mapping Study . . . 5
1.2.2 Quantity Implications of the Ovarian Cancer Mapping Study . . . 7
1.2.3 Profitability Implications of the Ovarian Cancer Mapping Study . . . 11
1.3 Empirical Setting and Data . . . 13
1.3.1 Scientific Background . . . 13
1.3.2 Large-Scale Cancer Genome Mapping Efforts . . . 14
1.3.3 Private Research Investments . . . 18
1.4 Effects on Quantity of Private Research Investments . . . 23
1.4.1 Empirical Strategy . . . 23
1.4.2 Sample and Descriptive Statistics . . . 24
1.4.3 Estimating Equation and Assumptions . . . 29
1.4.4 Results . . . 32
1.5 Effects on Profitability of Firms' Research Decisions . . . 44
1.5.1 Empirical Strategy . . . 44
1.5.2 Sample and Descriptive Statistics . . . 45
1.5.3 Results . . . 48
1.6 Conclusion . . . 55

2 Competition and Disclosure Incentives: Evidence from the Pharmaceutical Industry 57
2.1 Introduction . . . 57
2.2 Background and Conceptual Framework . . . 60
2.2.1 Results Reporting in the Pharmaceutical Industry . . . 60

2.2.2 ClinicalTrials.gov and Trial Results Disclosure . . . 61
2.2.3 Disclosure Implications of Competition . . . 64
2.3 Data and Description Statistics . . . 66
2.4 Empirical Equation and Assumptions . . . 75
2.4.1 Empirical Strategy . . . 75
2.4.2 Empirical Equation . . . 77
2.4.3 Assumptions . . . 77
2.5 Results . . . 78
2.5.1 Impact on Quality . . . 78
2.5.2 Impact on Disclosure . . . 81
2.6 Conclusion . . . 83

3 Medical Progress and Health Care Financing: Evidence from Academic Medical Centers 85
3.1 Introduction . . . 85
3.2 Background and Conceptual Framework . . . 89
3.2.1 Academic Medical Centers and Life Sciences Research . . . 89
3.2.2 Research Funding within Academic Medical Centers . . . 91
3.2.3 Medicare Payments and Financing Cuts . . . 92
3.3 Data and Descriptive Statistics . . . 94
3.4 Empirical Strategy and Results . . . 105
3.4.1 Impact on the Rate of Innovation . . . 108
3.4.2 Impact on the Quality of Innovation . . . 114
3.4.3 Impact on the Direction of Innovation . . . 114
3.5 Discussion and Conclusion . . . 120
3.5.1 Discussion . . . 120
3.5.2 Conclusion . . . 125

References 126

Appendix A Appendix to Chapter 1 139
A.1 Data Description . . . 139
A.1.1 Cancer Mapping Data . . . 139
A.1.2 Identifying Well-Designed and Non-Well-designed Trials . . . 141
A.2 Additional Figures and Tables . . . 142

Appendix B Appendix to Chapter 2 153
B.1 Identifying Clinical Trials . . . 153

B.2 Example ClinicalTrials.gov Trial Record 155
B.3 ClinicalTrials.gov Results Template 157
B.4 Additional Figures and Tables 158

Appendix C Appendix to Chapter 3 **163**
C.1 Medicare Inpatient Reimbursement 163
C.2 Mapping Outcomes to Hospitals 167
C.3 Academic Medical Center Sample Results 168

List of Tables

1.1 Overview of Gene-Cancer-Year Panel Construction 25
1.2 Summary Statistics: Gene-Cancer Level Data 27
1.3 Does Cancer Mapping Information Influence the Quantity of Trials? 33
1.4 Impact on Trial Quantity: Heterogeneity by Clinical Relevance of Cancer Mapping Information 38
1.5 What Types of Clinical Trials: New Uses or Novel Drugs? 42
1.6 Summary Statistics: Trial-Gene-Cancer Level Data 47
1.7 Impact of Cancer Mapping Information on Phase II Clinical Outcomes . . . 49
1.8 Impact of Cancer Mapping Information on Phase II to Phase III Transitions 52
1.9 Impact of Mapping Information on Phase III Clinical Outcomes 54

2.1 Summary Stats: Trial-Cancer Level 72
2.2 Differences Across High and Low Competition Trials 76
2.3 Competition and Trial Design 79
2.4 Competition and Trial Specialization 80
2.5 Competition and Disclosure 82

3.1 Summary Statistics 100
3.2 Impact on the Number of Grant Applications 109
3.3 Impact on the Number of Publications, Quality 112
3.4 Impact on Research Direction 116
3.5 Summary of 30-Day Risk-adjusted Survival Rates 121
3.6 Impact on Clinical Outcomes 124

A.1 Impact of Mapping Information, Excluding Genes Affected by Patent Regulation 146
A.2 Impact of Mapping Information, Excluding Genes Affected by Patent Regulation 147
A.3 Impact of Mapping Information in Same Gene, Related Cancer 148
A.4 Impact of Mapping Information for Different Types of Firms 149
A.5 Impact of Mapping Information for Different Types of Diseases 150

A.6 Impact of Mapping Information for Different Trial Design Types 151
A.7 Phase II Outcomes and Phase II to Phase III Transitions 152

B.1 Identifying Analytic Trial Sample 154
B.2 Summary Stats: Trial-Cancer Level 158
B.3 Competition and Trial Design, Competition Since 2000 159
B.4 Competition and Trial Specialization, Competition Since 2000 160
B.5 Competition and Disclosure, Competition Since 2000 161

C.1 Summary Statistics 171
C.2 Comparing the Primary Analytic and AMC Sample 172
C.3 Impact on the Number of Grant Applications 173
C.4 Impact on the Number of Publications, Quality 174
C.5 Impact on Research Direction 175
C.6 Impact on Clinical Outcomes 176

List of Figures

1.1 Trials Enrolling Patients with Ovarian Cancer and BRCA2 Gene Mutations . 10
1.2 Total Cancer Mapping Studies and Mapped Tumors by Year 16
1.3 Share of Cancer Trials that Enroll Patients Based on Genes, by Year 21
1.4 Cumulative Share of Gene-Cancer Pairs with Mutations Identified by Mapping Studies 28
1.5 Examining Cancer-Level Selection 31
1.6 Event Study Estimates—Impact of Cancer Mapping Information on Trial Quantity 36

2.1 Cumulative Percentage of Clinical Trials that Report Results 68
2.2 Diseases Tested and Trial Outcomes 70
2.3 Firm Size and Share of Trials Disclosed 74

3.1 Data Sources and Variable Construction 95
3.2 Distribution of Research Outcomes 101
3.3 Change in NIH Grant Applications, 1992-2007 102
3.4 Change in Publications, 1992-2007 103
3.5 Change in Industry-Sponsored Trials, 1994-2007 104
3.6 Distribution of Medicare Shares 106
3.7 Impact on the Number of Grant Applications 111
3.8 Impact on the Number of Grant Application 113
3.9 Impact on Bench Research (Number of Publications Cited by a Patent) . . . 117
3.10 Impact on Translational Research (Number of "Translational" Publications) 118
3.11 Impact on Bedside Research (Number of Industry-Sponsored Trial Sites) . . 119
3.12 Risk-Adjusted 30-Day Survival Rates 122

A.1 Trials Enrolling Patients with Ovarian Cancer and BRCA2 Gene Mutations, Olaparib Only 142
A.2 Overview of Scientific Background on Cancer Genome Sequencing 143
A.3 Event Study Estimates—Impact of Mapping on Trial Quantity, Trials with Non-Missing Intervention 144

A.4 Trial Advancement Rates, by Year 145

B.1 Example Clinical Trial Record 155
B.2 Example Clinical Trial Results Record 156
B.3 ClinicalTrials.gov Results Template 157

C.1 Medicare Share and Disproportionate Medical Share Payments 166
C.2 Medicare Share and Indirect Medical Education Payments 166
C.3 Mapping Research Activity to Hospitals: An Example 167
C.4 Impact on the Number of Grant Applications 168
C.5 Impact on the Number of Publications 169
C.6 Impact on Bedside Research (Number of Industry-Sponsored Trial Sites) . . 170

Acknowledgments

First, I thank my wonderful advisors David Cutler, Pierre Azoulay, and Amitabh Chandra. David Cutler's insightful guidance paired with his kindness were invaluable during my doctoral studies. I could not be more grateful for Pierre Azoulay's mentorship. From practical life advice to providing detailed feedback, Pierre was a major contribution to my development as a scholar. Amitabh Chandra encouraged me early on to work on topics related to health and innovation. He taught me to always think about the "big picture" and to bring excitement and curiosity to everything that I do.

I benefited immensely from conversations with Joshua Krieger, Leemore Dafny, Ramana Nanda, and Mark Shephard. I am particularly grateful to Ariel Stern, who was a mentor from the start. Ariel has shown me what it means to be a thoughtful, warm, and kind scholar. I am incredibly grateful to several undergraduate professors and graduate students who encouraged my interest in research. From my undergraduate training at UC Berkeley, I thank Sally K. Fairfax for introducing me to research (and for teaching me early on that research meetings should always include good food). I am so grateful to Sofia Berto Villas-Boas, Tiffany Shih, Catherine Wolfram, and Anne Joseph O'Connell for encouraging me to pursue the academic path.

I could not have completed this dissertation without my classmates in Cambridge. Participants at the NBER Productivity Lunch, NBER Pre-Doctoral Aging and Health Lunch, Harvard Medical School Health Economics Seminar, and Harvard Economics Department Labor Lunch provided invaluable feedback. For providing support and camaraderie, I thank Megan Bailey, Gabriel Tourek, Stephen Coussens, Caitlin Carroll, Abe Holland, Ariella Kahn-Lang, Rebecca Sachs, Maria Steenland, and Daniel Velez-Lopez.

This work was made possible with financial support from Harvard University, the National Institute on Aging (through Grant Number T32-AG000186 and Award Number R24AG048059 to the National Bureau of Economic Research), and the Harvard Environmental Economics

Program.

Last, but not least, thank you to my family and friends. My parents, Andrew and Haihwa, have shown me unconditional love and support throughout my life. I can't imagine a more caring and encouraging sibling than my sister, Christine. My deepest gratitude to Priya Chandrasekaran, Marianne Potvin, Sonia Tycko, and Alexandra Batchelor for supporting me and making me laugh until it hurt. Wolfgang Ridinger taught me so much and his sense of curiosity continues to inspire me to this day. I am so grateful for the love, support, and encouragement from John Connolly. Finally, my deepest love to Jenny Shankey, who was one of the most creative people I have ever known and whose spirit remains an inspiration.

To my family and Jenny

Chapter 1

Charted Territory: Evidence from Mapping the Cancer Genome and R&D Decisions in the Pharmaceutical Industry

1.1 Introduction

How does publicly available basic scientific information influence the quantity and profitability of private-sector research? On the one hand, fundamental shifts in basic scientific knowledge may spur and sharpen firms' research and development efforts, through improving technological search and lowering entry costs.[1] On the other hand, the public release of detailed scientific data may reinforce the value of existing products or lower the cost of entry for competitors. This may diminish the incentive of potential entrants to enter and cause subsequent private research efforts to be, on net, lowered or unchanged. Thus, the impact

[1]In this paper, "basic scientific knowledge" refers to knowledge about the "fundamental aspects of phenomena and of observable facts without specific applications towards processes or products in mind" (`https://grants.nih.gov/grants/glossary.htm`).

of comprehensive advances in publicly available basic scientific knowledge on private-sector research is ambiguous.

Most evaluations of basic science knowledge have focused on whether inputs to basic science, such as R&D grants or subsidies, increase private-sector research.[2] In this paper, I examine the effects of major outputs to basic science investments: publicly available basic scientific data from large-scale cancer genome mapping initiatives. Building on the foundation provided by the Human Genome Project (Jayaraj, 2018; Williams, 2013), cancer genome mapping initiatives systematically catalogue the genetic mutations that might drive the progression and growth of cancer.[3] These large-scale research initiatives aim to facilitate technological search through introducing and validating existing scientific knowledge about the disease (Fleming and Sorenson, 2004; Mardis, 2018). Large-scale cancer mapping efforts are believed to play an influential role in the development of novel cancer therapies: National Institutes of Health director Francis Collins and former National Cancer Institute deputy director Anna Barker noted that mapping the cancer genome would "help chart a new course across the complex landscape of human malignancies"(Barker and Collins, 2008).

This paper examines how the introduction of cancer "atlases" shapes the development of new therapies for cancer, a disease whose therapeutic market is the largest in terms of global spending—at $133 billion per year—and as the second leading cause of death in the United States, one in which advances yield tremendous value to society (CDC, 2018; IQVIA, 2018). Specifically, I focus on changes in clinical trials. Within the pharmaceutical industry, a key requirement for new product entry is the completion of U.S. Food and Drug Administration-required clinical trials, a risky and costly process. Only about 15 percent

[2]For example, Azoulay *et al.* (2018) examines the impact of National Institutes of Health funding on private sector patenting. For helpful reviews, see David *et al.* (2000) and Hall and Van Reenen (2000). A notable exception to the R&D subsidy literature is Nagaraj (2017), who examines the impact of publicly available satellite maps on discoveries by the gold exploration industry.

[3]In fact, the Human Genome Project (HGP) was largely motivated by a desire to enable future cancer mapping efforts and the development of cancer therapies. In one of the earliest commentaries calling for the HGP, the Nobel laureate Renato Dulbecco (1986) wrote "If we wish to learn more about cancer, we must now concentrate on the cellular genome."

of drug candidates successfully proceed from the start of clinical testing to approval, and estimated costs for bringing a drug to market are $2.6 billion (Danzon and Keuffel, 2014; DiMasi, 2001; DiMasi *et al.*, 2003, 2010; Hay *et al.*, 2014; Thomas *et al.*, 2016). The impact of the cancer mapping efforts on new product development depends in large part on the extent to which the basic science influences firms' clinical trials investment efforts.

In light of this, I focus on two issues: the *quantity* of clinical trials and the *profitability* of firms' research decisions. I assemble a new dataset of publicly available information produced by 168 large-scale cancer mapping efforts, linked to privately-funded clinical trials, over the period 2004-2016.[4] I observe many characteristics of the clinical trials, including the cancer types under investigation, the genetic criteria used for patient enrollment, the drug being tested, the sponsoring firm, the trial design, and the clinical outcomes.

I begin by investigating how the disclosure of publicly available cancer mapping information shapes the subsequent quantity of privately-funded clinical trials. Publication of results from large-scale cancer mapping efforts provides significant variation in the public disclosure that a mutation exists within a particular gene (e.g., BRCA2) in a specific cancer site (e.g., prostate). I isolate quasi-random variation in the timing that the information was submitted to prominent scientific journals using a gene-cancer-year level differences-in-difference framework. To address concerns over selection in the timing of mapping, I control for differences in "research potential" between different gene-cancers with gene-cancer fixed effects and control for secular changes in the pharmaceutical industry over time using year fixed effects and cancer-year linear time trends.

I find that mutation-related information disclosures from large-scale cancer genomic efforts increase private-sector investment in clinical trials by 50 percent. If gene-cancer pairs that received mutation-related information had counterfactually experienced the same level of investments as gene-cancer pairs that did not, there would have been up to 97 fewer privately-

[4]This includes mapping efforts from both government (e.g., National Institutes of Health) and non-government organizations (e.g., Johns Hopkins University) institutions.

funded clinical trials and 15 fewer cancer drugs.[5] These results are consistent with recent quasi-experimental work that shows how investments in basic science (Ahmadpoor and Jones, 2017; Azoulay *et al.*, 2018) and, in particular, how detailed *maps* of basic scientific information (Jayaraj, 2018; Nagaraj, 2017; Williams, 2013), can increase the level of private-sector research.

I next examine the hypothesis that one mechanism through which scientific maps increases private investment is by identifying linkages across research opportunities that were previously believed to be distantly related. In particular, genetic mapping can reveal that similar genetic aberrations underlie different cancers types and spur trials testing drugs approved or previously tested for one disease in an additional disease. To illustrate, cancer mapping may show that breast and prostate cancer share similar gene mutations, highlighting the potential for breast cancer drugs to be effective treatments for prostate cancer. Consistent with this hypothesis, I find that mapping significantly increases investment in trials testing drugs that are previously approved (140 percent) or tested (54 percent). In contrast, private investment in trials testing novel drugs remains unchanged. These findings contribute to a sparse body of research that considers incentives for firms to improve existing products.[6]

Finally, I explore whether cancer mapping information increases the profitability of firms' research decisions—in this case, the likelihood that firms make decisions that maximize their expected returns based on existing clinical information. Once firms initiate a clinical trial, they must complete a series of additional clinical trials, each with increasing cost and risk. At each point, firms must decide whether to continue or terminate investment. Clinical trial failures are expected, and one indicator of success is a firm's ability to "fail quickly"—i.e., maximizing their expected returns through minimizing resources allocated to drugs that are unlikely to be successful, and continuing investment in drugs that are most likely to generate promising clinical results and successfully come to market (Lendrem *et al.*, 2015; Spetzler *et al.*, 2016). In deciding whether to continue or terminate investment, firms may make low

[5]Here, "drugs" refers to an active ingredient treating a specific disease (See Section 1.3.3 for more details).

[6]This topic has largely been explored by legal scholars, see Eisenberg (2005) and Roin (2014).

profit decisions by dismissing or failing to understand existing clinical evidence: a review of AstraZeneca's drug pipeline revealed that 18% of failures occurred because a drug advanced to the next phase of clinical development despite weak evidence from earlier phases (Cook *et al.*, 2014).

Using trial-gene-cancer level data, I investigate whether mapping information is associated with increases in the profitability of firms' research decisions. I find that firms initiating trials in diseases with mutation-related information are 60 percent less likely to advance drugs with weak clinical evidence to the next phase, as compared to trials initiated in diseases without mutation-related information. Examining the outcomes of drugs chosen to advance, I find that cancer mapping information is associated with drugs that lead to greater improvements in patient survival in the next phase, even after controlling for disease and firm characteristics. These findings are consistent with anecdotal evidence that access to reliable scientific information increases the productivity of firms' research decisions by promoting earlier termination of drugs that are likely to fail and further investment in drugs that are most likely to be successful in the long run (Bujar *et al.*, 2017; Peck *et al.*, 2015; Sharpe and Keelin, 1998).

The remainder of the paper proceeds as follows. Section 1.2 presents a case study of my results using a single large-scale cancer mapping study. Section 1.3 introduces the empirical setting and the data. Section 1.4 analyzes the effect of cancer mapping on the quantity of privately-funded clinical trials. I examine the impact of cancer mapping on the profitability of firms' research decisions in Section 1.5. Finally, Section 1.6 concludes.

1.2 Case Study and Conceptual Framework

1.2.1 A Large-Scale Ovarian Cancer Mapping Study

The purpose of the cancer mapping efforts examined in this paper is to create a publicly available "mutational landscape" that serves as a foundation for subsequent cancer research. Large-scale cancer mapping efforts examine hundreds of patients in order to introduce novel

information about rare gene mutations that were previously overlooked by earlier, small-scale mapping efforts. By having a better understanding of a cancer's biological basis, firms can more easily develop drugs tailored to patient sub-groups with specific genetic features. These so-called "targeted" drugs may be more effective for those patients. This, in turn, may have ambiguous effects on private-sector research.

Consider, for example, the case of the *The Cancer Genome Atlas'* (TCGA) serous ovarian cancer study (TCGA, 2011a). Ovarian cancer is diagnosed in 22,000 women and is the fifth leading cause of cancer death among women in the United States (American Cancer Society, 2018). 85 percent of deaths are among patients with an aggressive ovarian cancer subtype called serous ovarian cancer (TCGA, 2011b). In this mapping study, TCGA researchers systematically catalogued the genetic mutations underlying more than 300 serous ovarian cancer tumors and submitted their findings to the journal *Nature* in 2010.

The TCGA ovarian study revealed that 21 percent of the tumors contained mutations in the BRCA1 and BRCA2 (collectively referred to here as "BRCA") genes, in addition to other genetic discoveries. Previous research had identified BRCA mutations in inherited ovarian cancer. However, the TCGA ovarian cancer study confirmed that mutations also occurred in non-inherited serous ovarian cancers. In light of these findings, TCGA researchers suggested that non-inherited serous ovarian cancer tumors could respond to poly (ADP-ribose) polymerase (or PARP) inhibitors. PARP inhibitors generate an anti-tumor effect: PARP and BRCA genes repair damaged DNA, which makes up genes. In tumors with mutated BRCA genes, PARP inhibitors prevent all potential DNA repair mechanisms, which can ultimately cause cancer cell death.[7] At the time of the TCGA study, PARP inhibitors were already being testing in clinical trials and used to treat other forms of ovarian and breast cancer with mutated BRCA genes.

[7]For more details on PARP inhibitors, BRCA mutations, and ovarian cancer, see: Bryant *et al.* (2005); Farmer *et al.* (2005); Lijima *et al.* (2017).

1.2.2 Quantity Implications of the Ovarian Cancer Mapping Study

The TCGA's ovarian cancer study has uncertain effects on the quantity of clinical trials enrolling BRCA-mutated non-inherited serous ovarian cancer patients. On the one hand, the TCGA's mapping information may introduce or validate existing information that assists firms in identifying subgroups of patients that respond most favorably to treatment. This is consistent with the theory that invention is a process of searching for better combinations of components (in this case, drugs and diseases) and that science facilitates the efficient identification of useful, new combinations (Fleming and Sorenson, 2004). A more efficient drug-disease (or drug-patient) match could enable firms to conduct trials with fewer patients and over a short duration, ultimately lowering the cost of bringing a novel drug to market (Chandra *et al.*, 2018).[8] As a result, mapping increase the net level of clinical trials testing drugs that are already in the pipeline (e.g., PARP inhibitors) or that have not yet entered clinical development.

On the other hand, mapping information may lower or not affect the net level of subsequent investment. This may occur through three mechanisms. First, the TCGA's ovarian cancer effort may reveal non-novel information about the relationship between genes and cancers. Previous sequencing or non-sequencing efforts may have already revealed relationships between BRCA genes and different forms of ovarian cancer. For example, firms with significant resources may have their own in-house genomics research efforts, or partner with firms that specialize in genomics research.[9] To illustrate, the drug Olaparib, the first approved PARP-inhibitor, was clinically tested in several forms of BRCA-mutated ovarian and breast cancer prior to 2010. This suggests that its manufacturer, AstraZeneca—a multinational pharmaceutical firm with $33.3 billion in sales in 2010—was already aware of several different

[8]Here, a drug refers to an active ingredient treating a specific disease.

[9]For example, FoundationMedicine—a firm that specializes in sequencing tumors and developing genetic tests for evaluating cancer—has a partnership with the pharmaceutical firm Pfizer which allows the company to benefit from access to FoundationMedicine's database of more than 200,000 tumor profiles.

diseases that could be effectively treated by Olaparib (AstraZeneca, 2011).[10] In the same vein, previous non-mapping efforts such as retrospective analyses may reveal that non-inherited serous ovarian cancer patients with BRCA mutations are most responsive to treatment, suggesting that this particular type of ovarian cancer is driven by BRCA mutations. Thus, publicly available mapping information may have no effect on the net level of private-sector innovation.

However, while the information produced by the TCGA's ovarian cancer effort may be non-novel, the information may still be useful: in describing the impact of TCGA's ovarian cancer study, a leading genomics expert at a pharmaceutical company, which manufacturers a PARP-inhibitor, confirmed that some of the information may have been known already. However, the TCGA's finding that 21 percent of ovarian cancers exhibit a BRCA mutation was helpful in validating existing hypotheses—regarding the share of serous ovarian cancer tumors with BRCA mutations—in a large sample.[11,12]

Second, mapping information may lower or not change the subsequent level of trials because certain firms may be able to take advantage of the information more readily and crowd out potential entrants. Specifically, manufacturers of existing PARP inhibitors may initiate clinical trials to treat individuals with non-inherited BRCA-mutated ovarian cancers. These firms have an advantage over new entrants (i.e., firms without a PARP inhibitors) because they may be able to skip several stages of the research and development process, such as earlier clinical trials that assess drug safety.[13]

---

[10]Indeed, an AstraZeneca's Annual Review notes: "In genomics, we have analysed more than 200,000 genomes (including data from *internal* and *external* databases) to inform investment decisions in drug discovery." Emphasis Added (AstraZeneca, 2017).

[11]Interviewed by author on April 3, 2018.

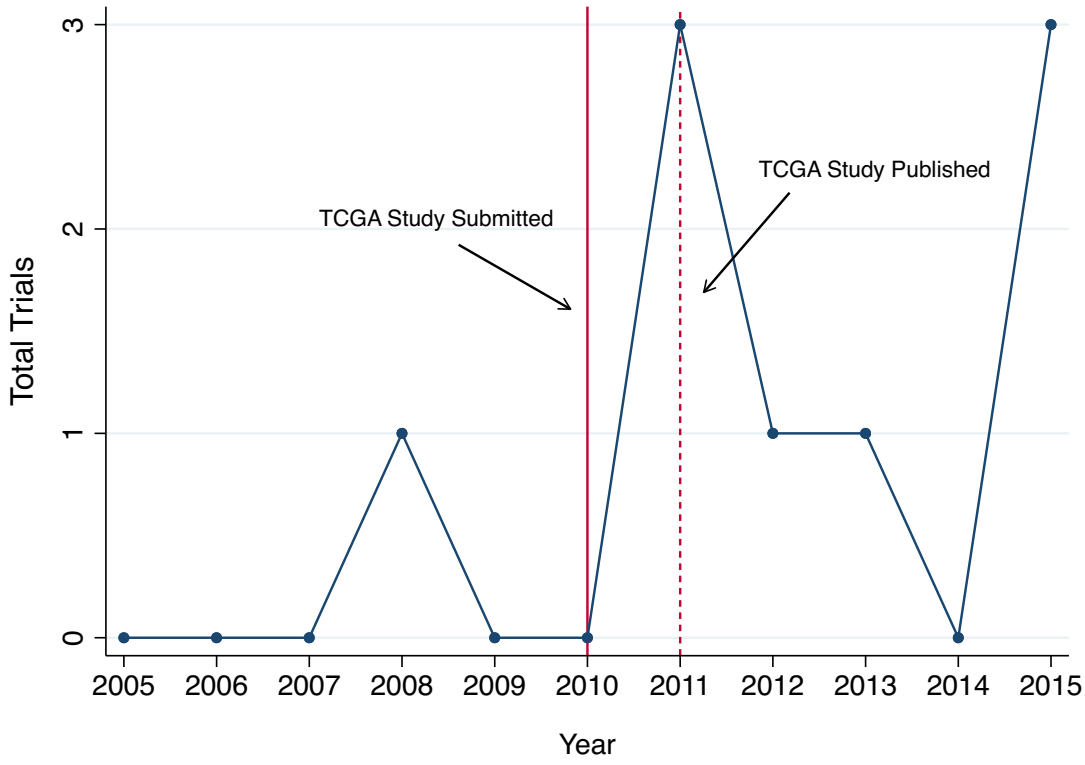
[12]To illustrate how pharmaceutical firms use the TCGA's results, a 2017 AstraZeneca study that examined the role of Olaparib in treating non-inherited mutations in serous ovarian cancer cited mutational prevalence estimates from the TCGA ovarian study (Dougherty *et al.*, 2012).

[13]This can apply to drugs that were previously tested or already approved. At the time of the TCGA study, there were no PARP inhibitors that were yet approved. The first approved PARP inhibitor, Olaparib, was not approved until 2014.

Third, in revealing new opportunities and lowering the cost of entry, placement of mapping information in the public domain may increase the level of competition—i.e., the number of firms with drugs treating non-inherited serous ovarian cancer with BRCA mutations. As a result, firms considering whether to invest may expect lower returns and have lowered incentives to invest.

Figure 1.1 shows the total number of privately-funded trials enrolling patients with BRCA2-mutated ovarian cancer by year, five years before and after 2010—the year in which the TCGA submitted its findings to the journal *Nature*. For simplicity, this figure excludes AstraZeneca's Olaparib which experiences relatively high levels of investment throughout the period, suggesting that AstraZeneca may have relied heavily on its own internal mapping database to guide its research efforts (see Appendix Figure A.1). I first observe that a trial was initiated in 2008—a fact that might be explained by the reasons outlined above (e.g., TCGA's mapping information may have been non-novel for some firms).

Figure 1.1: *Trials Enrolling Patients with Ovarian Cancer and BRCA2 Gene Mutations*



Notes: This figure shows the total number of clinical trials (privately-funded, phase II only) enrolling patients with BRCA2-mutated ovarian cancer in each year from 2005 to 2015. The vertical lines indicated the years in which the TCGA's ovarian cancer study (TCGA, 2011a) was submitted to (solid line) and published in (dashed line) the journal *Nature*. For simplicity, trials testing the drug Olaparib are omitted (see Appendix Figure A.1 for trials testing Olaparib).

The remainder of the figure reveals a striking relationship between the level of BRCA2-mutated ovarian cancer trials and the disclosure of the TCGA's ovarian study results: the level of trials increases in the same year in which the TCGA submitted its findings to *Nature*. Indeed, the increase is occurs in trials testing three different PARP inhibitors—all of which had been previously tested in breast cancer. Though causality has not yet been shown, this figure suggests that mapping information may positively influence the subsequent level of clinical trials, particularly among new uses of existing (previously-tested) drugs.

1.2.3 Profitability Implications of the Ovarian Cancer Mapping Study

Once firms initiate and complete clinical trials, they must decide whether to terminate or continue investment by advancing their drugs to the next clinical trial phase—a costly decision that involves significant uncertainty. In light of this, a natural follow-up question is to ask whether the TCGA's ovarian study also increases the profitability of firms' termination-or-continuation decisions. Demonstrating the impact of TCGA ovarian cancer study on the profitability of firms' decisions is difficult in this particular context due to incomplete data.[14] Therefore, my strategy in the remainder of the section is to provide a brief conceptual framework.

Suppose that results of a trial testing a drug on patients with BRCA-mutated non-inherited serous ovarian cancer reveals that the drug is ineffective. For example, the share of trial patients whose tumors shrink may be too low, or patients in the treatment group do not experience any additional gains in months of survival relative to those in the control group. Assume that the negative clinical results accurately reflect the drug's underlying value. Information from the TCGA ovarian cancer study has uncertain effects on the profitability of

[14]For example, data on trial results is required to understand whether, following the TCGA ovarian cancer study, firms are more likely to terminate trials with ambiguous trial outcomes. However, only three of the trials in Figure 1.1 have available data on trial results (see Section 1.3.3 for a discussion on trial results reporting).

the trial sponsor's termination-or-continuation decision.

On the one hand, a detailed scientific map may increase the profitability of the trial sponsor's decision by encouraging the firm to terminate investment in the drug and to save resources by minimizing further investment in a drug that is unlikely to be successful. Instead, mapping information may encourage the firm to direct resources towards drugs that are likely to successfully obtain regulatory approval in the long run (Peck *et al.*, 2015). These effect can result from two mechanisms. First, detailed basic scientific information from the TCGA's ovarian cancer study can lead to more informed-decision making (Arora and Gambardella, 1994; Bujar *et al.*, 2017; Cockburn and Henderson, 1997; Cohen and Levinthal, 1990; Cook *et al.*, 2014; Fleming and Sorenson, 2003, 2004; Morgan *et al.*, 2018; Nelson, 1982; Rosenberg, 1990; Sharpe and Keelin, 1998; Ward and Dranove, 1995). In this case, basic science can help the firm interpret the clinical trial outcomes, and clarify the costs and gains associated with the decision to terminate or continue investment.

Second, the TCGA's ovarian cancer study may improve the firm's decision-making quality. With access to a reliable, organized view of the ovarian cancer landscape, the firm may be less susceptible to biases that can lead to suboptimal outcomes. For example, mapping information may lower the likelihood that the firm computes payoffs incorrectly (e.g., due to confirmation bias, overconfidence, sunk-cost fallacy) (Bujar *et al.*, 2017; Donelan *et al.*, 2015; Tversky and Kahneman, 1974), fails to consider alternatives (Sharpe and Keelin, 1998), follows the decisions of the past or their peers (Bujar *et al.*, 2017), or overemphasizes progression-seeking behaviors (Cook *et al.*, 2014; Guedj and Scharfstein, 2004). This, in turn, may also lead to cost-saving trial terminations.

On the other hand, scientific mapping information may encourage the trial sponsor to *continue* investing in the drug, thus increasing the likelihood that firms incur the high development costs associated with late-stage failures (Peck *et al.*, 2015). Fleming and Sorenson (2004) suggests that science may motivate researchers to continue investing in a particular drug, despite negative clinical feedback. This may lead to perverse outcomes: by suggesting

that a drug-disease pairing should succeed theoretically, scientific information may encourage the firm to ignore clinical evidence that indicates otherwise.

1.3 Empirical Setting and Data

1.3.1 Scientific Background

Cancer—the disease I consider—is caused by changes in DNA.[15] A gene is a segment of DNA and a gene mutation is a type of DNA change that can modify normal cell behavior, causing excessive growth and tumor development (Stratton *et al.*, 2009). The average tumor contains 33 to 66 mutated genes; the number varies across different types of mutations (Vogelstein *et al.*, 2013). For example, the blood cancer, acute myeloid leukemia, is associated with a median number of 8 mutations. In contrast, non-small cell lung cancer is associated with 150-200 mutations per tumor. Mutations can cause a cell to produce proteins that can lead cells to grow quickly and cause damage to neighboring areas (TCGA, 2018).

I use gene-cancer pairs as my disease unit of analysis. First, I begin with a list of 80 cancer sites, based on the standard Surveillance, Epidemiology, and End Results (SEER) classification system. Next, I focus on a set of 627 genes listed in Cancer Gene Census, which are believed to be causally associated with cancer.[16] Each gene found in the Cancer Gene Census is listed along with a cancer for which there are at least two independent reports showing that mutations are found in patients with that particular cancer type and are considered to be likely implicated in driving other cancer types. This results in 50,160 gene-cancer (627 genes $\times$ 80 cancer sites) pairs possible.

[15]The underlying mechanics of genetics is much more complex. However, this is the scientific background needed for the purposes of this paper. For more details, please see `https://ghr.nlm.nih.gov/primer`.

[16]The original version of the Cancer Gene Census was first published in Futureal *et al.* (2004). The version used here comes from the Version 82 of the Catalogue of Somatic Mutations in Cancer database (For more details, see `https://cancer.sanger.ac.uk/cosmic/download`).

1.3.2 Large-Scale Cancer Genome Mapping Efforts

The purpose of cancer genome mapping is to identify the specific genes and mutations associated with different types of cancer. This is executed by comparing the DNA sequences of cancer cells to those of normal tissue (either from the same individual or a reference DNA). Appendix Figure A.2 graphically summarizes this scientific background.

In the past two decades, large-scale systematic cancer genome sequencing initiatives—efforts to catalogue and discover mutations in large numbers of tumors—have been an important source of genomic information. These large-scale efforts include *The Cancer Genome Atlas* (TCGA), the *Cancer Genome Project*, the *International Genome Consortium*, the *Pediatric Cancer Genome Project*, and cancer mapping efforts that occur in universities and other research institutions. Two key factors contributed to the rise of these initiatives (Wheeler and Wang, 2013). The first was the 2003 completion of the Human Genome Project, which sequenced the human genome and provided a reference for subsequent cancer mapping efforts. Williams (2013) finds that intellectual property restrictions that hampered subsequent use of mapping information led to a significant decrease in the level of follow-on innovation. The second factor was improvements in sequencing technology, which allowed for more accurate, faster, and cheaper sequencing. It is widely reported that the introduction of so-called next-generation sequencing allowed the cost of sequencing per genome (excluding the cost of data analysis) to fall from $95 million in 2001 to $1,000 in 2017 (Wetterstrand, 2018).[17]

I obtain the information produced through by these large-scale cancer sequencing efforts —mutation data at the gene-cancer-level—from the Catalogue of Somatic Mutations in Cancer (COSMIC) and the cBioPortal for Cancer Genomics (cBioPortal) databases. Similar to biological resource centers which act as "living libraries" for biological materials, both databases act as repositories of mapping data from hundreds of cancer mapping studies

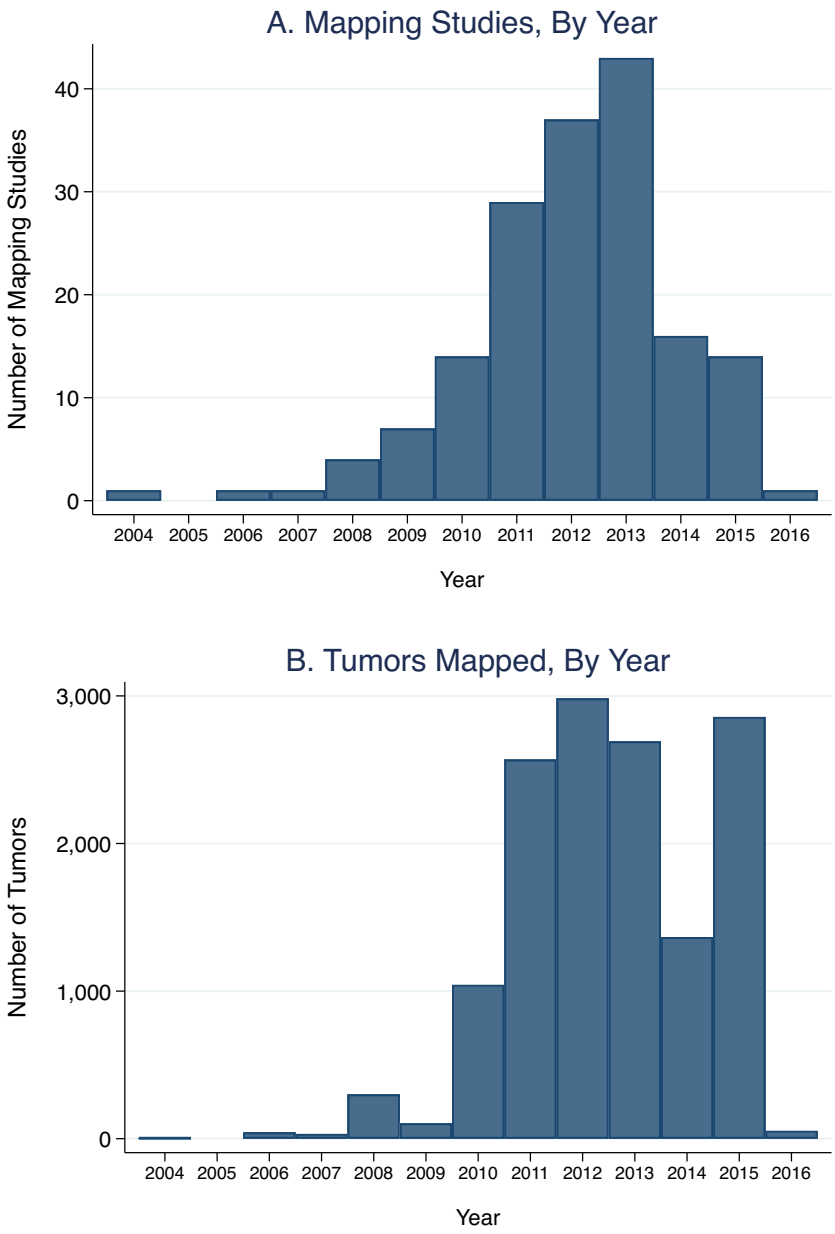
[17]Technologies have evolved from first-generation Sanger sequencing, a method that sequences a single DNA fragment at a time, to next-generation sequencing, which allows parallel mapping of millions of genes at one time.

(Furman and Stern, 2011). Further, COSMIC and cBioPortal curate and standardize cancer genome data for subsequent researchers (Yang *et al.*, 2015). Mapping data includes information about a sequenced tumor's cancer type (e.g., ovarian cancer), associated genetic mutations (gene BRCA2), and the date in which the associated mapping study was submitted to a scientific journal for publication (e.g., *Nature*-September 2010).[18]

I focus on mapping information from 168 cancer mapping efforts (see Appendix A1 for a description of how the mapping studies were selected). The cancer mapping studies used in this paper share three important characteristics. First, cancer mapping studies are cancer-site specific. For example, the TCGA ovarian cancer study described in Section 1.2 focused only on mapping ovarian cancer tumors. Second, the cancer mapping studies are large-scale and systematic. The cancer mapping studies examined in this paper typically examine hundreds of tumors. 91 percent of the mapping studies examine the entire or all of the protein-coding regions in DNA. Third, following a large literature that uses journal rankings as a proxy for publication impact, I focus on the set of large-scale mapping studies that are published in highly-ranked scientific journals. Figure 1.2 shows the number of cancer mapping studies and mapped cancer tumors between 2004 and 2016. The increase and fall likely reflects the finite number of cancer sites (e.g., the marginal value of the fifth large-scale ovarian cancer mapping study may be limited).

[18]I focus on non-silent somatic mutations, mutations that occur in the protein-coding region of the DNA and that are likely to lead to a change in biological structure. See the Appendix for more details.

Figure 1.2: *Total Cancer Mapping Studies and Mapped Tumors by Year*



Notes: These figures plot the total number of cancer mapping studies (Panel A) and mapped tumors (Panel B) in each year from 2004 to 2016. The x-axis indicates the year in which the mapping study was submitted to the journal it was ultimately published in. Mapping studies are large and published in a top 25 Genetic journal, based on rankings between 1999-2004. The increase in mapped tumors in 2015 is driven by a single study that sequenced 1144 lung cancer tumors and was submitted to the journal *Nature Genetics* in 2015 (Campbell *et al.*, 2016).

I am interested in research activity following the public disclosure that a mutation exists in a gene-cancer pair. Before describing the drug development process, I highlight two features of mutation-related information. First, I focus on the "positive" impact of mutation information on subsequent research activity—i.e., how disclosure that a mutation occurs in a gene-cancer pair may lead to an increase in private-sector research activity, relative to gene-cancer pairs that do not have mutation information. However, it is possible that cancer mapping efforts may lower subsequent research activity within a disease area. This can occur, for example, if cancer mapping reveals that a particular gene-cancer harbors a mutation that makes it more difficult to treat patients with the gene mutation and cancer. For example, a TCGA lung cancer study revealed that three percent of tumors contained a mutation that allows them to evade the immune system (TCGA, 2012). This suggest that drugs that work through activating the immune system would not be effective treatments for lung cancer patients with that specific gene mutation.

Second, information produced by large-scale cancer mapping efforts may be known before the cancer mapping study's official publication date: for instance, pharmaceutical firms may first become aware of preliminary mapping results at conferences. To approximate the earliest date that mapping information was publicly known, I identify, for each gene-cancer pair in my dataset, the first date that a mapping study containing information about a mutation in the gene-cancer is submitted to a journal.[19]

In a subset of the analysis that follow, I examine how the impact of mapping information varies across information with more (or less) clinical relevance. The scientific literature classifies mutations into two broad categories: mutations that are likely to drive the growth and progression of cancer (so-called driver mutations) and mutations that are unlikely to have a deleterious effect (so-called passenger mutations). It is not possible to definitively prove

[19]The submission date is likely to roughly approximate the time in which final results are presented at scientific conferences. For example, results from a TCGA bladder cancer mapping effort was submitted to the scientific journal *Cell* on March 23, 2017 (Robertson *et al.*, 2017). The mapping study's final results were presented at the American Society of Clinical Oncology annual meeting, a major cancer conference, on June 5, 2017 (`https://meetinglibrary.asco.org/record/153648/abstract`).

that a mutation is a driver or a passenger—instead, cancer sequencing researchers typically employ a variety of statistical methods to determine whether a given mutations is highly likely to be a driver mutation.[20] These probable driver mutations contain the strongest signal of cancer-causing behavior and are typically described in detail in the associated mapping publication.

1.3.3 Private Research Investments

Drug Development

Drug development typically begins with extensive preclinical laboratory research that involves testing a new candidate on animals and human cells. Once complete, the manufacturer begins the most expensive aspect of drug development: human testing of drugs in a series of clinical trials in which costs increase with each subsequent trial phase. Drugs that successfully demonstrate safety in phase I trials proceed to phase II trials in which their efficacy is tested in a few hundred patients. Phase III is the final stage of clinical development and involves assessing efficacy in thousands of patients and examining them over a longer period of time. Both phase II and phase III trials assess efficacy through measuring changes in overall survival and objective response rate. Once phase III is complete, manufacturers must submit a new drug application (NDA) for regulatory review. Overall, the average clinical development process is long (typically taking 8-12 years), costly (typically costing a manufacturer $800 million - $2.6 billion), and risky (only 9% of drugs that begin clinical development ultimately go to market) (CSDD, 2014; Danzon and Keuffel, 2014; DiMasi, 2001; DiMasi et al., 2003).[21]

The development and review process is indication specific—i.e., a drug receives regulatory approval for a specific therapeutic use. However, more than 60% of cancer drugs approved

[20]Common methods include: Mutation Significance (MutSig) algorithm (Lawrence *et al.*, 2014) or the Mutational Significance in Cancer (MuSiC) algorithm (Dees *et al.*, 2012).

[21]These costs estimates reflect the direct cost of research and the opportunity cost of capital. The estimates have been subject to criticism due to small sample size, assumptions about the cost of capital, and the confidential nature of the underlying data. Despite this, other efforts have generated similar cost estimates (Avorn, 2015).

have multiple uses. To expand a drug's label to include a new use, the manufacturer must undertake additional efficacy clinical trials and submit a supplemental new drug application (sNDA) (FDA, 1998b). The amount of resources involved depends on the similarity between the original and new use (FDA, 2004). For example, if manufacturer of a drug that is approved in one cancer type (e.g., gallbladder) is seeking approval in another tumor type with a common biological origin (e.g., colon), the manufacturer may skip phase I trials and rely on fewer phase II trials (FDA, 1998a). With less evidence for the FDA to review, average approval times are shorter for sNDAs for new indications and new patient populations relative to NDAs (DiMasi, 2013).

New use approvals have high expected social value (Berndt *et al.*, 2006; Roin, 2014). Francis Collins, the former director of the National Institutes of Health (NIH) describes the clinical testing of existing drugs for new uses is an opportunity to become "more efficient and effective at delivering therapies and diagnostics to patients" (Collins, 2011). Further, firms seeking new use approvals may generate scientific evidence that is useful for clinical decision making, particularly in contexts where off-label use is widespread. However, despite the relatively lower costs of seeking new use approvals, there is a widespread perception that there is too little investment in new uses of approved drugs. The so-called "problem of new uses" is caused by the limited patent protection for new uses and widespread off-label drug use (Eisenberg, 2005).

Clinical Trials Data

I collect data on privately-funded clinical trials. Data on clinical trials comes from Clarivate Analytics Cortellis Competitive Intelligence Database, which collects trials from public trial registries. Each clinical trial provides detailed information on the cancer being examined (e.g., prostate cancer), the drug being tested (e.g., Olaparib), and the sponsoring firm (AstraZeneca). The clinical trials also contain information on protein biomarkers (e.g., the gene EGFR).[22]

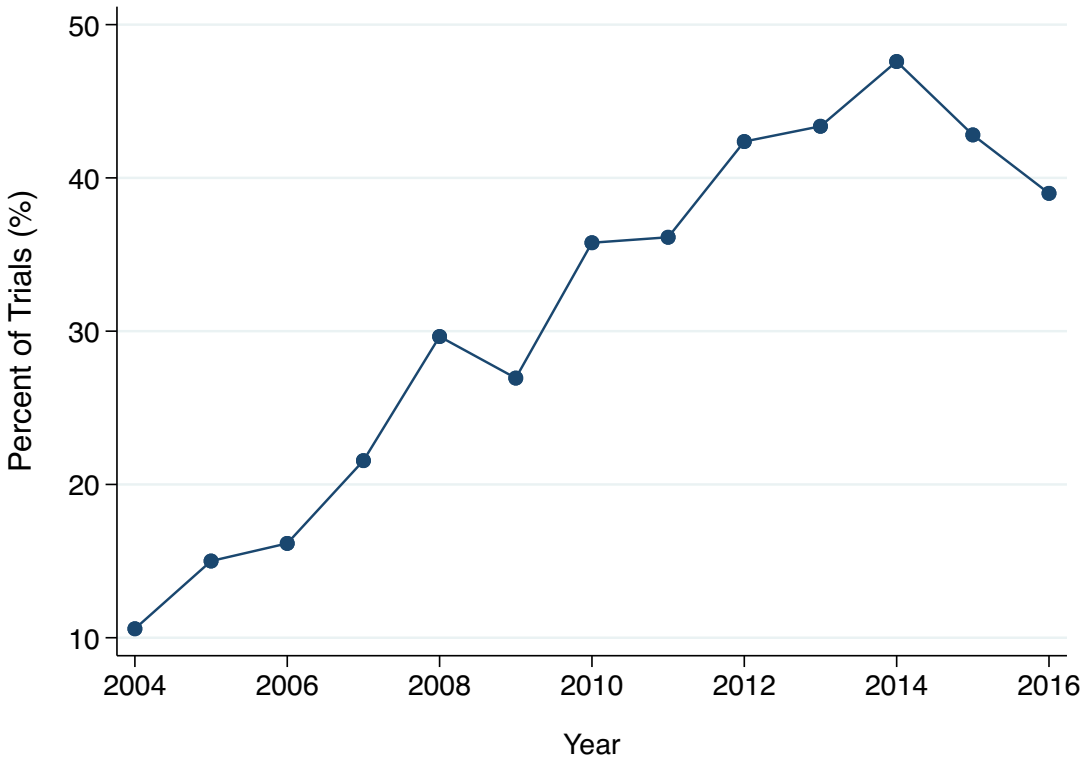
[22]I am grateful to Ariel D. Stern for sharing the cleaned data from Chandra *et al.* (2018) for this paper.

I restrict the set of clinical trials to those with biomarkers that are used to guide patient selection. Each patient biomarker can then be linked to genes using the Uniprot database to generate a dataset of trials at the gene-cancer level. Since I am interested in private-sector investments, I restrict my sample of clinical trials to those that are privately-sponsored.

To analyze the impact of mapping information on the quantity of subsequent research, I focus on investments in phase II trials—the first trials that measure efficacy and constitute a major investment for firms. This results in 30,137 privately-funded phase II clinical trials at the gene-cancer level. Figure 1.3 shows the growing share of cancer trials that are gene-related, or use gene characteristics to guide patient enrollment, over time. There is a notable increase in the share of gene-related trials before 2011, the year in which a large share of mutations was first identified in a given gene-cancer. As discussed in the ovarian cancer case study in Section 1.2, this increase may have been driven by several sources, including retrospective analyses of previous trial results or licensing relationships with genomic firms.[23] This paper aims to examine whether large-scale cancer mapping efforts lead to any additional effect on the level of privately-funded clinical trials, above and beyond these other factors.

[23]One interpretation is that the pre-2011 increase is driven by trials initiated in gene-cancer pairs that received mutation information before 2011. However, removing these trials does not change the overall trend.

Figure 1.3: *Share of Cancer Trials that Enroll Patients Based on Genes, by Year*



Notes: This figure plots the percent of cancer clinical trials (privately-funded, phase II only) that are gene-related—i.e., genes are used to enroll patients. Observations are at the trial-cancer level.

I supplement the clinical trial data in two ways:

(i) *Drug Approvals Data:* I link trial data to drug approval data to identify whether a trial is evaluating an approved drug. Data on anticancer drugs originally approved to treat cancer come from the CenterWatch, National Cancer Institute, and Memorial Sloan Kettering Cancer Center websites. This results in 187 drugs originally approved to treat cancer between 1977 and 2015, inclusive. For each drug, I obtain the date of approval and the approved cancer type.

I next classify a drug as being approved for a gene if it is approved with a companion diagnostic, a requirement for drugs aimed at targeting patients with specific genetic types.[24] For example, in 2014, the PARP-inhibitor, Olaparib was approved to treat ovarian cancer patients with BRCA1 and BRCA2 gene mutations. The drug was approved alongside the companion diagnostic BRACAnalysis CDx, a test used to detect mutations in the BRCA genes of ovarian cancer patients. I code this as being an approval in the "BRCA1-Ovarian" and "BRCA2-Ovarian" pairs in 2014.

Using the drug approvals data, I classify trials into three categories: trials testing approved, pipeline, and novel drugs. A trial-gene-cancer is classified as testing an "approved drug" if its intervention has already been approved in the same gene. For example, a trial enrolling ovarian cancer patients with BRCA2 gene mutations is classified as testing an approved drug if its intervention has been approved to treat patients with BRCA2 gene mutations prior to the start of focal trial. Similarly, a trial-gene-cancer is indicated as testing a "pipeline drug" if its intervention is not approved in the same gene but has been clinically tested previously. Finally, a trial-gene-cancer is classified as testing a "novel drug" if its intervention is not approved in the same gene and has never been clinically tested before.[25]

[24]For more details, see `https://www.fda.gov/medicaldevices/productsandmedicalprocedures/invitrodiagnostics/ucm301431.htm`

[25]Since firms are not required to report phase I trials to public trial registries, this classification scheme

(ii) *Clinical Trial Outcomes:* For a subset of the empirical exercises that follow, I examine the relationship between mapping information on common clinical trial outcomes, such as the share of patients who respond to treatment. Since the Food and Drug Administration Amendments Act (FDAAA) of 2007, most Phase II and Phase III clinical trials have been required to report results within one year of completion.[26] Despite this requirement, clinical trial results are significantly underreported (it is estimated that just 22 percent of trials meet this reporting requirement) (Anderson *et al.*, 2015; Prayle *et al.*, 2012).

To obtain data on clinical trial outcomes, I turn to abstracts submitted to the American Society of Clinical Oncology (ASCO) Annual Meeting. ASCO is the primary professional society for medical oncologists and most major research groups submit abstracts describing the findings of their clinical trials to their annual conference. Using abstracts from 2004 to 2017, I collect data on the two commonly used clinical outcomes in cancer drug development: treatment group gains in overall survival (the time between randomization and death) and objective response rates (the proportion of trial patients who experience a reduction in tumor size).

1.4 Effects on Quantity of Private Research Investments

1.4.1 Empirical Strategy

In an ideal experiment, I would estimate the impact of large-scale cancer mapping on the quantity of privately-funded trials by randomly assigning mutation information to different gene-cancer pairs. I would then compare the level of subsequently initiated clinical trials in gene-cancer pairs with mutation information, to gene-cancer pairs without mutation

may underestimate the number of trials testing pipeline drugs and overestimate the number of trials testing novel drugs.

[26]Trials covered by the FDAAA include those that have at least one site in the US and are testing a drug, device, or biological agent (FDA, 2007)

information. Motivated by the ovarian cancer case study in Section 1.2, I approximate this ideal experiment by using variation in the timing of publicly disclosed information about a mutation in a gene-cancer pair. The relative difference in clinical trials—between gene-cancer pairs with mutation information and gene-cancer pairs without—could be picking up one or both of two effects. First, the increase could represent an increase in clinical trials in gene-cancer pairs with mutation information. Second, the increase could represent a decrease in gene-cancer pairs without mutation information. I am interested in capturing both effects: the relative difference in clinical trials between gene-cancer pairs with and without mutation information.

This empirical strategy removes cancer-level differences in research potential through including gene-cancer fixed effects and estimates the impact of mapping information on clinical trials using variation in the timing of information shock—i.e., when the mutation information is disclosed—between gene-cancer pairs. By comparing gene-cancer pairs that receive an information shock early with those that receive an information shock late (or never received an information shock), I am able to estimate difference-in-difference regressions with gene-cancer, year fixed effects, and cancer-year linear trends.

1.4.2 Sample and Descriptive Statistics

I construct a balanced gene-cancer-year panel, over the period 2004-2016, inclusive. Since my analysis begins in 2004—the year in which the Cancer Gene Census (the source of the cancer genes used in this analysis) was first published—and I am interested in quantifying the effect of newly disclosed scientific information (mutation disclosures) on subsequent investment, I drop all gene-cancer pairs with known relationships as of 2004. This results in 49,542 (=50,160 - 618) gene-cancer pairs and 644,046 gene-cancer-year observations. Table 1.1 summarizes how the gene-cancer-year panel is constructed.

Table 1.1: *Overview of Gene-Cancer-Year Panel Construction*

	Count
# of genes (e.g., BRCA1, BRCA2)	627
# of cancer (e.g., ovarian, small intestine)	80
# of cancer groups (e.g., digestive)	19
# of gene - cancer (e.g., BRCA2 - prostate)	50,160
# of gene - cancer, excl. gene-cancer known in 2004	49,542
# of years (2004 to 2016)	13
# of gene - cancer - year (e.g., BRCA2 - prostate - 2004)	652,080
Final Panel: # of gene - cancer - year, excl. gene-cancer known in 2004	**644,046**

Notes: This table provides an overview of how the gene-cancer-year panel was constructed. See Appendix A.1 for more details.

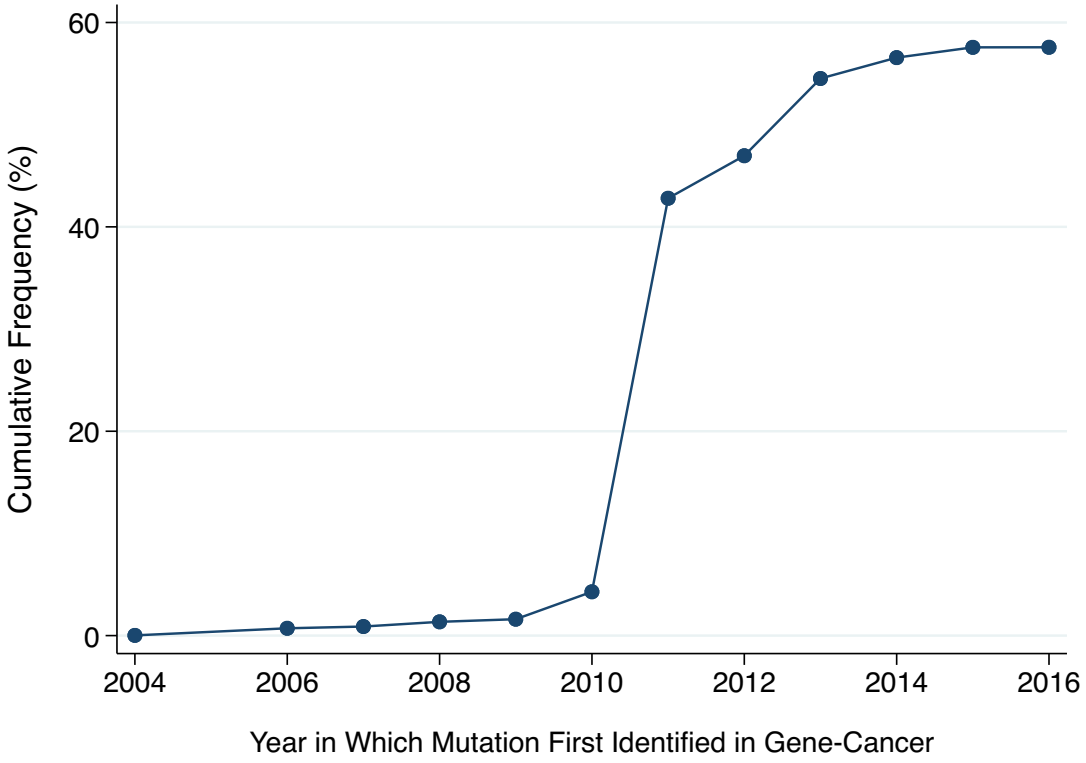
Summary statistics at the gene-cancer level are shown in Table 1.2. Panel A shows that by 2016, a mutation was identified in 58 percent of all 49,542 gene-cancer pairs. Figure 1.4 shows the cumulative distribution of the years in which mutations were first identified among the 168 mapping studies. The 2011 increase in the cumulative distribution reflects the disclosure of results from several cancer mapping studies that examined hundreds of tumors (as illustrated in Figure 1.2) and are therefore more likely to detect "rare" mutations. Consistent with these trends, Table 1.2 shows that the median year in which mutation information is first disclosed is 2011.

Table 1.2: *Summary Statistics: Gene-Cancer Level Data*

	Mean	Standard Deviation	Minimum	Maximum
A. Sequencing				
Share With Mutation (%)	57.58	49.42	0	100
Share With Mutation: Driver Mutation (%)	9.48	29.29	0	100
B. Sequencing Timing				
Year First Mutation	2011.36	1.26	2004	2016
Year First Mutation: Driver Mutation	2012.10	1.23	2008	2016
C. Outcome Variables				
Any Trial (%)	8.99	28.60	0	100
Any Trial With Approved Drug (%)	0.65	8.01	0	100
Any Trial With Pipeline Drug (%)	7.73	26.70	0	100
Any Trial With Novel Drug (%)	5.38	22.56	0	100

Notes: This table shows summary statistics at the gene-cancer level. There are 49,542 gene-cancer pairs in this sample. The period of analysis is 2004-2016. Share With Mutation: $0/1 = 1$ for gene-cancer pairs with mutations identified by cancer mapping studies. Share With Mutation: Driver Mutation: $0/1 = 1$ for gene-cancer pairs with driver mutations identified by cancer mapping studies. Cancer mapping studies are large and published in a top 25 Genetic journal, based on rankings between 1999-2004. All trials are privately-funded phase II trials. Any Trial With Approved Drug: $0/1 = 1$ for trials testing drugs that have already been approved in the same gene. Any Trial With Pipeline Drug: $0/1 = 1$ for trials testing drugs that have not been approved in the same gene, but have been tested previously. Any Trial With Novel Drug: $0/1 = 1$ for trials testing drugs that have not been approved in the same gene and have never previously been tested. See text and the appendix for more detailed data and variable descriptions.

Figure 1.4: *Cumulative Share of Gene-Cancer Pairs with Mutations Identified by Mapping Studies*



Notes: This figure plots the cumulative share of gene-cancer pairs with mutations identified by cancer mapping studies. As discussed in Section 1.3, there are 49,542 gene-cancer pairs possible. The period of analysis is 2004-2016. Cancer mapping studies are large and published in a top 25 Genetic journal, based on rankings between 1999-2004.

Table 1.2 also shows that only a minority of mutations are likely cancer-causing: Table 1.2 shows that driver mutations are identified in only 9.5 percent of gene-cancer pairs. Panel C shows that nine percent of all gene-cancers experience at least one privately-funded phase II clinical trial, between 2004-2016, inclusive. Of this nine-percent, the share of gene-cancer pairs that experience a trial testing a pipeline drug (eight percent) is higher than the share of trials testing an approved drug (less than one percent) and a novel drug (five percent).

1.4.3 Estimating Equation and Assumptions

Estimating Equation

My empirical analysis uses variation in the timing of publicly disclosed mapping information to estimate the effect of mapping information on the level of subsequent research investment within a gene-cancer pair:

$$Y_{g,c,t} = \alpha + \beta PostDisclGeneCancer_{g,c,t} + \delta_{g,c} + \tau_t + \theta_{c,t} + \epsilon_{g,c,t} \tag{1.1}$$

where $Y_{g,c,t}$ is an indicator for a clinical trial in gene g, cancer c in year t.The use of The $PostDisclGeneCancer$ variable is an indicator for whether gene-cancer gc has been publicly known to be mutated as of that year. This variable varies within gene-cancers over time, and a transition from 0 to 1 represents the fact that a mutation in a gene-cancer has been publicly disclosed. I include gene-cancer fixed effects, $\delta_{g,c}$, to control for time-invariant differences across gene-cancers, such as a gene-cancer's inherent commercial potential. Year fixed effects τ_t control for year-specific shocks that are common across gene-cancers. Finally, cancer-linear year trends (or cancer-year fixed effects) $\theta_{c,t}$ control for cancer-specific changes that are common across genes within the same cancer. I perform estimates using OLS models and cluster standard errors at the gene and cancer level.[27]

My coefficient of interest is β. β compares the average level of clinical trial investments

[27]Relative to non-linear models, such as probit or logit regressions, ordinary least square regressions generate estimates that are less prone to the incidental parameters problem (Angrist and Pischke, 2009).

in gene-cancers that received mapping information early to those that received mapping information late (or never received mapping information).

Assumptions

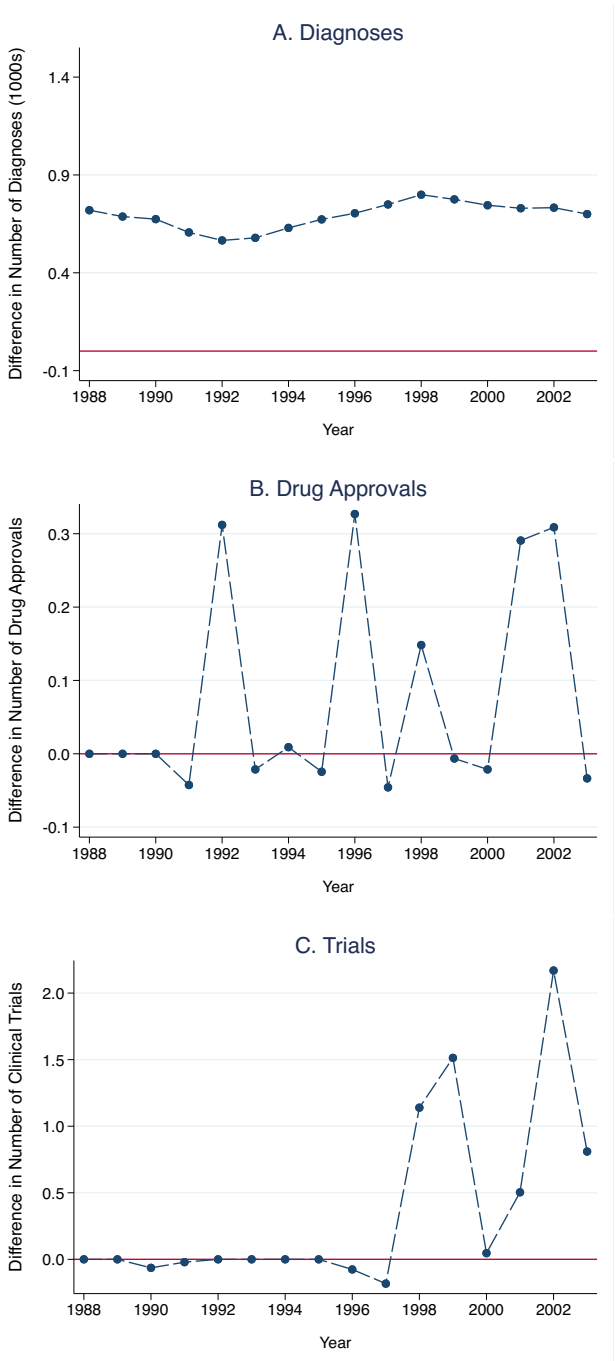
A key concern is that the research potential of gene-cancer pairs that were sequenced early are significantly different those that are sequenced late, and that those differences are changing over time. There are two types of potential selection. The first type of selection is at the cancer-level: large-scale cancer mapping studies (which are typically cancer-site specific) may be more likely to examine tumors that have higher ex ante expected research value. For example, the TCGA prioritized cancers with tumor samples that are more readily available, suggesting that the TCGA was directed towards cancers with large market sizes and the resulting estimates may be upward biased.[28,29]

I explore whether there is cancer-level selection in Figure 1.5, by comparing proxies for research potential (diagnoses, drugs approvals, trials) among cancers that were first sequenced before 2011 (the median sequencing year) and cancers that were first sequenced in/after 2011. I examine how the differences in research proxies for these two groups of cancers vary over time. While the difference in diagnoses (Panel A) remains relatively flat, the increasing difference in drug approvals (Panel B) and trials (Panel C) suggest that cancer-level selection is present. However, including cancer-year linear time trends (or cancer-year fixed effects) attenuates these concerns by controlling for cancer-level secular changes.

[28]For more details, see `https://cancergenome.nih.gov/cancersselected`

[29]A large literature documents the positive relationship between market size and pharmaceutical research. See e.g., Acemoglu and Linn (2004) and Dubois *et al.* (2015)

Figure 1.5: *Examining Cancer-Level Selection*



Notes: This figure examines baseline differences between cancers that are first sequenced early (before 2011) and cancers that are first sequenced late (in/after 2011). For each panel, difference in means of the outcome variable is calculated between the two cancer groups in each year from 1998 (the earliest year in which data for all three outcomes are available) to 2003. The outcome variables are: number of diagnoses (Panel A), number of drug approvals (Panel B), and number of phase 2, privately-funded clinical trials (Panel C).

The second type of selection is at the gene level—i.e., conditional on selecting a particular cancer, researchers may choose to sequence particular genes with higher ex ante research value. Due to the mapping technology used, this is unlikely to be a major impediment: of the 168 mapping studies used in this analysis, 91 percent employ mapping techniques that are unbiased at the gene level in the sense that they search across 100 percent of the protein-coding genes in the DNA to identify mutations.[30] The remaining nine percent of mapping studies use a strategy called targeted sequencing where select genes are targeted ex ante. While gene-level selection is a concern for these studies, the relatively low number of genes this paper focuses on (627 "at risk" cancer genes) and the large number of genes examined in the targeted sequencing studies included in this paper's analysis (3,000 genes, on average) suggest that the potential bias from gene-level selection is relatively low.

1.4.4 Results

Table 1.3 documents a positive relationship between mapping information and subsequent levels of privately funded clinical trials. The first specification in column 1 includes gene-cancer and year fixed effects, and then in subsequent columns I add cancer-year linear trends (column 2) and cancer-year fixed effects (column 3). In all cases, I estimate a strong, positive, and statistically significant effect of mapping on the relative level of subsequently initiated privately-funded clinical trials. The estimates show that information about a mutation in a gene-cancer is associated with a 0.00874-0.00915 percentage point relative increase on average in clinical trials per year. This translates into an increase in the rate of investment on the order of 50 percent of the pre-mapping information sample mean and 37 percent of the full sample mean. Appendix Figure A.2 shows that these results are robust to restricting the sample to only those genes and cancers that are unlikely to be affected by changes in

[30]The specific mapping strategies are: whole-genome sequencing and whole-exome sequencing. Whole genome sequencing reads both protein coding and non-coding regions, while whole exome sequencing focuses on protein coding regions.

intellectual property regulation that may subsequently influence researchers' and firms' efforts to identify mutations and conduct clinical trials using gene-based criteria.

Table 1.3: *Does Cancer Mapping Information Influence the Quantity of Trials?*

	Dependent Variable: 1(Any Clinical Trials)		
	(1)	(2)	(3)
1(PostDisclGeneCancer)	0.00585**	0.00874***	0.00915**
	(0.00173)	(0.00255)	(0.00302)
Mean of Dep. Var.	0.017	0.017	0.017
Percent Gain	34.41%	51.43%	53.82%
Gene-cancer FEs	X	X	X
Year FEs	X	X	X
Cancer × Linear Year Trend		X	
Cancer × Year FEs			X
Observations	644,046	644,046	644,046

Notes: This table shows the relationship between cancer mapping information and the quantity of subsequent trials. Gene-cancer-year level observations. All estimates are from OLS models. The sample includes all gene-cancer-years (excluding gene-cancer pairs known in 2004) from 2004-2016 (N = 644,046 gene-cancer-year observations). Controls include gene-cancer fixed effects, year fixed effects, and cancer-year linear time trends. Robust standard errors, clustered at the gene and cancer level, are shown in parenthesis. Cancer mapping studies are large and published in a top 25 Genetic journal, based on rankings between 1999-2004. Outcomes: 0/1 = 1 if a privately-funded phase II clinical trial is reported in a gene-cancer-year. PostDisclGeneCancer: 0/1 = 1 for the year after the mutation was disclosed. Mean of Dep. Var. is the mean number of trials in a gene-cancer before the first disclosure of a mutation. Percent Gain is calculated using the pre-mutation information trial mean.
*p<0.10, p**<0.05, ***p<0.01.

One interpretation of my findings is that if gene-cancer pairs that received mutation-related information had counterfactually experienced the same level of investments as gene-cancer pairs that did not, there would have been up to 97 fewer privately-funded clinical trials at the trial level (as opposed to trial-gene-cancer). This translates into roughly 15 fewer cancer drugs, or a 6 percent decrease between 2004 and 2016.[31]

To explore the timing of the estimated effects, I estimate:

$$Y_{g,c,t} = \alpha + \sum_{z} \beta_z \times 1(z) + \delta_{g,c} + \tau_t + \theta_{c,t} + \epsilon_{g,c,t} \tag{1.2}$$

where $\delta_{g,c}$, τ_t, and $\theta_{c,t}$ represent gene-cancer fixed effects, year fixed effects, and cancer-year linear time trends, respectively, for gene g, cancer c, and year t. z represents the "lag," or the years relative to a "zero" relative year, which marks the last year a gene-cancer was not known to be mutated (i.e., year 1 marks the first year that a mutation for a gene-cancer was disclosed).

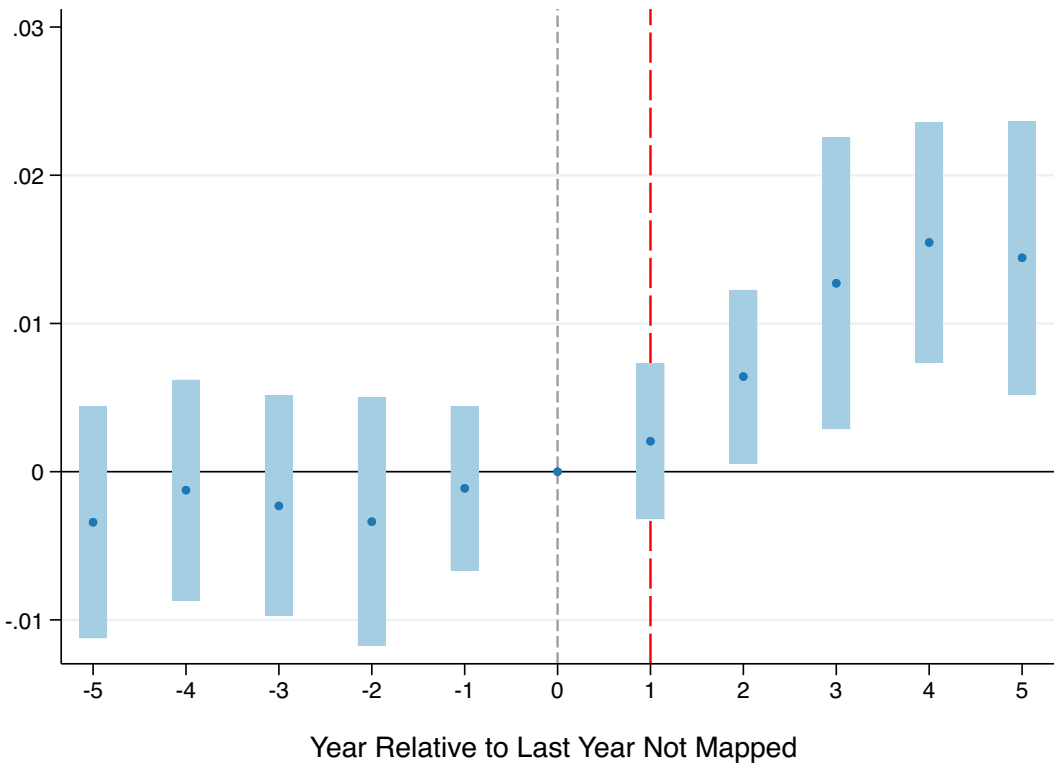
Figure 1.6 presents estimates of β_z from this regression and corresponds to a dynamic version of Table 1.3, Column 2. The light blue colored lines represent 95-percent confidence intervals and the dashed red line indicates the first year in which a mutation in a gene-cancer is publicly disclosed. The figure shows in the second year a gene-cancer is publicly disclosed (t = 2) in the graph, there is a persistent increase in the level of subsequently initiated phase

[31]I calculate these estimates using the pre-mutation information trial averages as my counterfactual. As of 2016, there are 28,524 gene-cancers that receive mutation information (or 28,524 "mapped" gene-cancers). The likelihood of obtaining experiencing a trial in any given year prior to receiving mutation information is 0.017. This suggests that if the mapped gene-cancers experienced this pre-mutation information likelihood of obtaining a trial, there would be 484.908 (28,524 × 0.017) trial-gene-cancer observations in each year. Mapping increases the likelihood of a trial by 0.0074 to 0.0244 (0.017 + 0.0074). This suggests that if the mapped gene-cancers had this likelihood of experiencing a trial, there would be 695.986 (28,524 × 0.0244) trial-gene-cancer observations in each year. This suggests that mapping leads to a 211.078 (695.986 - 484.908) yearly increase in the number of trial-gene-cancer observations. Since the majority of gene-cancers are experienced in 2011, to be conservative, I allow mapped gene-cancers to be "mapped" for 6 (2016-2011+1) years, resulting in a total of 1266.466 (6 × 211.078) trial-gene-cancers. To convert this to the trial level, I note that trials are typically associated with 13 trial-gene-cancers (trials may enroll patients with a variety of genes or cancers. For example, trials may enroll patients with BRCA1-mutated and BRCA2-mutated breast and ovarian cancer patients. This trial would appear 4 times). Converting 1266.466 trial-gene-cancer level observations to the trial level gives 97 unique trials. To obtain the estimated number of approved drugs, I take the estimated probability of successfully advancing from phase 2 to regulatory approval (15.2%) from Thomas et al. (2016), which results in an estimated 15 cancer drugs.

II clinical trials in the same gene and cancer. This delay is consistent with the view that firms may be initially testing drugs that are "on the shelf," or that have been previously tested in related diseases.

Together, these estimates suggest that information from mapping efforts within a particular disease has a positive and significant impact on the subsequent level of clinical trials in the same disease. Having shown that mapping information increases the likelihood of a privately-funded clinical by 50 percent, I now examine what types of mapping information and clinical trials drive these effects.

Figure 1.6: *Event Study Estimates—Impact of Cancer Mapping Information on Trial Quantity*



Notes: This figure plots coefficients (and 95 percent confidence intervals) from the event study specification described in Equation 1.2 and listed in Table 1.3, Column 2. On the x-axis are years z relative to a "zero" relative year that marks the last year the gene-cancer was not known to be mutated based on the cancer mapping studies (i.e., year 1 marks the first year a mutation in a gene-cancer was publicly disclosed by a cancer mapping study). As in the specifications in Table 1.3, this specification is based on gene-cancer-year level observations, the coefficients are estimates from OLS models, the sample includes all gene-cancer-years (excluding gene-cancer pairs known in 2004) from 2004-2016, and the standard errors are robust and clustered at the gene and cancer level. Gene-cancer fixed effects, year fixed effects, and cancer-year linear time trends are included. All trials are privately-funded phase II trials. Cancer mapping studies are large and published in a top 25 Genetic journal, based on rankings between 1999-2004.

Heterogeneity by Clinical Relevance of Mapping Information

The previous analysis rests on the assumption that mapping information contains useful scientific information for drug developers. In this section, I examine this assumption more closely. In particular, I ask: are firms more likely to respond to mutations that are more clinically relevant—i.e., more likely to contribute to the progression and growth of cancer?

Table 1.4 shows how the relationship between mapping information and trial quantity varies with differing levels of clinical relevance. Specifically, using Equation 1.1, I estimate how investment responds to the first appearance of a driver mutation (column 1) and the first appearance of a passenger mutation (column 2). Column 1 shows that information about a driver mutation leads to a 106 percent increase in the probability of a clinical trial. In contrast, news of a passenger mutation increases the probability of a clinical trial by 31 percent. The difference in percent gains is statistically significant. These estimates support the view that firms are more responsive to information that is more clinically relevant.

Table 1.4: *Impact on Trial Quantity: Heterogeneity by Clinical Relevance of Cancer Mapping Information*

	Dependent Variable: 1(Any Clinical Trials)	
	Driver Mutation (Strong Clinical Relevance)	Passenger Mutation (Weak Clinical Relevance)
	(1)	(2)
1(PostDisclGeneCancer)	0.0392***	0.00511**
	(0.00883)	(0.00233)
Mean of Dep. Var.	0.037	0.017
Percent Gain	106.1%	30.96%
Gene-cancer FEs	X	X
Year FEs	X	X
Cancer × Linear Year Trend	X	X
Observations	644,046	644,046

Notes: This table examines how the relationship between cancer mapping information and quantity of subsequent trials varies across mapping information with differing levels of clinical relevance. Gene-cancer-year level observations. All estimates are from OLS models. The sample includes all gene-cancer-years (excluding gene-cancer pairs known in 2004) from 2004-2016 (N = 644,046 gene-cancer-year observations). Controls include gene-cancer fixed effects, year fixed effects, and cancer-year linear time trends. Robust standard errors, clustered at the gene and cancer level, are shown in parenthesis. Cancer mapping studies are large studies and published in a top 25 Genetic journal, based on rankings between 1999-2004. Outcomes: 0/1 = 1 if a privately-funded phase II clinical trial is reported in a gene-cancer-year. PostDisclGeneCancer: 0/1 = 1 for the year after the mutation was disclosed. Mean of Dep. Var. is the mean number of trials in a gene-cancer before the first disclosure of a mutation. Percent Gain is calculated using the pre-mutation information trial mean. Column 1 shows the effect of the first driver mutation in a gene-cancer, where driver mutations are identified in two ways: 1) the mapping authors list the mutation as a likely driver mutation, or 2) the gene-cancer mutation has occurred in at least ten patients of the same cancer type. All remaining mutations are classified as passenger mutations. Column 2 shows the effect of the first passenger mutation in a gene-cancer.
*p<0.10, p**<0.05, ***p<0.01.

To further examine how the relationship between mapping information and private investment varies by information strength, I investigate whether information about one disease may affect research in a different, but closely related disease (Henderson and Cockburn, 1996; Sampat, 2012). For example, small intestine and large intestine cancer are both in the same cancer site group ("digestive system"). News that the KRAS gene is mutated in small intestine cancer may indicate that KRAS mutations are likely to occur in large intestine cancer. Appendix Table A.3 provides support for this hypothesis. Column 1 shows that clinical trial investment increases by 35 percent in response to mapping information in the same gene and a different, but closely related cancer. As expected, this effect is smaller than the direct effect of information in the same disease (50 percent). Column 2 shows that once the regression controls for mapping information in the same disease, the additional effect of mapping information in a different disease becomes statistically insignificant.

Composition of Research Investments: New Uses or New Drugs?

The increase in the quantity of privately-funded clinical trials could reflect several types of innovation. First, the increase could represent trials testing drugs approved for one disease in an additional disease ("approved drugs"). Second, as in the case of the PARP inhibitors described in Section 1.2, the increase could represent testing drugs that are not approved, but have been previously tested in another disease ("pipeline drugs"). Finally, the increase could represent trials testing drugs that have never been tested in any disease before ("novel drugs").

In theory, the relative impact of cancer mapping information on these three types of trials is ambiguous. On the one hand, a key benefit of cancer mapping is that it reveals similarities across different cancers. As a result, cancer mapping may reveal that a drug approved to treat or previously tested in one cancer, may also be effective for treating other cancers. For example, in 2013, TCGA published the results of a large-scale effort to map nearly 400 endometrial tumors. The results revealed "that the worst endometrial tumors were so similar to the most lethal ovarian and breast cancers, raising the tantalizing possibility that the three

deadly cancers might respond to the same drugs" (Kolata, 2013). This, in turn, may lead to a disproportionate increase in trials testing new uses of previously-tested (approved or pipeline) drugs.

However, it's possible that mapping information may not shift the level of investment in new uses of approved or pipeline drugs at all. First, as described in the ovarian cancer case study in Section 1.2, manufacturers of approved and pipeline drugs with substantial resources may have their internal mapping effort which may have already encouraged firms to test their approved or pipeline drugs in multiple diseases. Second, manufacturers of approved drugs may decide against running an additional trial, and instead use the publicly available information to expand demand for off-label drug use.

With this motivation, I examine how large-scale cancer genome mapping efforts influences investment in trials testing new uses of previously-tested (approved or pipeline drugs) and novel drugs. It should be noted that this comparison is primarily relevant for understanding how the composition of research shifts in the short-run. Specifically, it is possible that cancer mapping spurs additional phase II clinical trials testing novel drugs, but that the effect simply takes more time to observe (relative to investment in trials testing new uses). With this caveats in mind, I estimate regressions similar to Equation 1.1. In this analysis, the dependent variable is set to one if trial is a trial testing an approved drug, pipeline drug, or novel drug.[32]

Estimates are presented in Table 1.5. Results for trials testing previously-tested drugs are shown in Column 1 (approved drugs) and Column 2 (pipeline drugs). Mapping information increases investment in products that already exist at the time the genetic information is publicly disclosed: investment in trials testing approved drugs increases by 142 percent and

[32]This analysis categorizes trials based on the novelty of the drug(s) being tested. As a result, the analysis uses the subset (96%) of privately-funded phase II trials with a listed drug intervention. 4% of privately-funded phase II trials have missing drug intervention data. Re-running the previous analysis using the subset of trials with drug intervention leads to similar results. See Appendix Figure A.3.

in trials testing pipeline drugs by 54 percent.[33] In contrast, Column 3 shows that cancer mapping does not significantly change the rate of investment in trials testing novel drugs.

The results are consistent with prior evidence on the relationship between openness and the composition of subsequent R&D: for example, Murray *et al.* (2016) find that policies which increased access to existing research shifted the composition of follow-on research towards more diverse projects. The findings in this section suggest that, at least in the short term, publicly available basic scientific data can reinforce the value of existing products and encourage firms to engage in R&D investments that make the most of the products they already have.

[33]I can examine how the effect varies across drug approval novelty, by splitting the trials examined in column 1 into those that test drugs that were approved recently (within 100 days of the clinical trial start date) and drugs that were approved non-recently (more than 100 days before the clinical trial start date). The difference between the effect of information on recently approved and non-recently drugs is not statistically significant.

Table 1.5: *What Types of Clinical Trials: New Uses or Novel Drugs?*

	Dependent Variable: 1(Any Clinical Trials)		
Clinical Trial Testing:	Previously-Tested Drugs		Novel Drugs
	Approved	Non-Approved	
	(1)	(2)	(3)
1(PostDisclGeneCancer)	0.000920*	0.00632**	0.00249
	(0.000467)	(0.00250)	(0.00164)
Mean of Dep. Var.	0.0006	0.012	0.007
Percent gain	141.98%	53.86%	36.41%
Gene-cancer FEs	X	X	X
Year FEs	X	X	X
Cancer × Linear Year Trend	X	X	X
Observations	644,046	644,046	644,046

Notes: This table examines how the relationship between mapping information and quantity of subsequent trials varies across trials testing previously-tested (approved, pipeline) drugs in additional diseases and trials testing novel drugs. Gene-cancer-year level observations. All estimates are from OLS models. The sample includes all gene-cancer-years (excluding gene-cancer pairs known in 2004) from 2004-2016 (N = 644,046 gene-cancer-year observations). Controls include gene-cancer fixed effects, year fixed effects, and cancer-year linear time trends. Robust standard errors, clustered at the gene-cancer level, are shown in parenthesis. Cancer mapping studies are large studies and published in a top 25 Genetic journal, based on rankings between 1999-2004. Mean of Dep. Var. is the mean number of trials in a gene-cancer before the first disclosure of a mutation. Percent Gain is calculated using the pre-mutation information trial mean. Outcomes: 0/1 = 1 if a phase II clinical trial is reported in a gene-cancer-year and tests: an approved drug (column 1); a pipeline drug (column 2); and a novel drug (column 3). PostDisclGeneCancer: 0/1 = 1 for the year after the mutation was disclosed.
*p<0.10, p**<0.05, ***p<0.01.

Composition of Research Investments: Additional Dimensions

I explore how mapping information shifts the composition of subsequent trials across four additional dimensions.

(i) *Firm Heterogeneity:* It is expected that financially-constrained firms are less likely to make investments in basic science, such as cancer mapping studies. This suggests that financially-constrained firms' views of potential research opportunities are more likely to shift in response to publicly available cancer mapping information (Nagaraj, 2017). I investigate how the impact of mapping information varies across firms with different levels of financial constraints. As a proxy for firm resources, I identify the total number of patents owned by the firm, as of 2003. Firms with more than 100 patents are classified as "large," and the remaining are classified as "small." Based on this classification, 68% of trials are conducted by large firms. Appendix Table A.4 shows that, in terms of percentage gain, small (financially-constrained) firms exhibit a larger response than large firms: among small firms, clinical trial investment increases by 89 percent, relative to a 31 percent increase among large firms.

(ii) *Disease Type:* I next investigate how private research investment shifts across different disease types. For example, did mapping disproportionately benefit diseases with historically low levels of research investment, or diseases with smaller or larger market sizes? Appendix Table A.5 shows that, in terms of percentage gains, mapping information disproportionately increases investment in diseases with historically low levels of clinical trials, and equally benefits cancers with low and high market sizes.

(iii) *Trial Design Type:* Finally, I explore how cancer mapping shifts private investment in well-designed and non-well-designed trials, where well-designed trials are those that are designed to generate reliable, unbiased scientific evidence. For example, in a randomized controlled trial design, patients are randomly allocated to treatment and control arms. Randomization aims to reduce biases that can be introduced through

patient selection. Appendix Table A.6 confirms that mapping leads to a similar increase in both well-designed and non-well-designed trials.

1.5 Effects on Profitability of Firms' Research Decisions

1.5.1 Empirical Strategy

In the first set of results, I used a gene-cancer-year panel differences-in-difference research design to show that publicly available, large-scale cancer mapping efforts increases the likelihood that firms initiate phase II clinical trials. As discussed in the ovarian cancer case study in Section 1.2, a natural follow-up question is to ask how cancer mapping information shapes the profitability of firms' decisions. To perform this analysis, I first establish patterns in phase II trial outcomes among trials initiated in gene-cancer pairs where mutation is available (hereafter, "trials with information") and those initiated in gene-cancer pairs where genetic information is not yet available (hereafter, "trials without information"). I then consider firms are more likely to terminate phase II trials with weak or ambiguous clinical outcomes when genetic information is available. Finally, to assess whether mapping is associated with an increased likelihood that firms make choices that meet their objectives, I consider whether drugs that are chosen to advance to phase III ultimately result in better clinical outcomes.

To perform this analysis, I estimate OLS cross-sectional regressions and Cox proportional hazard models on trial-gene-cancer level data. In this analysis, I focus on phase II and phase III because, compared to phase I trials, both trial types are relatively well-reported and have standardized outcomes.[34] Further, using a trial-gene-cancer dataset (as opposed to the gene-cancer-year panel used in my previous analysis), allows me to examine the relationship between mapping information and *any given* trial's likelihood of generating a promising

[34]For example, a common phase II and phase III outcomes is objective response rate, which is commonly assessed using Response Evaluation Criteria in Solid Tumors (RECIST) criteria (For more details, see: `http://recist.eortc.org/`).

clinical outcomes or advancement rate.[35] To isolate the impact of mapping information that is most likely to impact the profitability of firm investment, I focus on the impact of mutations that are more likely to be clinically valuable (i.e., driver mutations).

1.5.2 Sample and Descriptive Statistics

To generate the trial-gene-cancer level dataset used in this analysis, I focus on phase II and phase III trials that satisfy two criteria. First, I restrict the analysis to the set of trials must be completed or terminated status.[36] Figure A.4 shows trends in phase II advancement rates over time. Panel A shows that the share of phase II trials that successfully advance to phase III is falling over time, a finding consistent with widespread reports about declining productivity in the pharmaceutical industry (Cook *et al.*, 2014; Peck *et al.*, 2015). Panel B indicates that the share of advanced phase II trials that are initiated in gene-cancer pairs with mutation information increases significantly in 2011, the year in which a large share of gene-cancers first experience mutation information.

Second, I restrict the analysis to trials that have available data on clinical trial outcomes. I use the most commonly measured clinical trial outcomes for each phase. For phase II trials, I use the objective response rate, or the share of the trial's patients whose tumors respond to treatment. For phase III, I use gains in overall survival, defined as the gains in time between randomization and death for the treatment group and widely considered to be the traditional gold standard for demonstrating clinical benefit of a drug. In total, this results in 2,354 phase

[35]Specifically, I use a trial-gene-cancer dataset to avoid any compositional effects that might arise with a gene-cancer-year panel. For example, suppose that a gene-cancer-year panel is used to examine the relationship between mapping information and the likelihood that a trial demonstrates a statistically significant improvement in overall survival (i.e., is "successful"). Suppose that gene-cancers with mapping information are associated with an increased likelihood of having a successful trial or an increased number of successful trials. This result can be picking up one or two effects: first, mapping increases the likelihood of success, holding the total number of trials constant. Alternatively, mapping increases the total number of trials, holding success constant. With the caveat that the estimates are correlations, using a trial-gene-cancer dataset allows me to examine the relationship between mapping information and trial success, holding the total number of trials constant.

[36]This refers to the trial's status as of July 14, 2017. This excludes a large share of firms that are "in-progress."

II trials and 422 phase III trials, at the trial-gene-cancer level.

Table 1.6 describes the final trial-gene-cancer level dataset. The table describes trial outcomes, phase II to phase III advancement rates, as well as trial sponsor characteristics. As a proxy for the trial sponsor's R&D experience, I take the inverse hyperbolic sine of the number of clinical trials that the firm initiated in the same cancer, prior to the start of the focal trial. Taking an inverse hyperbolic sine of a variable is similar to a natural logarithm transformation, but the inverse hyperbolic transformation is defined at 0 (Burbidge *et al.*, 1988). Table 1.6 shows that, among the trials used in this analysis, phase II trials have advancement rates of 57 percent.[37] Phase II trials with information are significantly less likely to advance to phase III and phase III trials with information are significantly more likely to demonstrate a statistically significant improvement in overall survival.[38]

[37]This is higher than the most comparable estimates in Wong et al. (2018), which estimates transition rates of 39 percent. This is likely due to selective reporting of trial results: all trial-gene-cancers in my dataset are required to have information on clinical trial results. The phase II to phase III transition rates of all phase II trials (including those without clinical trial results information) is 46 percent. Firms may be more likely to report positive clinical trial results (and therefore, trials that are more likely to advance to the next phase) to public trial registries or at ASCO. However, it is unlikely that this reporting bias is correlated with the presence of mapping information, suggesting the resulting estimates should be minimally biased.

[38]Specifically, whether the difference in the overall survival between the treatment group and the control (in the trial, or a historical control) is positive with the p-value < 0.05.

Table 1.6: *Summary Statistics: Trial-Gene-Cancer Level Data*

	Full	Trials With Info	Trials With No Info	Difference (2)-(3)
	(1)	(2)	(3)	(4)
A. Phase 2 (N = 2,354)				
Trial Outcome: Log(Response Rate)	2.58	2.63	2.58	0.05
1(Advance to Phase III)	0.57	0.21	0.60	-0.39***
1(Advance to Phase III, Within 4 Years)	0.57	0.20	0.60	-0.39***
Firm Experience (IHS(# Clinical Trials))	31.13	80.10	27.51	52.59***
B. Phase 3 (N = 422)				
Trial Outcome: 1(Overall Survival)	0.54	0.69	0.52	0.17**
Firm Experience (IHS(# Clinical Trials))	21.93	25.74	21.43	4.30

Notes: This table shows summary statistics at the trial-gene-cancer level. The sample includes trial-gene-cancers that are between from 2004-2016 (excluding gene-cancer pairs known in 2004), have available clinical outcomes data, and are completed or terminated as of July 14, 2017. Column 2 describes trials initiated in gene-cancer pairs where driver (clinically-relevant) mutation information was publicly available before the start of the trial. Column 3 describes trials initiated in gene-cancer pairs where driver (clinically-relevant) mutation information was not yet available at the start of the trial. Column 4 shows the difference in means. Cancer mapping studies are large studies and published in a top 25 Genetic journal, based on rankings between 1999-2004. Response Rate is the trial's objective response rate. Advance to Phase III: 0/1 = 1 for phase II trials that successfully advance to phase III. Advance to Phase III, Within 4 Years: 0/1 = 1 for phase II trials that successfully advance to phase III, within four years of the trial start date. FirmExperience is the inverse hyperbolic sine of the total number of clinical trials the trial sponsor has conducted in the focal cancer, one month prior to the trial start date. Overall Survival: 0/1 = 1 if the trial's treatment group demonstrates a statistically significant (p-value < 0.05) improvement in overall survival relative to the trial's control group or a historical control.
*p<0.10, p**<0.05, ***p<0.01.

1.5.3 Results

Phase II Outcomes

Before turning to analysis of firms' termination-or-continuation decisions, I first establish that firms with access to genetic information are choosing among drug investments whose clinical quality is similar to those of firms without access to genetic information. Formally, I estimate the following OLS specification:

$$Y_{i,g,s,c} = \beta PostDisclGeneCancer_{g,c} + \boldsymbol{X}_i + \epsilon_{i,g,c} \tag{1.3}$$

where $Y_{i,g,s,c}$ is the inverse hyperbolic sine objective response rate for trial i, gene g, cancer site s, and cancer c. The main coefficient of interest, $PostDisclGeneCancer_{g,c}$, is an indicator for whether information about a clinically relevant mutation is available for gene g, cancer c, at least one month prior to the start of trial i. $\boldsymbol{X}_i$ is a vector of trial characteristics including the trial sponsor's R&D experience, disease (gene and cancer) fixed effects, and trial start year linear trends.[39] Standard errors are clustered at the gene and cancer level.

Table 1.7 shows that phase II trials with information are not more likely to have higher objective response rates, relative to phase II trials without information. The results suggests that the quality of drug investments are similar across gene-cancer pairs with and without mapping information.

[39]Due to the small sample size, gene-cancer fixed effects are not included in the analysis

Table 1.7: *Impact of Cancer Mapping Information on Phase II Clinical Outcomes*

	Dependent Variable: IHS(Response Rate)	
	(1)	(2)
1(PostDisclGeneCancer)	0.343	0.350
	(0.258)	(0.261)
Firm Experience (IHS(# Clinical Trials))		-0.112
		(0.139)
Mean of Dep. Var.	2.583	2.583
Cancer FEs	X	X
Gene FEs	X	X
Linear Year Trend	X	X
Nb. Trial-Gene-Cancers	2,323	2,323
Nb. Trials	159	159
Nb. Genes	61	61
Nb. Cancers	80	80
R^2	0.666	0.672

Notes: This table shows the relationship between mapping information and phase II clinical outcomes, as measured by objective response rate. Trial-gene-cancer level observations. All estimates are from OLS regressions. The sample includes all phase II trial-gene-cancers that are between from 2004-2016 (excluding gene-cancer pairs known in 2004), have available clinical outcomes data, and are completed or terminated as of July 14, 2017. There are fewer than 2,354 observations because the estimation drops observations with a gene or cancer that just shows up once. Controls include cancer fixed effects and gene fixed effects. Robust standard errors, clustered at the gene and cancer level, are shown in parenthesis. Cancer mapping studies are large studies and published in a top 25 Genetic journal, based on rankings between 1999-2004. PostDisclGeneCancer: 0/1 = 1 for whether driver (clinically-relevant) mutation information was disclosed for the gene-cancer by the start of the clinical trial. FirmExperience is the inverse hyperbolic sine of the total number of clinical trials the trial sponsor has conducted in the focal cancer, one month prior to the trial start date.
*p<0.10, p**<0.05, ***p<0.01.

Termination Rates for Trials with Weak Outcomes

To understand the relationship between mapping information, phase II outcomes, and phase II advancement rates, I estimate Cox proportional hazard model regressions of the form:

$$h_{i,c,f}(t) = h_{c,f,0}(t) \times exp[\beta PostDisclGeneCancer_{g,c} + \lambda ResponseRate_i + \boldsymbol{X}_i] \qquad (1.4)$$

where $h_{c,f,0}(t)$ is the baseline hazard rate of trial advancement, stratified by cancer and sponsoring firm f's "high R&D experience status."[40] I consider a trial sponsor as having "high R&D experience" if its R&D experience is above the median of the firm experience distribution. $ResponseRate_i$ reflects trial i's phase II clinical outcomes and is the inverse hyperbolic sine of trial i's response rate. As before, $\boldsymbol{X}_i$ is a set of trial characteristics, including the trial sponsor's R&D experience and a trial start year linear trend. Standard errors are clustered at the gene and cancer level.

Table 1.8 presents the estimates. Before examining how the relationship between mapping information and phase II advancement rates varies by phase II outcomes, I examine the relationship between mapping information and phase II advancement rates across the full phase II trial-gene-cancer sample. Column 1 includes only a mapping information indicator and a linear year trend, and then in Column 2 and Column 3, I incrementally add baseline controls. Column 3 shows that, holding phase II clinical trial outcomes constant, phase II trials with information are 49 percent less likely to advance to phase III. As expected, phase II trials with higher response rates are more likely to successfully proceed. Appendix Table A.7 provides additional support for the positive relationship between promising phase II outcomes and phase II to phase III transition rates.

To examine how the relationship between mapping information and phase II advancement rates vary by phase II trial outcomes, I split the sample of phase II trials into those whose response rates were below or equal to median of the cancer-specific response rate distribution

[40]Testing the proportional-hazards assumption yielded non-significant results, suggesting that the proportionality assumption holds.

(Column 4) and those with response rates above the median (Column 5). Columns 4 suggests that conditional on having weak phase II clinical results, phase II trials with information are significantly less (60 percent) likely to advance to phase III. In contrast, column 5 shows that there is no statistically significant relationship between mapping information and advancement rates among phase II trials with positive clinical results. The results indicate that on average, firms with access to mapping information are more likely to terminate phase II trials with relatively weak or ambiguous trial outcomes.

Table 1.8: *Impact of Cancer Mapping Information on Phase II to Phase III Transitions*

	Dependent Variable: Advancing from Phase II to Phase III				
	Full Sample			Split Sample	
				Response ≤ Median	Response > Median
	(1)	(2)	(3)	(4)	(5)
1(PostDisclGeneCancer)	-0.696	-0.616	-0.672*	-0.913**	-0.0803
	(0.478)	(0.436)	(0.394)	(0.451)	(0.453)
Firm Experience (IHS(# Clinical Trials))		-0.257*	-0.187	-0.330	-0.00479
		(0.155)	(0.135)	(0.238)	(0.136)
Phase II Outcome (IHS(Response Rate))			0.235***	-0.0365	1.717***
			(0.0832)	(0.155)	(0.387)
Percent Change	-50.15%	-46.00%	-48.93%	-59.86%	-7.72%
Linear Year Trend	X	X	X	X	X
Nb. Observations	2,354	2,354	2,354	1,287	1,067
Nb. Trials	164	164	164	78	94
Nb. Genes	92	92	92	61	77
Nb. Cancers	80	80	80	80	74
Log Likelihood	-3520	-3500	-3436	-1559	-1258

Notes: This table shows the relationship between mapping information and phase II transition rates. Trial-gene-cancer level observations. Estimates are from Cox proportional hazard models, stratified by cancer and large firm status. The sample includes all phase II trial-gene-cancers that are between from 2004-2016 (excluding gene-cancer pairs known in 2004), have available clinical outcomes data, and are completed or terminated as of July 14, 2017. Column 4 shows estimates for phase II trials that have a response rate equal to or below the median of the cancer-specific distribution of response rates. Column 5 show estimates for phase II trials that have a response rate above the median of the cancer-specific distribution of response rates. Controls include a linear year time trend. Robust standard errors, clustered at the gene and cancer level, are shown in parenthesis. Cancer mapping studies are large studies and published in a top 25 Genetic journal, based on rankings between 1999-2004. PostDisclGeneCancer: 0/1 = 1 for whether driver (clinically-relevant) mutation information was disclosed for the gene-cancer by the start of the clinical trial. FirmExperience is the inverse hyperbolic sine of the total number of clinical trials the trial sponsor has conducted in the focal cancer, one month prior to the trial start date. Response Rate refers to the trial's objective response rate, or the share of patients who respond to treatment.
*p<0.10, p**<0.05, ***p<0.01.

Outcomes of Drugs that Experience Continued Investment

The previous set of results suggest that upon completing phase II trials, firms can observe where their drug falls in the quality distribution. Firms with access to mapping information are more likely to terminate drugs whose quality falls below a certain threshold. This section asks: do drugs that fall above the quality threshold (as of phase II) and are advanced to phase III, ultimately demonstrate clinical benefit based on the gold-standard measure of efficacy in cancer: overall survival?

Using a specification similar to that outlined in Equation 1.3, I examine whether phase III trials with information are more likely to demonstrate improvements in overall survival, relative to phase III trials without information. I focus on measuring improvements in overall survival as opposed to assessing whether the drug successfully completes the phase III trial and receives approval because the timing of the mapping initiatives (the median mapping year is 2011) and relatively long length of phase III trials (up to four years) indicate that regulatory approvals are rare my setting and data. Table 1.9 shows that, even after controlling for disease and firm characteristics, conditional on advancing to phase III, trials with mapping information are 40 percent more likely to demonstrate a statistically significant improvement in overall survival.

Together, this analysis shows that firms with mapping information are more likely to terminate phase II drug investments with weak clinical outcomes. Drugs advanced by firms with access to mapping information are more likely to demonstrate improvements in clinical outcomes (and therefore, more likely to successfully receive approval). This analysis does not establish causation and is estimated on a relatively small sample size. However, the significant correlations lends a basic level of credence to the idea that when firms have access to detailed, reliable scientific information, firms make more profitable investment decisions.

Table 1.9: *Impact of Mapping Information on Phase III Clinical Outcomes*

	Dependent Variable: 1(Increase in Overall Survival))	
	(1)	(2)
1(PostDisclGeneCancer)	0.183**	0.217**
	(0.0767)	(0.0936)
Firm Experience (IHS(# Clinical Trials))		0.0413
		(0.0572)
Mean of Dep. Var.	0.540	0.540
Percent Gain	33.87%	40.21%
Gene FEs	X	X
Cancer FEs	X	X
Year Linear Trends	X	X
Nb. Trial-Gene-Cancers	394	394
Nb. Trials	71	71
Nb. Genes	31	31
Nb. Cancers	31	31
R^2	0.410	0.417

Notes: This table shows the relationship between mapping information and phase III clinical outcomes, as measured by whether the phase III trial's treatment group demonstrates a statistically significant increase in overall survival. Trial-gene-cancer level observations. All estimates are from OLS regressions. The sample includes all phase II trial-gene-cancers that are between from 2004-2016 (excluding gene-cancer pairs known in 2004), have available clinical outcomes data, and are completed or terminated as of July 14, 2017. There are fewer than 422 observations because the estimation drops observations with a gene or cancer that just shows up once. Controls include gene fixed effects and cancer fixed effects. Robust standard errors, clustered at the gene and cancer level, are shown in parenthesis. Cancer mapping studies are large studies and published in a top 25 Genetic journal, based on rankings between 1999-2004. Outcome: 0/1 = 1 if the trial's treatment group demonstrates a statistically significant (p-value < 0.05) improvement in overall survival relative to the trial's control group or a historical control. PostDisclGeneCancer: 0/1 = 1 for whether driver (clinically-relevant) mutation information was disclosed for the gene-cancer by the start of the clinical trial. FirmExperience is the inverse hyperbolic sine of the total number of clinical trials the trial sponsor has conducted in the focal cancer, one month prior to the trial start date.
*p<0.10, p**<0.05, ***p<0.01.

1.6 Conclusion

This paper shows that large-scale publicly available scientific maps have important effects on the quantity and profitability of private-sector innovation. Data from large-scale cancer sequencing efforts and privately-funded clinical trials reveal that cancer mapping information leads to an estimated 50 percent increase in privately-funded clinical trials. I estimate that this translates into up to 97 additional trials and approximately 15 additional drugs. These results are driven by response to information about mutations most likely to propel cancer, a result consistent with the prediction that mapping information produces information through helping firms address scientific challenges, thus lowering the cost of clinical development. Further, cancer mapping significantly increases investment in previously-tested drugs, suggesting that one way in which large-scale scientific mapping efforts boost private innovation is through identifying clear paths across research opportunities that were previously believed to be distantly related, and through encouraging firms to make the most of the products they already have.

I analyze the relationship between cancer mapping information and the profitability of firms' research investments by looking at whether cancer mapping information leads firms to make choices that are more informed and consistent with their objectives: I find that firms are more likely to terminate phase II trials with weak outcomes, and to continue drug investments that ultimately more likely to demonstrate promising clinical outcomes. These results complement other research into the importance of understanding what guides productivity in research and development: for example, case studies suggest that access to detailed scientific information can increase the likelihood that any given investment is successful (Cook *et al.*, 2014; Morgan *et al.*, 2018). Guedj and Scharfstein (2004) find that agency problems play a large role in predicting pharmaceutical firms' continue-or-terminate decisions.

My analysis suggests avenues for future research. As governments consider investments

and policies to spur subsequent innovation, understanding the effects of investments in basic scientific knowledge is essential for structuring policy that encourages the efficient development of effective medical technologies. I study one response to large-scale cancer mapping efforts: firm investments in clinical trials. My findings on cancer mapping and trials testing new uses of approved drugs suggest that cancer mapping may also affect off-label drug use, a widespread practice that is poised to continue to grow in importance over the coming years.[41] Future work should focus on understanding how large-scale cancer mapping initiatives directly shape off-label drug demand among patients and health care providers and how this, in turn, affects firms' investment strategies. Further, the focus in this paper has been on cancer drug development and the paper is motivated by the fact that scientific mapping can improve technological search and innovation. However, the logic of the paper may apply to other diseases, including different types of brain disorders.[42]

Finally, large increases in R&D spending and persistent declines in research productivity have been widely-documented across the pharmaceutical industry (Cockburn, 2007; Scott Morton and Kyle, 2011). This study suggests that the public provision of basic scientific data in the form of scientific maps have the potential to boost medical research productivity. Declining research productivity, however, is ubiquitous across many industries, such as computers and agriculture (Bloom *et al.*, 2018; Jones, 2009). Future work should examine the extent to which publicly available scientific information can help firms in these industries navigate the research and development process.

---

[41]Off-label use is estimated to comprise approximately 50% of cancer treatments (Bach, 2015; Conti *et al.*, 2013; Molitor and Agha, 2012; Pfister, 2012).

[42]For example, Alzheimer's Genome Project (`https://curealz.org/the-research/areas-of-focus/alz-genome-project/`), the European Human Brain Project (`https://www.humanbrainproject.eu/en/`), and the US BRAIN Initiative (`https://www.braininitiative.nih.gov/`)

Chapter 2

Competition and Disclosure Incentives: Evidence from the Pharmaceutical Industry

2.1 Introduction

A central challenge for consumers is the need to access accurate information about the quality of new products. In the health care sector, as well as other industries such as transportation and chemicals, firm suppression and selective reporting of product quality information has important consequences for physicians and patients. For example, the *New York Times* revealed that GlaxoSmithKline had hidden clinical evidence of increased heart attack risk associated with its widely-used diabetes drug, Avandia, raising concerns about the reliability and representativeness of publicly available scientific information used to support medical decision making (Harris, 2010; Zarin and Tse, 2008). A key concern is that the selective disclosure of information about a product's quality, such as its safety and efficacy, can lead to inappropriate matches between patients and drugs.

This paper examines pharmaceutical firms' incentives to disclose clinical information about drug quality. In the United States, drug manufacturers must generate evidence of drug safety

and efficacy in clinical trials as a prerequisite for market entry. While firm must disclose trial results to the regulator, they can choose to selectively disclose trial findings to consumers, such as physicians and payers (Department of Health and Human Services, 2016b). Concerned about the selective disclosure of trial results, in 2007, Congress mandated that firms disclose their trial findings to the public trial registry and results database, ClinicalTrials.gov. In spite of mandatory results reporting regulation, just 13 percent of trials report results to ClinicalTrials.gov within the mandated time frame of 12 months, and concerns about the selective disclosure of trial results persist (Anderson *et al.*, 2015).

I focus on the relationship between competition and firms' drug quality disclosure incentives. The impact of competition on firms' disclosure decisions is uncertain. On the one hand, firms may need to educate consumers about the meaning of drug "quality," raising firm concerns that competitors may free-ride on their educational investments. In this instance, competition may dampen firms' disclosure incentives (Board, 2009).

On the other hand, firms may invest in positively influencing consumers' perceptions of drug safety and efficacy. In the pharmaceutical sector, consumers view a drug as "high quality" if its use is supported by rigorous evidence of efficacy and safety. There are several ways firms can shape perceived quality (from the consumers' point of view), such as demonstrating that a drug provides substantial therapeutic benefits in a well-designed, unbiased clinical trial. Under this scenario, investment in perceived product quality and the subsequent disclosure of quality information may be an important firm strategy for market entry. In light of this, I ask two empirical questions: first, what role does competition play in shaping firms' efforts to increase perceived quality? Second, how does competition influence firms' decisions to disclose the results of their clinical trials?

I examine clinical trials to treat cancer, the largest specialty medicine market in terms of global spending, and one that is the second leading cause of death in the United States (CDC, 2018; IQVIA, 2015). I assemble a dataset of privately-funded clinical trials, over the period 2000-2016. I observe many clinical trial characteristics, including the cancer type

under investigation, sponsoring firm, trial design, clinical outcomes, and importantly, whether the results have been publicly disclosed on ClinicalTrials.gov. For each drug that is clinically tested in a given cancer, I identify its competitors by determining the number of drugs that were previously tested in the same cancer.

To analyze the empirical connection between competition, investments in perceived product quality, and disclosure, I use a fixed effects model that controls for disease, firm, and product characteristics. To start, I examine the relationship between competition and firms' investments in perceived drug quality. Specifically, I examine investments along two dimensions: trial design and specialization. I find that firms facing more competition are more likely to invest in trials that are designed to generate reliable, unbiased results, as indicated by whether the trial is controlled, double-blind, and includes a disease-appropriate clinical outcome measure. In additional, I find that higher levels of competition are associated with increased likelihood that a trial's patient enrollment criteria is aimed at demonstrating a statistically significant improvement in clinical outcomes, as indicated by whether the trial is specialized. More precisely, a 1 percent increase in competition is associated with a 0.4 percent decrease in the number of patient disease types enrolled in the trial. Together, this suggests that firms facing more competition are more likely to invest in improving perceived product quality.

Finally, I analyze how competition ultimately shape firms' disclosure decisions. I find that, even after controlling for drug quality, a 10 percent increase in disclosure is associated with up to a 24 percent increase in the likelihood that a trial result will be disclosed within 12 months of the primary completion date. Together, the results suggest that investment in perceived product quality and the subsequent disclosure of quality information is an important firm strategy for market entry.

The remainder of the paper proceeds as follows. Section 2.2 provides background information about the pharmaceutical industry and clinical trial results reporting. Section 2.3 describes the data. Section 2.4 discusses the empirical strategy used in Section 2.5. Finally,

Section 2.6 concludes and discusses next steps.

2.2 Background and Conceptual Framework

2.2.1 Results Reporting in the Pharmaceutical Industry

In the United States, the U.S. Food and Drug Administration (FDA) oversees the drug development process. Drug development consists of a series of phases: the process typically begins with extensive preclinical laboratory research that involves testing a new drug candidate on animals and human cells. Once complete, the manufacturer begins the most expensive aspect of drug development: human testing of drugs in a series of clinical trials in which costs increase with each subsequent phase. Upon successful completion of the clinical trials, the sponsor will submit a drug application containing clinical evidence of drug safety and efficacy to the FDA for final approval.

While firms seeking to obtain regulatory approval must disclose all trial results to the FDA, not all trials results are disclosed to the public: the FDA does not disclose information about a drug, prior to its approval. Once the drug is approved, the FDA releases a summary of the clinical data included in its drug application, but does not release a full report on the drug's clinical results as it considers some information to be proprietary (Kesselheim and Mello, 2007; Steinbrook, 2004). Pharmaceutical firms may selectively disclose results to maintain their competitive advantage, and it has long been suggested that the drug industry engages in selective disclosure of its trials results, such as the omission of negative clinical findings (Lexchin *et al.*, 2008).

Indeed, a large literature examines the existence of these reporting biases, which include publication bias (defined by the Cochrane Handbook as "the publication or non-publication of research findings, depending on the nature and direction of the results") and outcome reporting bias (defined as "the selective reporting of some outcomes but not others, depending

on the nature and direction of the results").[1] For example, Turner *et al.* (2008) compare the results of trials published in the scientific literature to those submitted to the FDA for review. The authors find that effect sizes in the published literature are 30 percent larger than the effect sizes reported to the FDA. To assess outcomes reporting bias, Mathieu *et al.* (2009) compare the outcomes of trials published in top scientific journals to those that were registered and find that 31 percent of registered trials show evidence of selective outcome reporting, such as favoring statistically significant results.

The issue of selective reporting has been explored in a variety of contexts, ranging from psychology to economics (Delong and Lang, 1992; Ioannidis *et al.*, 2017). In the health care setting, a key concern is that selective reporting from clinical research studies dampens the ability of physicians to make accurate assessments about a drug's safety and efficacy. This, in turn, can ultimately lead to wasteful spending and inappropriate patient-drug matches. Indeed, government policies aimed at increasing transparency have largely arisen due to rising "concerns about ethical and scientific issues affecting the design, conduct, and reporting of clinical trials, including the suppression and selective reporting of results based on the interests of sponsors, unacknowledged alterations of pre-specified outcome measures, 'offshoring' of human-subjects research, and failure to report relevant adverse events" (Zarin *et al.*, 2011).

This paper is therefore related to an important set of policy questions aimed increasing access to scientific knowledge, and understanding the factors that encourage firms and researchers to selectively disclose information about the quality of their products.

2.2.2 ClinicalTrials.gov and Trial Results Disclosure

Established under the Food and Drug Administration Modernization Act of 1997 (FDAMA), ClinicalTrials.gov is the primary mechanism through the government enforces the disclosure and dissemination of trial information. Developed by the FDA and the National Institutes of Health (NIH), ClinicalTrials.gov performs two key functions: trial registration and results

[1] See `https://methods.cochrane.org/bias/reporting-biases`

reporting.

Trial Registration. Policies mandating clinical trial registration are motivated by the idea that access to information can minimize reporting biases and increase access to new treatments (Steinbrook, 2004; Zarin *et al.*, 2011). Since the FDAMA, firms have been required to register all trials used to support the FDA approval of a novel drug or a new use of an already approved drug.[2] Under existing trial registration requirements, trial sponsors must prospectively report basic trial information, such as the trial's purpose, design, patient eligibility criteria, location of trial sites, and contact information for those wanting to enroll in the trial (see Appendix Figure B.1). Trial sponsors are required to submit information when the trial is first initiated, and ClinicalTrials.gov use both automated and manual methods to ensure data accuracy.[3] In addition to allowing to minimizing reporting biases, ClinicalTrials.gov provides an opportunity for potential trial participants to learn about opportunities to enroll in clinical trials.

Other policies reinforce ClinicalTrials.gov registration requirements: since 2005, the International Committee of Medical Journal Editors (ICMJE) has required prospective trial registration in ClinicalTrials.gov, or another major trial registry, as a condition of publication. Additionally, in 2006, the World Health Organization established its own trial registry which provides access to various international registries.[4]

Trial Results Reporting. While the trial registration of trials has become standard practice, concerns about the selective reporting of trial results persist. In response, the Food

[2]Specifically, the FDAMA mandated that trials sponsors had to register any trial that was used to support the application of a drug aimed a treating a serious and life-threatening disease. In 2007, the Food and Drug Amendments Act of 2007 (FDAAA) expanded trial registration requirements to include all non-phase 1 interventional studies of drugs, biologics, and devices that have at least one U.S. site, is manufactured in and exported from the United States, or is conducted under an FDA drug application. The policy applies to all types of trial sponsors, including private firms and government organizations, such as the National Institutes of Health (Clinical Trials Registration and Results Information Submission Final Rule, 2016). For a comprehensive timeline, see `https://clinicaltrials.gov/ct2/about-site/background`.

[3]Specifically, under the FDAAA, trial sponsors must registry the trial within 21 days of enrollment of the first trial participant.

[4]For a detailed timeline, see `https://clinicaltrials.gov/ct2/about-site/history`

and Drug Amendments Act of 2007 (FDAAA) mandated that trial sponsors report summary trial results to ClinicalTrials.gov. Under the FDAAA, trial sponsors must submit trial results in a simple form (see Appendix Figure B.3 for a template). Once submitted, the results are then publicly posted in a structured, tabular format (see Appendix Figure B.2). This information includes data on pre-specified primary and secondary outcome measures and statistical analyses (Tse *et al.*, 2009). Initially, the FDAAA applied only to trials testing drugs that had been previously approved, but the mandate was later expanded to include all drugs, regardless of their approval status. Trial sponsors must submit clinical trial findings no later than 12 months after the date of final data collection for the primary outcome measure (the "primary completion date").[5] Failure to submit trial results may results in the FDA imposing fines of $10,000 a day or withholding of NIH grants funds (FDAAA, 2007).

Despite the threat of a financial penalty, results reporting remains low. Prayle *et al.* (2012) examine 738 trials and find that just 22 percent of trials report results within 2 years of the primary completion date. In a subsequent study, Anderson *et al.* (2015) examine 13,000 trials and find that 38 percent of trials reported results at any time prior to September 2013. In addition to competition, reasons for the relatively low levels of trial results reporting include the low perceived penalties for non-reporting: the FDA has never enforced the $10,000 civil monetary penalty. Further, while the ICMJE mandates trial registration, results reporting is not a prerequisite for publication.

While ClinicalTrials.gov is the world's largest database of trial results data, firms may reveal with trial results through other channels, such as industry conferences and their own publicly available registries. While the focus of this paper is trials results disclosure in ClinicalTrials.gov, future work will incorporate trial results that are disclosed through these alternative channels.

[5]Delays of up to 2 additional years are permitted in some cases, such as if the trial sponsors is aiming to seek approval of a new use.

2.2.3 Disclosure Implications of Competition

The theory of firm incentives to disclose quality information is rooted in the "unravelling result" (see Dranove and Jin (2010) for an overview). Grossman *et al.* (2017) and Milgrom (1981) describe a case where monopolists facing no disclosure cost will disclose the quality of their products. Fearing that rational consumers will believe that non-disclosing firms have low quality products, all firms will disclose.

Contrary to theory, however, firms often decide to withhold information about the quality of their products. Indeed, as Jin (2005) notes, the theories' implication that firms' disclosure decisions are independent of their competitive environment is surprising and rests on several strong assumptions. For example, the models assume that consumers have perfect knowledge of the quality distribution, that quality is exogenously given, and that the distribution of quality is independent of the level of competition. However, a large theoretical and empirical literature suggests that competition spurs improvements in quality.[6] For example, Gaynor *et al.* (2013) exploit an NHS policy change and find that competition improves hospital quality among UK hospitals.

Relaxing assumptions suggest that competition has uncertain effects on quality disclosure. To illustrate, Jin (2005) describes two scenarios: under the first scenario, consumers have imperfect knowledge of the quality distribution and disclosing firms need to consider the costs of educating consumers about the value of the information disclosed. These firms may be concerned that competitors will free-ride off efforts to educate consumers and ultimately choose not to disclose their product quality information (Board, 2009). Empirical evidence suggests that firms learn from their competitors within the pharmaceutical industry: in recent work, Krieger (2017) shows that competitors learn from each other's clinical trial failures. Under this first scenario, increased levels of competition are associated with lower levels of disclosure.

[6]See, for example, Abbott (1955); Bloom *et al.* (2015); Gaynor (2011); Matsa (2011).

Under the second scenario, firms' disclosure decisions are related to a two-stage game of product differentiation (Shaked and Sutton, 1982). In the first stage, firms invest in a certain level of perceived drug quality. Consumers consider a drug to be "high quality" if its use in treating a specific disease is supported by *rigorous* evidence of *safety and efficacy*. Thus, firms seeking to positively influence consumers' perceptions of drug quality can make two types of investments, the first focused on generating *rigorous* evidence and the second aimed at generating meaningful *safety and efficacy* evidence.

The first type of quality investment is in trial design that generates unbiased clinical evidence. It is widely viewed that a trial's design plays an important role in shaping the reliability of the trial's results (Ioannidis, 2005; Prasad and Berger, 2015). For example, randomization, in which patients are assigned to treatment and control arms by chance, can minimize potential patient selection biases that can be introduced through patient selection. Similarly, the use of control arms, in which subset of patients receive a placebo or reference drug instead of the experimental treatment, allows researchers the ability to isolate the true (unbiased) effect of the treatment. For these reasons, randomized-controlled trials are widely viewed as the highest form of evidence on which to base treatment decisions.

An additional trial design feature is the clinical endpoint. A key goal of clinical research and a requirement for FDA approval is to demonstrate an improvement in patient-centered outcomes or an established surrogate for clinical benefit (Kemp and Prasad, 2017). The "universally accepted direct measure of benefit" is overall survival, which is defined as the time from randomization until death.[7] Increasingly, firms have been permitted to show improvements in established surrogates for clinical benefit.[8] These so-called "surrogate endpoints" are believed to significantly reduce the length of clinical trials and are widely used

[7]https://www.fda.gov/downloads/drugsGuidanceComplianceRegulatoyInformation/Guidance/UCM071590.pdf

[8]Indeed, in response to pharmaceutical firms' concerns about the high time and financial cost of measuring overall survival, in recent years, the FDA has changed the set of acceptable clinical trial endpoints to allow for improvements in established surrogates for clinical benefit (Fleming, 2005; Kemp and Prasad, 2017; Prasad *et al.*, 2015).

in the development and approval of cancer drugs. However, for surrogate endpoints to be reliable for a specific cancer (e.g., prostate), they must be "validated"—i.e., they must be correlated with survival outcomes, such as overall survival (Fleming and Powers, 2012).

The second type of quality investment is in specialized trials—i.e., trials whose patient population is identified in a way so as to demonstrate greater therapeutic benefits (Chandra *et al.*, 2018). Specifically, firms may invest in running targeted trials by identifying patient subgroups that are more responsive to treatment. With an effective drug and patient match, firms can more easily show that a drug leads to statistically significant improvements in outcomes.

Once firms choose a certain level of perceived drug quality, they must make their product quality disclosure decisions. If the benefits of disclosing are sufficiently high, firms will disclose. In this setting, firms will invest in perceived product quality and the subsequent disclosure of quality information as a market entry strategy. Under this scenario, firms facing more competition will increase investments in both perceived quality and disclosure. The logical extension of this—that settings with less competition have less disclosure—finds theoretical support in Jovanovic (1982), whose model suggests that firms will only disclose if quality (which may be lower in a setting with the low competition) exceeds a certain threshold.

2.3 Data and Description Statistics

I begin with all clinical trials from ClinicalTrials.gov that begin between 2000—the first calendar year in which ClinicalTrials.gov was made publicly available—and include all clinical trials that begin up until 2016. Each clinical trial provides detailed information on the disease being examined (e.g., breast cancer), the drug being tested (e.g., Olaparib), the sponsoring firm (AstraZeneca), and the trial design (randomized, controlled, blinded). Using the algorithm outlined in Anderson *et al.* (2015), I limit the sample of trials to those that are most likely to be subject to the FDAAA's results reporting requirements (see Appendix A for details).

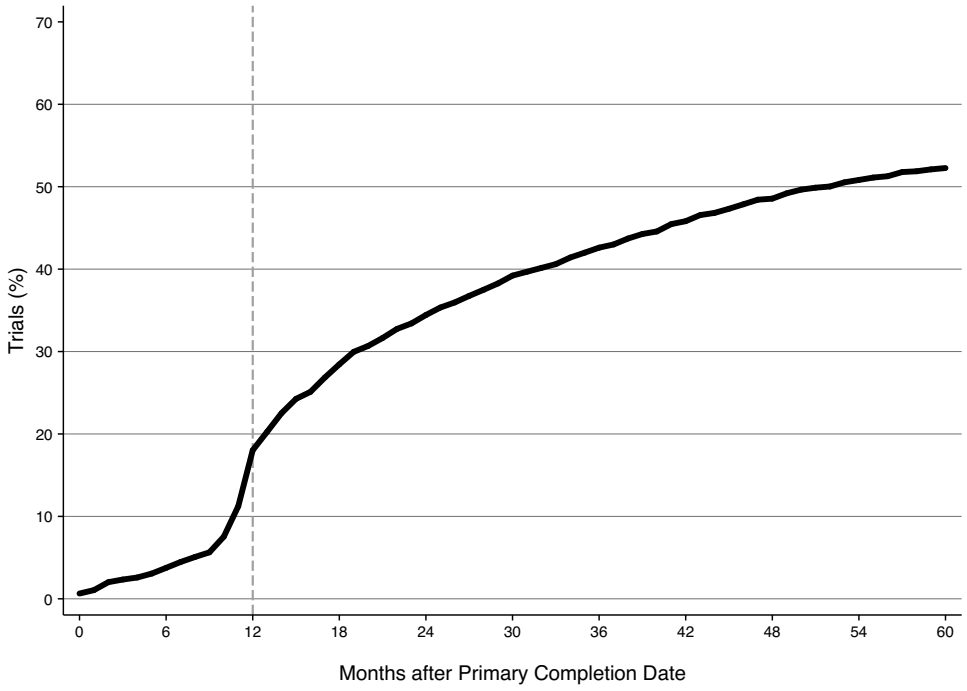
Next, I restrict the trial sample to those testing drugs to treat cancer, which comprise 25 percent of all trials. Restricting my analysis to cancer is useful for two reasons: first, the existence of relatively standardized outcome measures which provide a clear way to compare differences in clinical outcomes across diseases (Howard *et al.*, 2015). For example, the most common outcome measured in phase II trials is objective response rate (ORR), or the share of patients whose tumors respond to treatment. ORR is measured using a standardized criteria, the Response Evaluation Criteria in Solid Tumors (RECIST) criteria.[9] Second, the standardized categorization of cancers into cancer sites provides a tractable way for researchers to measure research investments. I follow Budish *et al.* (2015) in using the Surveillance, Epidemiology, and End Results (SEER) classification system to classify cancers into different cancers sites (such as breast and lung). However, it should be noted that the linkage between conditions listed in ClinicalTrials.gov and the SEER cancer classification scheme may be over generous: for example, a ClinicalTrials.gov condition of "Solid Tumor" or "Neoplasm" maps into 65 cancer sites (such as breast, lung, and prostate cancer). As a robustness check, future work will exclude trials with generalized ClinicalTrials.gov conditions.

Next, I collect data on trial results disclosure, investments in perceived quality, and competition.

Results Disclosure. For each trial, I obtain the primary completion date. I then identify the date that trial results were first submitted (if at all) to ClinicalTrials.gov. Across all trial phases, only 17 percent of trials in the sample have submitted trial results within 12 months of the study's primary completion date (the deadline mandated by the FDAAA) and 55 percent of trials have submitted by July 2018 (Appendix Table B.2). Figure 2.1 shows the cumulative percentage of clinical trials reporting results to ClinicalTrials.gov according to time from its primary completion date.

[9]See: `http://recist.eortc.org/`.

Figure 2.1: *Cumulative Percentage of Clinical Trials that Report Results*



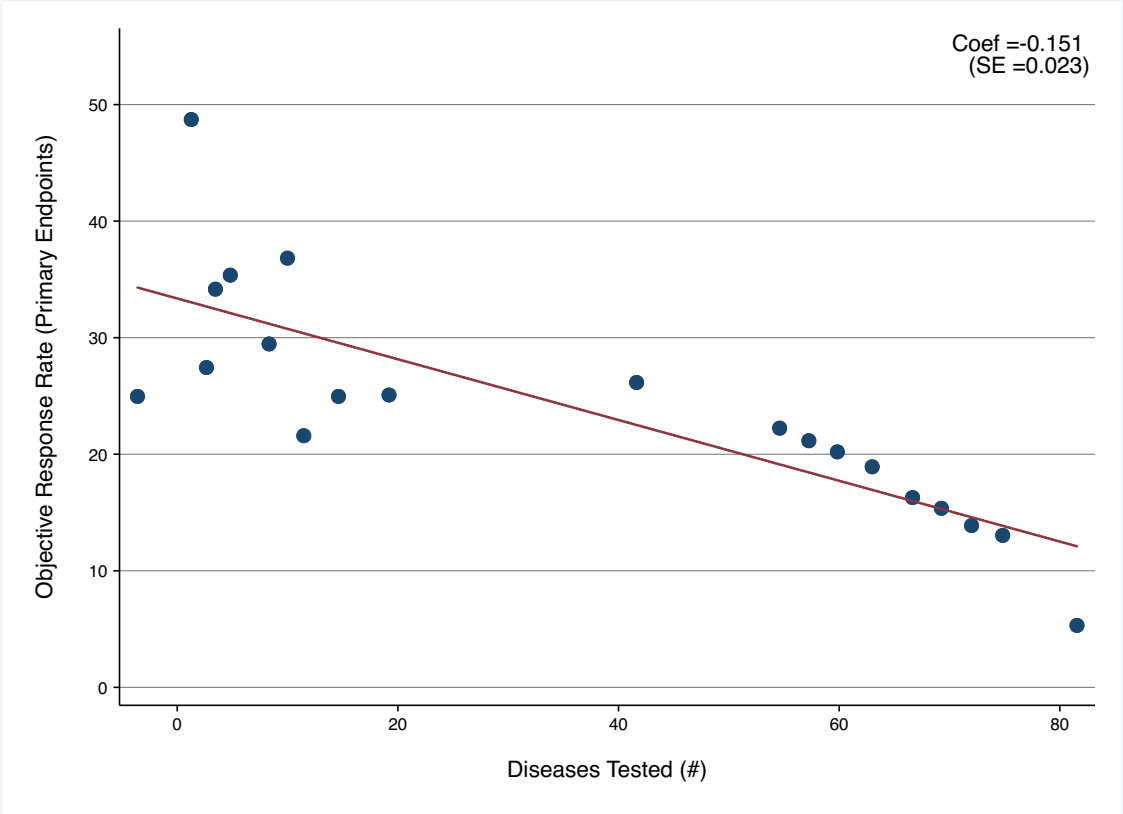
Notes: Observation at the trial-level. Phase 2 only with trial design characteristics.

Perceived Quality. To characterize firms' efforts to positive influence consumers' perception of drug quality, I collect data on a trial's design and specialization. To characterize a trial's design, I collect information on whether the trial is randomized, controlled, and blinded from ClinicalTrials.gov. Next, I use data from Prasad *et al.* (2015)—systematic review of cancer surrogate endpoints and survival—to assess whether the clinical outcome assessed in the trial is appropriate (i.e., whether the endpoint is validated).

My primary measure of trial specialization comes from determining, for each trial, the number of patient disease subtypes and the number of patients being tested. Trials can examine a large number of diseases (in this sample, the median number of conditions being tested in the sample is 39). Firms aimed at generating favorable clinical outcomes may devote the resources to identifying the patient disease subgroup most likely to demonstrate a therapeutic benefit. These specialized or targeted trials are likely to involve fewer disease subtypes and patients.

To explore the extent to which an increase in trial specialization (as measured by the decrease in the number of diseases tested) is associated with improvements in clinical outcomes, Figure 2.2 considers the relationship between the number of diseases tested and the share of trial patients whose tumors respond to treatment (i.e., the trial's ORR). Among this sample of trials, there is a negative and statistically significant relationship between the number of diseases tested in a trial and its ORR. Though this analysis is performed on the set of trials with selectively disclosed results, this provides compelling evidence that the number of patient disease types examined a trial is a good proxy for trial specialization and a good predictor of the trials' eventual clinical outcome.

Figure 2.2: *Diseases Tested and Trial Outcomes*



Notes: This figure is a bin scatter of objective response rates (y-axis) against the number of diseases tested (on the x-axis), after residualizing on cancer fixed effects. Each dot represents 139 phase II trials, on average. Trial-cancer level observations. The sample includes privately-funded phase II clinical trials from 2000-2016.

Competition. My main measure of competition comes from identifying, for each trial in a cancer, the number of unique drugs (i.e., active ingredients) previously tested in a phase II trial in the same cancer. This measure is inspired by a large empirical literature that treats drugs within the same disease as likely substitutes and their manufacturers as likely competitors (Allain, Henry, and Kyle, 2016; Gilchrist, 2016; Krieger, 2017 are some recent examples). I count the number of previously tested drugs over two time windows: since 2000 —the first calendar year that ClinicalTrials.gov was first publicly available—and five years prior to start of the focal clinical trial.

I am interested in how competition shapes firms' trial investment behavior, so I restrict my analysis to clinical trials that are industry-sponsored which make up 60 percent of all trials in my sample (Appendix Table B.2). Following a growing literature (Kao, 2019; Krieger, 2017), I focus on phase II trials since they constitute the first major investment for firms. This final sample (hereafter the "Final Sample") contains 16,210 trial-cancer observations.

Table 2.1 provides summary statistics for the key disclosure, quality, and competition variables in the Full Sample. In addition to the variables described above, the table provides information on trial sponsor experience, as measured by the cumulative number of patents associated with a trial's sponsor.

Table 2.1: *Summary Stats: Trial-Cancer Level*

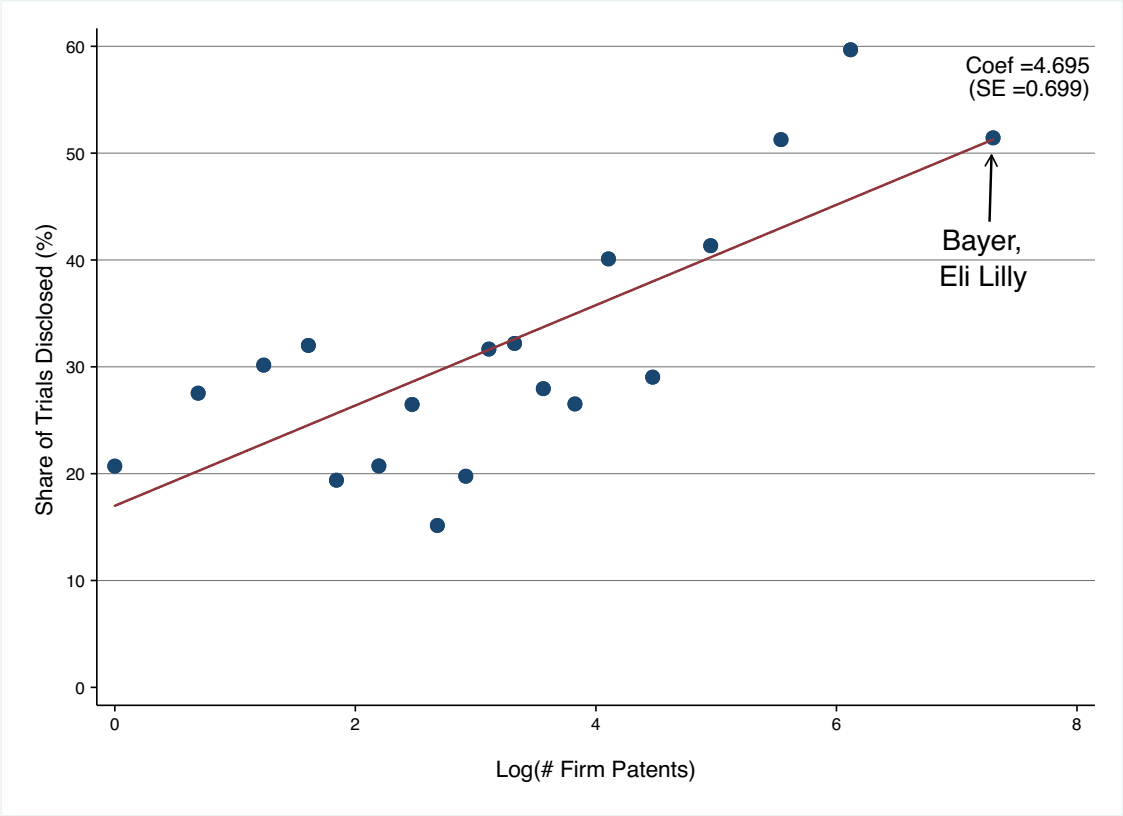
	Observations	Mean	Standard Deviation	Minimum	Maximum
Results Disclosure					
0/1: Disclosed Results Ever	16,210	0.51	0.50	0	1
0/1: Disclosed Results Within 12 Months	16,210	0.15	0.36	0	1
Clinical Results: Objective Response Rate (Phase 2 Only)					
Primary Endpoints	1,814	24.11	25.29	0	100
All Endpoints	3,528	21.78	24.19	0	100
Design					
0/1: Validated Endpoint	16,210	0.23	0.42	0	1
0/1: Randomized	16,210	0.22	0.41	0	1
0/1: Doubled Blinded	16,210	0.02	0.15	0	1
0/1: Control (Placebo)	16,210	0.05	0.21	0	1
0/1: Control (Active Comparator)	16,210	0.12	0.32	0	1
0/1: Control (Sham Comparator)	16,210	0.00	0.01	0	1
Specialization					
# Diseases Examined	16,210	38.56	29.72	1	80
# Patients Enrolled	16,210	68.73	98.11	1	1814
Competition					
# Drugs (Prior 5 Years)	16,210	271.87	95.13	58	709
# Drugs (Since 2000)	16,210	438.17	228.58	0	1,572
Other Characteristics					
# Trial Sites	15,273	11.94	19.64	1	283
0/1: Small Molecule Intervention	16,210	0.91	0.29	0	1
Patient Age (Mean)	16,053	20.86	10.51	0	90
Trial Sponsor Experience (# Patents)	16,210	1272.92	1531.59	0	8,602

Notes: This table shows summary statistics at the trial-cancer level in the Full Sample. The sample consists of phase II, industry-funded clinical trials that begin between 2000-2016. "Disclosed Results Ever" = 1 if a trial's result is disclosed by July 2018. See the text and the appendix for more detailed data and variable descriptions.

For a subset of the empirical analyses that follow, I exclude trials conducted by firms that are likely to always disclose trial results. Figure 2.3 shows that, as of 2017, firms that are large (as measured by their number of patents) are more likely to disclose a higher share of their trial results. These firms' results disclosure decisions are more likely to be motivated by factors unrelated to competition, such as fear of potential regulatory action or access to financial resources that facilitates results disclosure to ClinicalTrials.gov. For example, the large pharmaceutical firm Bayer discloses the results of more than 70 percent of its cancer-related trials on ClinicalTrials.gov and has its own clinical trial registry and results database.[10] As a result, I create a "Restricted Sample" which excludes trials conducted by large firms (those with more than 100 patents) and that disclose a large share (proxied by 90 percent) of their trial results on ClinicalTrials.gov, as of 2017.

[10]http://pharma.bayer.com/en/innovation-partnering/clinical-trials/transparency-policy/

Figure 2.3: *Firm Size and Share of Trials Disclosed*



Notes: This figure is a bin scatter of a firm's share of trials disclosed (y-axis) against the number of patents (on the x-axis), after residualizing on cancer fixed effects. Each dot represents 139 trials, on average. Trial-cancer level observations. The sample includes privately-funded phase II clinical trials from 2000-2016.

2.4 Empirical Equation and Assumptions

2.4.1 Empirical Strategy

To begin, I show a simple breakdown in the distribution of disclosure and quality variables by percentile of competition in Table 2.2. Trials are categorized as experiencing high levels of drug competition if they fall in the top 50th percentile of the year-specific drug competition distribution. Trials that fall in the bottom 50th percentile are categorized as experience low levels of drug competition. The first two columns replicate the first two columns in Table 2.1 and describe the full sample of trials. The next two columns compare trials with high and low levels of competition, as measured by the number of unique drugs tested in the same disease, five years prior to the start of the focal trial.

There are clear differences across the two trial types: in this simple comparison of means, trials experiencing high levels of competition are more likely to disclose by July 2018. In general, high competition trials are also associated with improved design and increased specialization: high competition trials are more likely to have a validated endpoint and to be randomized, controlled, and double-blind. They are also more likely to examine fewer disease types, but enroll more patients.

Table 2.2: *Differences Across High and Low Competition Trials*

	All Trials		Trials with High Competition	Trials with Low Competition	High - Low
	Observations	Mean	Mean	Mean	
Results Disclosure					
0/1: Disclosed Results Ever	16,210	0.51	0.52	0.50	0.020**
0/1: Disclosed Results Within 12 Months	16,210	0.15	0.16	0.15	0.009
Clinical Results: Objective Response Rate					
Primary Endpoints	1,814	24.11	27.55	20.67	6.874***
All Endpoints	3,528	21.78	25.68	18.05	7.630***
Design					
0/1: Validated Endpoint	16,210	0.23	0.26	0.20	0.064***
0/1: Randomized	16,210	0.22	0.26	0.18	0.085***
0/1: Doubled Blinded	16,210	0.02	0.03	0.01	0.018***
0/1: Control (Placebo)	16,210	0.05	0.06	0.03	0.031***
0/1: Control (Active Comparator)	16,210	0.12	0.15	0.09	0.058***
0/1: Control (Sham Comparator)	16,210	0.00	0.00	0.00	0.000
Specialization					
# Diseases Examined	16,210	38.56	28.86	47.50	-18.641***
# Patients Enrolled	16,210	68.73	71.34	66.32	5.017***
Other Characteristics					
# Trial Sites	15,273	11.94	13.57	10.43	3.142***
0/1: Small Molecule Intervention	16,210	0.91	0.89	0.92	-0.031***
Patient Age (Mean)	16,053	20.86	21.27	20.48	0.790***
Trial Sponsor Experience (# Patents)	16,210	1272.92	1250.02	1294.02	-43.996*
0/1: Principal Investigator Employed by Sponsor	8,236	0.26	0.25	0.27	-0.018*
0/1: No Results Restrictions	8,236	0.41	0.40	0.42	-0.022**

Notes: This table shows differences in key variables across trials with high and low levels of drug competition in the Full Sample. Trial-cancer level observations. The sample consists of phase II, industry-funded clinical trials that begin between 2000-2016. The first two columns describe all trial-cancer observations in the Full Sample. The third column describes trials with high levels of competition (as measured by the number of drugs tested in the same disease, in the past 5 years). The fourth column describes trials with low levels of drug competition. The fifth column takes the difference between trials with high and low competition. "Disclosed Results Ever" = 1 if a trial's result is disclosed by July 2018. See the text and the appendix for more detailed data and variable descriptions.
*p<0.10, p**<0.05, ***p<0.01.

I now examine the relationship between competition, perceived quality, and results disclosure more formally in the regression analysis that follows.

2.4.2 Empirical Equation

My empirical analysis uses variation in the level of competition across trials to estimate the relationship between competition on perceived drug quality and disclosure:

$$Y_i = \alpha + \beta Competition_i + \gamma' X_i + \epsilon_i, \tag{2.1}$$

where Y_i is a measure of drug quality or disclosure for trial i. $Competition_i$ is a measure of drug competition (as measured by the number of molecules being tested in the same cancer). X_i is a vector of controls, such as the drug type (measured by an indicator for whether it is a small molecule or biologic) and trial sponsor experience (as measured by the number of patents the firm owns). All specifications include cancer fixed effects, to control for differences in research opportunities and demand across diseases, and fixed effects for the year the trial begins, to control for differences in technological change and changing disclosure regulations over time. Standard errors are clustered by cancer.

2.4.3 Assumptions

One concern is that competition is endogenous. However, as Allain *et al.* (2016) note, competition is generally fixed in the short run. High barriers to entry in the pharmaceutical industry and long time frames from pre-clinical testing to phase II (which I use as my measure of competition), minimize the ability for entry. As such, the main regression results use competition as measured by the number of active ingredients previously tested in phase II trials within the same disease in the past five years.

An important exception is the testing of new uses of existing drugs, which can allow firms to move from testing an active ingredient in one disease (such as breast cancer) to another, closely related disease (such as ovarian cancer) relatively quickly. Future work will address this issue more directly.

A second concern is the presence of omitted variables that may be correlated with differences in both competition, drug quality, and disclosure decisions. For example, it's possible that a cancer-specific technological advance may trigger both an increase in competition and improvements in trial outcomes. While some concerns can be mitigated through defining competition as the number of active ingredients (as opposed to the number of trials—which might respond more), it is not possible to control for all omitted variables (Allain *et al.*, 2016). As a result, the results should not be viewed as causal.

2.5 Results

2.5.1 Impact on Quality

Table 2.3 considers the relationship between competition and trial design, with estimates from OLS models. Competition has a positive relationship with several measures of improved trial design. Column 1 shows that in the Full Sample, a 10 percent increase in drug competition is associated with a 0.009 percentage point increase in the likelihood that any given trial has a validated endpoint. Performing the regression on the Restricted Sample reveals that the relationship between competition and the propensity to have a validated endpoint is positive, though not statistically significant. Similarly, Columns 3 and 4 show that there is a positive relationship between competition and randomization, though the relationship is not statistically significant. Columns 5-8 show that trials experiencing more competition are more likely to be double-blinded and have placebo controls.

Table 2.3: *Competition and Trial Design*

	0/1: Validated Endpoint		0/1: Randomized		0/1: Double Blinded		0/1: Controlled (Placebo)		0/1: Controlled (Active Comparator)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
ln(Competition, # drugs + 1)	0.094**	0.035	0.041	0.052	0.052**	0.046**	0.055*	0.055*	-0.029	-0.018
	(0.045)	(0.045)	(0.062)	(0.062)	(0.016)	(0.018)	(0.027)	(0.029)	(0.027)	(0.029)
Mean of Dep. Var.	0.23	0.23	0.22	0.22	0.02	0.02	0.05	0.05	0.12	0.12
Controls	X	X	X	X	X	X	X	X	X	X
Full Sample	X		X		X		X		X	
Restricted Sample		X		X		X		X		X
Observations	16,210	15,027	16,210	15,027	16,210	15,027	16,210	15,027	16,210	15,027
R^2	0.09	0.11	0.06	0.06	0.02	0.02	0.03	0.04	0.04	0.04

Notes: This table examines the relationship between competition and trial design. Trial-cancer level observations. Estimates are from OLS models. Competition is measured by the number of drugs tested in the same disease, in the past 5 years. Odd-numbered columns show estimates using the Full Sample and even-numbered columns show estimates using the Restricted Sample. Controls include cancer fixed effects, year of trial start fixed effects, firm experience, and drug type. Robust standard errors, clustered at the cancer level, are shown in parenthesis.
*p<0.10, p**<0.05, ***p<0.01.

Next, I explore the relationship between competition and trial specialization in Table 2.4. Columns 1-2 show that more competition is associated with higher levels of trial specialization, as measured by the number of patient disease types examined. Specifically, a 1 percent increase in drug competition is associated with 4-5 percent fewer diseases being tested. In contrast, Columns 3-4 show that there is not statistically significant relationship between competition and the number of patients enrolled.

Recall that trials that are more specialized may be more likely to demonstrate improvements in clinical outcomes. Columns 5-8 examine the relationship between competition and clinical outcomes among trials with disclosed results. With the caveat that this sample of trials suffers from selection issues for the reasons mentioned above, I find strong evidence that an increase in competition is associated with a therapeutic benefits, as measured by ORR.

Table 2.4: *Competition and Trial Specialization*

	ln(# Diseases Examined)		ln(# Patients Enrolled)		Primary Endpoints		All Endpoints	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(Competition, # drugs + 1)	-0.444**	-0.470**	0.078	-0.029	43.364***	31.171***	35.078***	21.405***
	(0.177)	(0.183)	(0.125)	(0.130)	(9.827)	(8.248)	(7.654)	(6.080)
Mean of Dep. Var.	2.95	2.95	3.63	3.63	24.11	24.11	21.78	21.78
Controls	X	X	X	X	X	X	X	X
Full Sample	X		X					
Restricted Sample		X		X				
Disclosed, Full Sample					X		X	
Disclosed, Restricted Sample						X		X
Observations	16,210	15,027	16,210	15,027	1,814	1,583	3,528	3,084
R^2	0.42	0.42	0.14	0.15	0.33	0.35	0.27	0.29

Notes: This table examines the relationship between competition and trial specialization. Trial-cancer level observations. Estimates are from OLS models. Competition is measured by the number of drugs tested in the same disease, in the past 5 years. Columns 1 and 3 show estimates using the Full Sample and Columns 2 and 4 show estimates using the Restricted Sample. Columns 5 and 7 show estimates using the subset of Full Sample trials with disclosed trial results. Columns 6 and 8 show estimates using the subset of Restricted Sample trials with disclosed trial results. Controls include cancer fixed effects, year of trial start fixed effects, firm experience, and drug type. Robust standard errors, clustered at the cancer level, are shown in parenthesis.
*p<0.10, p**<0.05, ***p<0.01.

Measuring competition as the number of drugs tested in the same disease since 2000 reveals similar results (Appendix Tables B.3- B.4). Taken together, these estimates suggest that competition is associated with improvements along several dimensions of trial design and in trial specialization.

2.5.2 Impact on Disclosure

The final empirical section of this paper asks how competition influences firms' results disclosure decisions, conditional on drug quality. I perform this analysis by estimating equation 2.1, but this time I examine the relationship between competition and the likelihood that a trial's outcomes are disclosed. For ease of interpretation, I continue to use OLS models. Due to concerns about right-hand censoring, I examine the impact on disclosure within 12 months of trial reporting.

Table 2.5 reports the results. Using the Full Sample, the first column analyzes the relationship between competition and results disclosure, while the second column adds controls for drug quality. I find that, even controlling for drug quality, there is a positive and statistically significant relationship between competition and results disclosure in both the Full and Restricted Sample. In particular, a 10 percent increase in competition is associated with up to a 0.036 percentage point, or 24 percent increase in the likelihood of trial results disclosure within 12 months of the trial completion date. Taken together, this evidence suggests that competition is associated with higher levels of trial results disclosure.

Table 2.5: *Competition and Disclosure*

	Dependent Variable: 0/1: Disclosure Within 12 Months of Completion			
	(1)	(2)	(3)	(4)
ln(Competition, # drugs + 1)	0.366***	0.317***	0.397***	0.363***
	(0.039)	(0.035)	(0.041)	(0.037)
0/1: Validated Endpoint		0.190***		0.220***
		(0.008)		(0.009)
0/1: Randomized		0.026***		0.030***
		(0.007)		(0.007)
0/1: Double Blinded		0.131***		0.136***
		(0.026)		(0.026)
0/1: Control (Placebo)		-0.047**		-0.049**
		(0.018)		(0.017)
0/1: Control (Active Comparator)		-0.023*		-0.019
		(0.013)		(0.013)
ln(Diseases Tested (#))		-0.008***		-0.006**
		(0.002)		(0.002)
ln(# Patients Enrolled)		0.009***		0.007**
		(0.002)		(0.003)
0/1: Intervention: Small Molecule		-0.027**		-0.018*
		(0.010)		(0.009)
ln(Trial Sponsor Experience, # patents + 1)		0.152***		0.141***
		(0.006)		(0.006)
Mean of Dep. Var.	0.15	0.15	0.15	0.15
Controls	X	X	X	X
Full Sample	X	X		
Restricted Sample			X	X
Observations	16,210	15,642	15,027	14,461
R^2	0.05	0.14	0.05	0.15

Notes: This table examines the relationship between competition and trial design. Trial-cancer level observations. Estimates are from OLS models. Competition is measured by the number of drugs tested in the same disease, in the past 5 years. The first two columns show estimates using the Full Sample and the last two columns show estimates using the Restricted Sample. Controls include cancer fixed effects and year of trial start fixed effects. Robust standard errors, clustered at the cancer level, are shown in parenthesis.
*p<0.10, p**<0.05, ***p<0.01.

2.6 Conclusion

This paper provides evidence that competition is related to firms' investments in perceived quality and product quality disclosure decisions. Using the pharmaceutical industry as my setting, I show that firms facing more competition are more likely to invest in positively shaping consumers' perceptions of drug quality, as measured by investments in well-designed and specialized trials. I find that, even after controlling for drug quality, competition is associated with more disclosure of trial results. These findings are non-trivial: a 10 percent increase in competition is associated with a 24 percent increase in the likelihood of trial results disclosure within 12 months of the trial completion date. Taken together, the results suggest that investment in perceived product quality and the subsequent disclosure of quality information is an important firm strategy for market entry.

These findings relate to efforts to increase transparency in health care. Efforts to increase transparency are increasing in prevalence within the pharmaceutical industry. For example, the Physician Payments Sunshine Act was introduced to increase publicly available information about financial relationships between physicians and drug manufacturers. As governments seek to craft transparency policies to increase information exchange between firms and consumers, understanding how market mechanisms shape firm disclosure behavior is essential.

This paper has limitations and suggests avenue for future research. First, I explore one channel of trial results disclosure: clinical trial results reporting in ClinicalTrials.gov. In the pharmaceutical setting, firms will often disclose their results through other avenues, such as professional conferences. For example, the American Society of Clinical Oncology (ASCO) is the primary professional society for medical oncologists and most major research groups submit abstracts describing the findings of their clinical trials to their annual conference. In future work, I will incorporate firms' trial results disclosure at the ASCO annual meetings. In addition, I will examine the factors that motivate firms to disclose research results in some settings (e.g., publicly-available registries), but not others (e.g., professional and academic

settings).

Second, future work will broaden and refine my measures of competition. To measure competition, this paper uses the total number of drugs previously tested in a specific disease. Future work will refine my measure of competition to be a function of previous trials' clinical outcomes and design, such as the number of drugs that were previously tested in a well-designed trial and demonstrated a significant increase in overall survival. In addition, this paper focuses on the number of drugs previously tested *in the same disease.* In cancer, off-label drug use is common and widespread, comprising up to 50 percent of cancer treatments (Bach, 2015; Conti *et al.*, 2013; Molitor and Agha, 2012; Pfister, 2012). Future work will incorporate competition from off-label drug use by examining the number of drugs tested in different, but closely related diseases. Similarly, current measures of trial design and specialization can be expanded. For example, firms are increasingly running specialized trials through the use of biomarkers (Chandra *et al.*, 2018). In future work, I plan to examine how increased levels of competition shape firms' incentives to run clinical trials whose patients are selected based on gene-based criteria.

Finally, I find that competition is associated with an increase in placebo-controlled, but not active comparator-controlled trials. In most cases, it is easier to demonstrate superior clinical benefit over a placebo control, relative to an active comparator, raising questions about firms' strategic behavior with respect to investments in trial design quality. While the broad notion of trade-offs between speed and quality has been examined from the regulator perspective (Mostaghim *et al.*, 2017; Olson, 2008), less is known about how market structure influences firms' strategic investments in speed and quality. An interesting area for future research will be to explore how manufacturers in markets with competition strike a balance between pushing new products to market and investing in product quality.

Chapter 3

Medical Progress and Health Care Financing: Evidence from Academic Medical Centers[1]

3.1 Introduction

Research dating back at least to Romer (1990) has highlighted the important role that innovation plays in driving economic growth. With this motivation, scholars have acknowledged the key role that institutions play in translating knowledge into welfare-enhancing innovations (Dasgupta and David, 1994; Mokyr, 2006; Rosenberg, 1963, 1979). The question of how institutions shape the development of new products has important implications for policy aimed at promoting economic growth. Yet, we know surprisingly little about the ways in which institutions—specifically, the skills, resources, and norms they encompass—support the transformation of ideas into technological advances.[2]

One mechanism through which technological change spurs long-term growth is through

[1]Co-authored with Pierre Azoulay and Misty Heggeness

[2]Furman and Stern (2011) which examines the impact of biological resource centers on cumulative scientific discovery is an important exception.

improving health care outcomes. Cutler and McClellan (2001) find that medical technological change can lead to improvements in survival and quality of life across a broad range of diseases. The main contribution of this paper to examine how the development of novel medical advances is shaped by a specific type of institution: academic medical centers (AMC). In the United States, 30 percent of health-related research is performed inside academic medical centers (Commonwealth Fund, 1999). AMCs are institutions that consist of a medical school and an owned or closely affiliated clinical facility and have a triple mission of patient care, teaching, and research.

The AMC's complex institutional arrangement brings the "ideas sector" (i.e., biomedical research) in close contact with the "production sector" (i.e., clinical care). This setting facilitates the bidirectional flow of knowledge exchange between the laboratory bench and the patient bedside. This, in turn, can ultimately lead to development of effective treatments in at least two ways. First, AMCs play a key role in supporting research that utilizes basic science to support the development of promising new treatments—so-called "translational" research. The effective translation of basic science into novel treatments is widely viewed as being necessary for generating health improvements (Fontanarosa and DeAngelis, 2002; Woolf, 2008). Second, in numerous cases the first biological insight is acquired in a clinical setting, and only subsequently do "basic" scientists make sense of the mechanisms by which treatment is effective (Gershon, 1998). To illustrate, scientists discovered the first antidepressant drug, iproniazid, because a related compound used to treat tuberculosis made patients so euphoric that they stopped taking it. Subsequent research on iproniazid led to the chemical theories of depression that have generated all later antidepressant agents Wurtman and Bettiker (1995).

In this paper, we analyze how exogenous shifts to the "production sector" influences activity within the "ideas sector." Traditionally, financial support for research inside AMCs has come from three main sources: grants from the National Institutes of Health (NIH) and private foundations, contracts with the pharmaceutical industry, and importantly, cross-subsidies from patient care activities. Building on recent empirical work that examines

resource allocation within firms and, in particular, how multi-divisional firms respond when a business unit experiences a negative shock to its cash flows Giroud and Mueller (2017), we examine how a shock to clinical care revenues influences the rate, quality, and direction of research within AMCs.

Health care financing cuts have uncertain impacts on the level of subsequent research within AMCs. On the one hand, health care financing cuts may encourage hospitals and physicians to substitute effort towards patient care activities and away from research. In addition, low levels of cross-subsidies may make it harder for the hospital to attract talented investigators, resulting in a net decrease in subsequent research levels. On the other hand, hospitals and physicians may reduce the level of patient care activities in response to a price reduction. Instead of providing patient care, hospitals and physicians may increase time spent on research, leading to a net increase in subsequent research levels. The implications for the subsequent quality and direction of research are similarly unclear: for example, financing cuts may cause researchers to decrease both high and low quality projects, or focus their research efforts towards high-value research.

To analyze how changes in institutional funds from the "production sector" influences outcomes in the "ideas sector," we exploit quasi-experimental variation in cuts to clinical care revenues induced by the Balanced Budget Act of 1997 (BBA). The BBA was a major reform that led to considerable reductions in the level of Medicare reimbursements to hospitals. Following a period of growth due to the introduction of Medicare and Medicaid, clinical revenues slowed in the 1990s. This slowdown was partially due to the BBA and other federal efforts that were aimed at containing rising U.S. health care expenditures. Our analysis exploits the fact that the BBA decreased add-on payments made to support graduate medical education, suggesting that teaching hospitals were disproportionately affected by the reform. Among teaching hospitals, some institutions were harder hit by the reform than others because of differences in their reliance on Medicare (MedPAC, 2003).

Our empirical analysis focuses on two samples of hospitals—one that includes all teaching

hospitals and one that focuses on a research-intensive set of hospitals. We assemble a rich dataset that includes, for each hospital over the period 1992-2007, grant applications, funded grants, publications, clinical trials, and patient outcomes.

Using a difference-in-differences model that utilizes cross-hospital variation in the exposure to the reform, we find that cuts to hospital financing meaningfully reduce the subsequent quantity of research outcomes. We show that hospitals most affected by the BBA experience a nearly 4 percent decrease in subsequent grant applications and publications. These findings are consistent with previous empirical work suggesting that restrictions to the funding environment can dampen subsequent research efforts (Furman *et al.*, 2012; Tabakovic and Wollmann, 2018).

To further characterize the impact of health care financing cuts on the subsequent level of innovative activity, we examine the impact of the BBA on the quality and direction of subsequent research. We find that the BBA affects both low and high quality research equally, but leads to a "hollowing out" of basic, patient-oriented research. Specifically, in response to a health care financing shock, hospitals redirect efforts away from "translational" research, towards "bench" and "bedside" research. These findings are consistent with the view that institutions and researchers that rely most heavily on the cross-subsidies are those that bridge the gap between fundamental and patient-centered research. Taken together, these findings suggest that cuts to health care financing can impinge effectiveness of the American system of biomedical innovation.

Finally, to better understand the effects of financing shocks on consumer welfare, we explore whether the BBA led physicians to substitute away from research, towards improving patient care activities. Looking at 30-day risk-adjusted survival rates for four conditions, we do not find any association between the BBA and subsequent clinical outcomes, suggesting that the negative impact of the BBA on subsequent research was not offset by improvements in patient outcomes.

The remainder of the paper proceeds as follows. Section 3.2 provides background informa-

tion about Academic Medical Centers and financing shocks. Section 3.3 describes the data. Section 3.4 analyzes the effect of health care financing cuts on the rate, quality, and direction of subsequent research. Finally, Section 3.5 provides a discussion and concludes.

3.2 Background and Conceptual Framework

3.2.1 Academic Medical Centers and Life Sciences Research

Medical innovation arises from a complex interplay of research in the fundamental drivers of basic scientific phenomenon and investigations into how scientific insights might be applied to patient care. In the traditional model of biomedical research, research moves linearly from the "bench" to the "bedside." On one end of the bench-to-bedside (or basic-to-applied) continuum, researchers trained in the basic life sciences first discover a new molecule, and show that it inhibits a particular disease pathway in vitro.[3] Then, they develop animal models and gather initial data on safety and efficacy. On the "bedside" end of the continuum, physicians take the new molecule and test the purported treatment on a large sample of patients.

However, as has been emphasized by Rosenberg, this linear model of innovation, "however flattering to the scientist and the academic, is economically naive and simplistic in the extreme" (Rosenberg, 1994). Indeed, in the biomedical field, a closer examination of major treatment discoveries reveals a significantly more complex picture. In addition to moving from the bench to the bedside, in many cases, the flow of knowledge moves from the bedside to the bench. For example, new clinical uses for approved drugs often discovered in the clinical setting, and only subsequently do "basic" scientists make sense of the mechanisms by which these new uses are effective (DeMonaco *et al.*, 2012). As medical advances become

[3]The NIH defines "basic" research as the "systematic study directed toward greater knowledge or understanding of the fundamental aspects of phenomena and of observable facts without specific applications towards processes or products in mind." "Applied" research, in contrast, is defined as the "systematic study to gain knowledge or understanding necessary to determine the means by which a recognized and specific need may be met." (`https://nexus.od.nih.gov/all/2016/03/25/nihs-commitment-to-basic-science/`) In this paper, we use the terms "bench," "basic," and "fundamental" interchangeably. Similarly, "bedside," "applied," and "clinical" research are used interchangeably.

more complex, institutional structures that bridge the gap between medical research and medical care have become increasingly important.

AMCs bring, under the same roof, the skills and resources necessary for supporting the translation of basic science into tangible medical advances (and vice versa). Since the early 1950s, AMCs have played a central role in the American system of biomedical innovation (Ali and Gittelman, 2016; Rosenberg, 2008). The American Hospital Association (AHA) Annual Survey indicates that more than 600 U.S. hospitals are affiliated with a medical school.[4] Consisting of a hospital and closely-affiliated medical school, AMCs include basic scientists, clinicians, physician-scientists, physicians-in-training, students, and patients.

Through the physical integration of basic science and clinical research, AMCs encourage interdisciplinary collaborations not typically found in other research environments, such as private R&D labs (Ali and Gittelman, 2016). In particular, the AMC's unique organizational environment facilitates "translational research"—research that translates the "new knowledge, mechanisms, and techniques generated by advances in basic science research into new approaches for prevention, diagnosis, and treatment of disease ... essential for improving health" (Fontanarosa and DeAngelis, 2002). Examples of translational research include identifying a new protein target for the development of drugs to treat cancer or Alzheimer's.

In addition to advancing the development of "translational research," AMCs play a key role in increasing the flow of knowledge from the "bedside" to the "bench." In the case of medical devices, Gelijns and Rosenberg (1995) note that AMCs facilitate feedback from early technology users (e.g., physicians and patients) which, in turn, can shape the development and improvement of novel technologies. They write "it has been the clinicians rather than basic medical researchers who have most commonly been crucial to the invention of medical devices. They not only identify the clinical need for a new device or for improvements in existing devices but, because of their role as eventual users, they may also be the innovators

[4]This figure is based on the 2007 AHA Annual Survey. We categorize a hospital as being "affiliated with a medical school" if had a medical school affiliation reported to the American Medical Association.

and builders of the original prototype."

3.2.2 Research Funding within Academic Medical Centers

To support their research function, AMCs rely on funding from the government and the pharmaceutical industry. Government funds and industry funds typically serve different, non-substitutable roles: it is widely believed that government funds, which typically take the form of NIH grants, are used to support basic research, or research without immediate commercial value (Hellerstein, 1998). Funding from industry, in contrast, typically take the form of contracts for clinical trials: researchers within AMCs will often work with pharmaceutical and medical device firms to generate clinical evidence of product safety and efficacy to meet regulatory requirements.

This paper examines the role of a third source of research funding: revenues from clinical care activities. While these patient care cross-subsidies are not directly observable, survey and anecdotal evidence suggest that they exist and play an important role in supporting activities, ranging from research to unprofitable care (David *et al.*, 2014). A survey conduced by the Association of American Medical Colleges found that 10 percent of the faculty-practice plan revenues were used to support research (Robert and Sanderson, 1996). Interviews with AMC physicians and researchers describe the set of informal agreements guiding the use of cross-subsidies to support the institution's research function. One physician-researcher noted "in the good old days, the department of medicine's chairman would place a tax on everybody, like 20 percent on clinical care revenues, then this money could be used to support internal research."[5]

The use of clinical care revenues to support research is controversial. Proponents believe cross-subsidies may also allow investigators to continue to pursue their research agenda when faced with limited funding due to external factors (e.g., changes in government funding agencies' budgets or research priorities). Critics argue that the use of cross-subsidies is

[5]Personal interview.

wasteful—according to the physician-researcher, "the kinds of research supported [by clinical care revenues] was by and large poor quality research" and would "never would have been funded by peer review process."[6] Though of substantiative interest, the overall welfare impact of clinical care cross-subsidies for research is outside the scope of this paper. Instead, we ask: what are the empirical implications of a disruption to these patient care cross-subsidies?

3.2.3 Medicare Payments and Financing Cuts

In this paper, we examine how research activity responds to a shift in funds for research in the form of changes in clinical care revenues. We use a plausibly exogenous shock to AMC cash flow in the form of BBA-induced cuts to patient care activities funded by Medicare, which typically makes up 30 percent of patient care revenues (Reinhardt, 2006).[7]

Under a system known as the Prospective Payment System, Medicare reimburses most hospitals on a per-admission basis. In turn, each admission payment is a function of three types of adjustments: teaching subsidies, disproportionate payment subsidies, and outliers payments. The indirect medical education (IME) subsidy, is meant to compensate teaching hospitals for indirect expenses stemming for example from use of diagnostic services by clinically inexperienced residents or decreased productivity of nurses and support staff involved in the teaching of residents. The disproportionate share subsidy (DSH) corresponds to payment received by hospitals for the additional cost of treating poor patients. Finally, outlier adjustments are reimbursements to compensate providers for patients with exceptionally costly stays (Keeler *et al.*, 1998).

The BBA altered these adjustments (See Appendix A for details). As a result, Medicare inpatient payments fell about 5 percent between 1998 and 2000 and many hospitals saw their financial status deteriorate significantly (Dickler and Shaw, 2000; Iglehart, 1999; Shen and Wu, 2013). In response, the Balanced Budget Refinement Act (BBRA) of 1998 slowed

[6] Ibid [4].

[7] Other revenue sources include Medicaid and private insurers.

down the transition set by the BBA, a process continued by the Benefits Improvement and Protection Act (BIPA) of 2000. Despite these adjustments, it was widely believed that 35 percent of hospitals still experienced negative profit margins Bazzoli *et al.* (2004).

Empirical evidence on the effect of this increased financial strain on hospital activity is inconclusive. In theory, hospitals may respond to reduced Medicare payments by cost-shifting (i.e., increasing prices for privately-insured patients) or cost-cutting (i.e., lowering hospital costs, decreasing support for unprofitable services) (David *et al.*, 2014). Existing empirical work has provided evidence that hospitals cut costs by reducing payments to physicians and nurses (Bazzoli *et al.*, 2004; Cutler, 1998; Lindrooth *et al.*, 2005a) and cutting back on the quality of care (Lindrooth *et al.*, 2005b).

One dimension that is less well-understood is how cuts to hospital financing shape research activities. Reduced funding may shape the level of research activity along the intensive or extensive margin. Looking to the intensive margin, existing theory and empirical evidence suggests that the net effect of financing cuts on research levels among existing researchers largely depends on the relative strength of the income and substitution effects (Jacobson *et al.*, 2010; McGuire and Pauly, 1991; Yip, 1998). If the substitution effect dominates, physician-investigators may substitute clinical care activities for research activities (i.e., subsequent research levels will increase). If the income effect dominates, however, researchers may direct more effort towards patient care activities (i.e., subsequent research levels will decrease). In an interview, one researcher noted that declining clinical revenues caused physicians to "see more patients for less money," thus "jeopardizing ongoing research."[8] Subsequent research levels may further decrease if clinical care revenues act as complements to external sources of research funding (e.g., NIH grants). Indeed, clinical care revenues may provide young investigators time to establish their research careers while they acquire necessary grant-writing skills.

In addition, financing cuts may cause research to decline on the extensive margin: hospitals

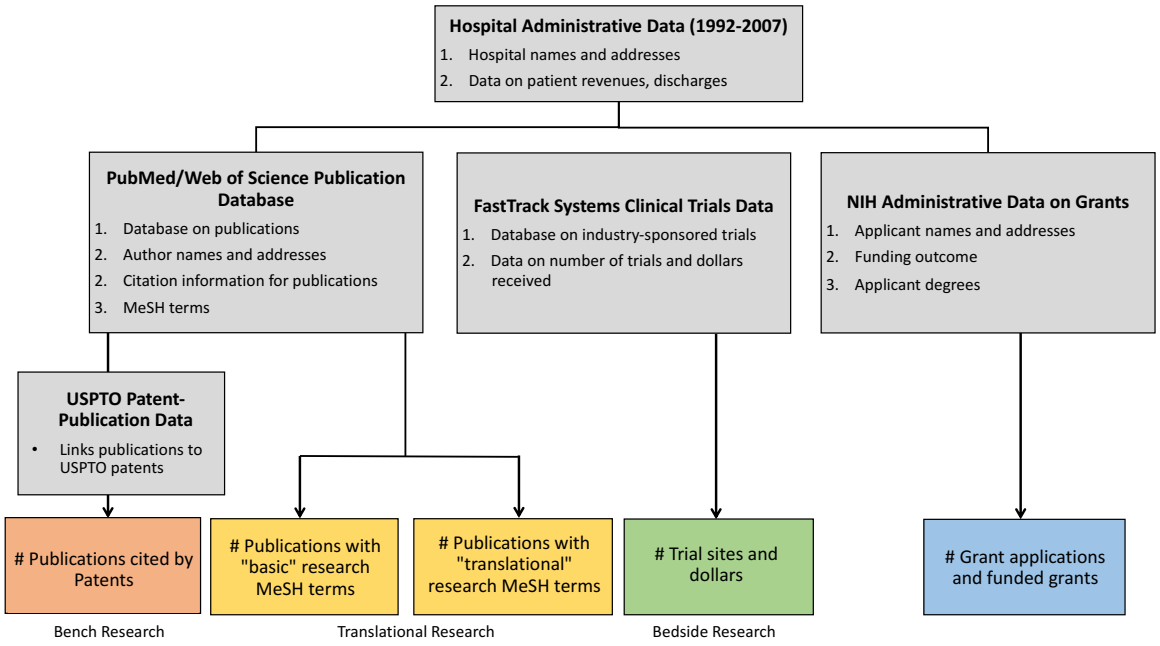
[8]Personal interview.

may decrease the rate at which they hire physician-investigators that can perform both "bench" and "bedside" research. Instead, hospitals may redirect their hiring efforts towards individuals that can be more easily obtain funding from the government (e.g., basic science researchers from the NIH) or the private sector (e.g., applied science researchers from pharmaceutical firms to run clinical trials). In response, the subsequent direction of research may away from "translational" research. Future versions of this paper will examine the extent to which the BBA shifts the rate and composition of the AMC workforce.

3.3 Data and Descriptive Statistics

Our analysis combines data from several primary sources (i) hospital characteristics from the Healthcare Cost Report Information System (HCRIS); (ii) administrative data on NIH grants from the IMPAC II database; (iii) publication data from PubMed; (iv) patent data from the USPTO; and (v) clinical trial data from FastTrack Systems, Inc. Figure 3.1 summarizes how these data sources fit together and how we construct the variables used in this analysis.

Figure 3.1: *Data Sources and Variable Construction*



We begin with data from the Healthcare Cost Report Information System (HCRIS), which is administrative data that covers the universe of Medicare-certified hospitals. We identify hospitals that have information from 1992-2007.[9] All payments are in terms of constant 1997 dollars. For each hospital record, we observe data on annual revenue from inpatient discharges (total and Medicare), number of inpatient discharges, indirect medical education payments (IME), and disproportionate share payments (DSH). Pediatric, psychiatric, cancer, and rehabilitation hospitals not paid under Medicare's Prospective Payment System are excluded from the sample.[10]

We supplement this hospital-level dataset with several measures of research activity. First, we link the hospital dataset to grants data from the National Institutes of Health (NIH) IMPAC II database.[11] For each grant, we obtain information on amounts awarded, investigators, their institutions, and a number of project characteristics. Next, we collect data on publications from PubMed, the public-access database which indexes the scientific literature. We obtain publication citation data from Thomson Reuters' Web of Science (up to 2017).

A key challenge for estimating the causal impact of the BBA is that the level of the shock (changes to Medicare funding due to the BBA) and the measures of research outcomes (NIH grants and publications) do not coincide. An analysis that examines how Medicare payments at the individual hospital-level affects NIH grants allocated at the medical-school level is likely to produce estimates that suffer from bias in a number of ways. We overcome these challenges by employing an outcome assignment mechanism that utilizes principal investigator (or author) addresses to allocate each grant (or publication) to the "correct" hospital.

[9]We require that hospitals have 15 or 16 years of data.

[10]Specifically, "specialty hospitals" include: long term care, rehabilitation, psychiatric, pediatric, cancer centers.

[11]To construct the set of relevant grants, we limit our analysis to research project awards (NIH activity code R), research career awards (NIH activity code K), program projects and centers (NIH activity codes M and P), cooperative agreements between NIH and a group of investigators (typically, NIH activity code U01), and R&D contracts to evaluate a product or device (NIH activity code N01).

As an example, the University of California, San Francisco medical center includes the Parnassus Heights Campus, Mt. Zion Hospital and Medical Center, and San Francisco General Hospital (see Appendix Figure B.1). Each of these locations has a unique Medicare provider number and therefore receive an independent Medicare payment. However, the three campuses share a single, common NIH institutional code. Our strategy consists of looking at each of the PI addresses affiliated with the UCSF NIH institutional code and allocating each PI (and grant) to one of the three hospitals.

The final analytic dataset consists of all hospitals that show evidence of teaching and research activity. To identify the set of research intensive hospitals, we make several restrictions. First, we start with a list of 1195 unique hospitals (as measured by unique Medicare provider numbers) from 1992-2007 HCRIS data.[12] Next, we exclude any hospitals that close during the 16 year period between 1992-2007 by restricting our hospital sample to those with at least 15 years of observations. Finally, we identify those hospitals that were likely to be affected by the BBA and to engage in research by restricting our sample to hospitals that receive (1) at least one indirect medical education payment and (2) produce at least one publication or submit at least one NIH grant application between 1992-2007. This results in a primary analytical sample consisting of 725 hospitals.

We also consider a second, more research and teaching-focused sample (the "AMC" sample). To create this second hospital sample, we follow the definition of a "major teaching hospital" used by (Burke *et al.*, 2017). Major teaching hospitals are members of the Council of Teaching Hospitals (COTH) and have a medical school affiliation reported to the American Medical Association.[13] Following our strategy for constructing our primary analytic hospital dataset, we restrict this second hospital sample to those with at least 15 years of observations. This results in 282 hospitals. All hospitals included in this second hospital sample except for

[12]We address mergers and acquisitions based on the following: If Hospital A and Hospital B merge at any point between 1992 and 2007, they are considered to be a single hospital.

[13]Hospital COTH status and AMA medical school affiliations are obtained from the American Hospital Association Annual Surveys from 1992-2007.

one (St. Luke's Hospitals-Meritcare) are also found in the primary analytic hospital dataset.

Further, we are interested in understanding how cuts to health care financing influence the allocation of research along the bench-to-bedside continuum. We construct two measures of "bench" research. First, we identify publications that were cited by a patent. This strategy is motivated by the idea that publications cited by patents are more likely to represent basic, or fundamental research (Trajtenberg *et al.*, 1992). Second, we follow Azoulay *et al.* (2019) in using disease MeSH terms to identify publications whose researchers employ a basic science methodology. Specifically, we identify publications using either a molecular biology technique or that use a model organism.

Next, we identify the set of publications that are linked to "translational" research, or research that links basic scientific research to novel treatments. Using the MeSH-based definitions outlined in Azoulay *et al.* (2019), we generate three measures. First, we denote a publication as representing "translational" research if it it is disease-oriented, relies on either a molecular biology technology technique or a model organism, and is not a clinical trial publication. Next, we denote a publication as "inspiring translational research" if it is translational according to the above criteria and is cited by a clinical trial publication. We identify work that "builds on translational research" as publications that report the results of a clinical trial and list a translational publication in their references.

Last, we identify "bedside" research investments by linking the hospital dataset to clinical trials come from FastTrack Systems, Inc, which gathers trial information from subscribing pharmaceutical companies in order to help them plan and negotiate investigator grants. Subscribers include nearly all large pharmaceutical companies (US and foreign), as well as most biotechnology firms.

Table 3.1 provides descriptive statistics for the primary analytic hospital sample. The table consists of five panels which respectively present statistics by hospital characteristics (panel A), next by grants (panel B) and publications (panel C), and finally industry-sponsored clinical trials (panel D). Looking first to the hospital characteristics in panel A, the first row

shows that hospital sizes (as measured by the number of patient discharges) are skewed—the annual average number of patient discharges is 18,600 while the maximum is 65,800. The second row shows that approximately a third of these visits are funded by Medicare.

Looking next to Panel B, the grant data is highly skewed: the mean number of grant applications is 9, while one hospital (Massachusetts General Hospital (MGH)) has 444. The statistics also show significant investment in new research: the majority of grant applications are for novel research proposals, as opposed to grant renewals. Splitting the grants by the degree of the principle investigator (MD, PhD, or MD/PhD) reveals that most grant applicants have PhDs.

Panel C shows that the publication data is also highly skewed: the average hospital produces 46 publications in a year, while one hospital (MGH) produces 1683. Both our patent and MeSH measures of "bench" research reveal that approximately 15 percent of publications use basic science techniques. Consistent with the view that the unique institutional features of AMCs foster the development of "translational"' research, we find that nearly 20 percent of the publications (10 publications) are associated with "translational" research methods. In addition, panel C also shows that the average AMC is associated with 3 publications that builds on "translational" research, and 4 publications that inspire "translational" research.

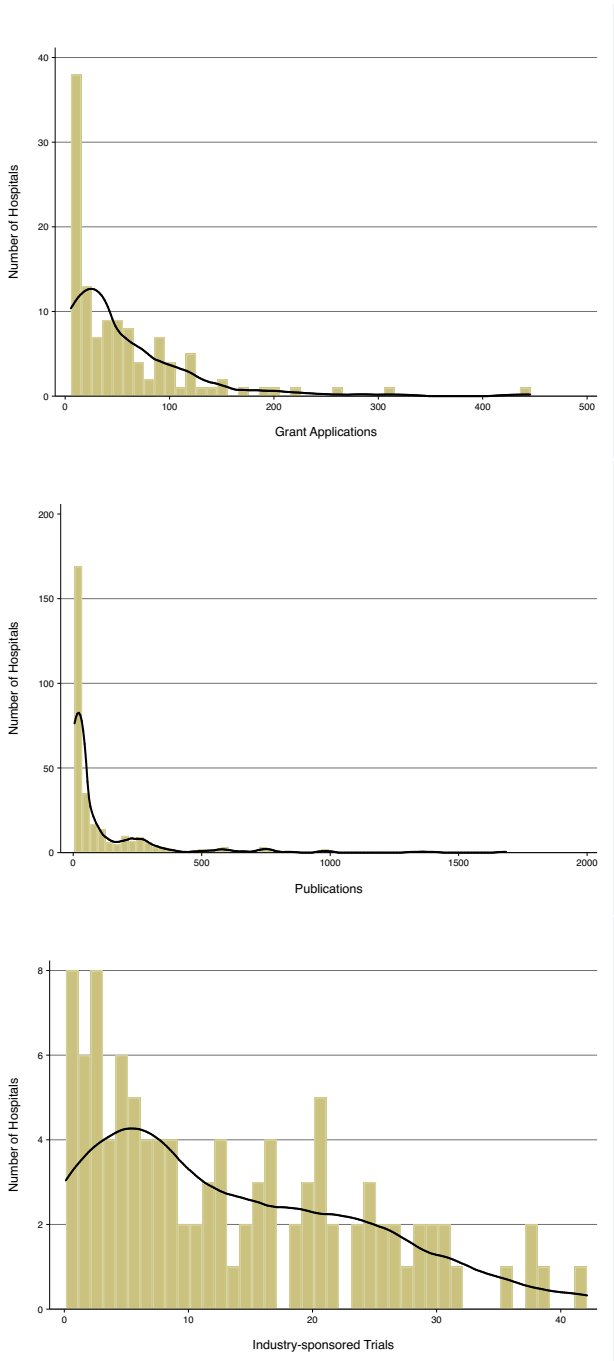
Looking to the other end of the bench-to-bedside continuum, Panel D provides an overview of the level of funding that hospitals receive from contracts to run clinical trials. On average, hospital facilitate and/or act as sites for 1.8 industry-sponsored clinical trials and receive $163,000 in return. The dispersion across hospitals is large: while some hospitals do not conduct any trials, some hospitals conduct more than 40. The skewness of these research outcomes are demonstrated in Figure 3.2. Figures 3.3 and 3.4 show that across all hospitals, there is a general increase in the level of grant applications and publications over our study period. Figure 3.5 shows that the level of industry-sponsored trial sites has remained relatively constant.

Table 3.1: *Summary Statistics*

	Observations	Mean	Median	Standard Deviation	Minimum	Maximum
Panel A. Hospital Characteristics						
Discharges (1,000s)	725	18.57	16.38	10.78	0.47	65.82
Medicare Share of Discharges	725	0.33	0.34	0.11	0.02	0.64
Medicare Teaching Payment ($1,000,000s)	725	5.67	2.52	7.97	0.00	59.76
Medicare Disproportionate Share Payment ($1,000,000s)	725	4.37	2.69	4.83	0.00	36.93
Panel B. Grants						
Number of Grant Applications						
Total	725	8.87	0.06	33.10	0.00	444.00
Renewal	725	1.68	0.00	6.62	0.00	88.25
New	725	7.19	0.06	26.54	0.00	355.75
MD	725	3.06	0.00	11.80	0.00	158.62
PhD	725	4.33	0.00	16.45	0.00	193.38
MD-PhD	725	1.36	0.00	5.82	0.00	87.75
Funding Amount ($ Mill. Dollars)						
Total	725	13.09	0.00	51.93	0.00	709.37
Renewal	725	3.59	0.00	15.11	0.00	211.80
New	725	9.51	0.00	37.10	0.00	497.57
MD	725	5.38	0.00	22.32	0.00	297.43
PhD	725	5.34	0.00	20.89	0.00	256.55
MD-PhD	725	2.30	0.00	10.79	0.00	153.27
Panel C. Publications						
Total	725	46.44	2.38	150.84	0.00	1,683.62
Citation Ranking: $\leq$25	725	11.44	0.94	31.37	0.00	306.12
Citation Ranking: 26-50	725	10.65	0.56	32.49	0.00	333.38
Citation Ranking: 51-75	725	11.30	0.38	37.84	0.00	413.19
Citation Ranking: 76-90	725	7.37	0.19	27.16	0.00	325.19
Citation Ranking: 91-95	725	2.68	0.06	10.77	0.00	131.44
Citation Ranking: 96-99	725	2.32	0.06	10.15	0.00	130.56
Citation Ranking: >99	725	0.69	0.00	3.26	0.00	45.12
"Basic Research": Cited in Patent	725	6.73	0.19	25.67	0.00	316.44
"Basic Research": MeSH	725	6.89	0.00	26.68	0.00	269.44
"Translational" Research	725	9.73	0.19	33.82	0.00	403.94
Builds on "Translational" Research	725	2.79	0.19	8.65	0.00	85.75
Inspires "Translational" Research	725	4.44	0.06	16.45	0.00	208.94
Panel D. Trials						
Number of Trials	725	1.80	0.00	6.01	0.00	41.62
Trials Sites (1,000s)	725	163.27	0.00	567.82	0.00	4,522.10

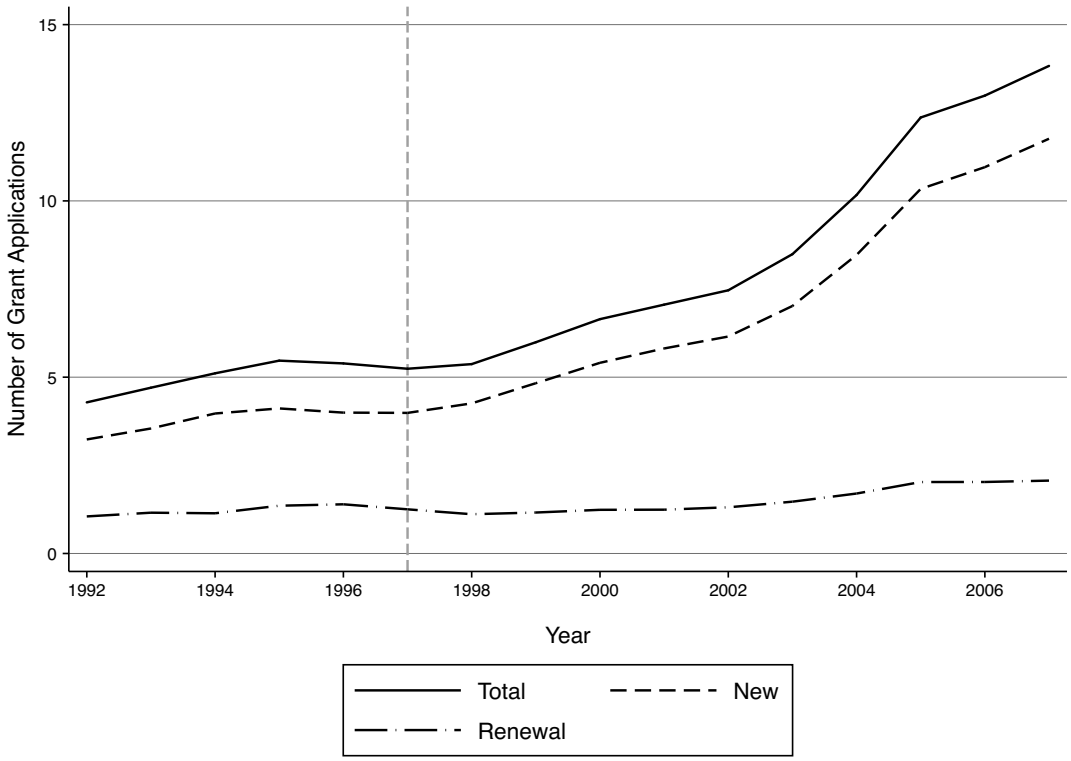
Notes: This table shows summary statistics for the primary analytic hospital sample between 1992-2007. All variables are measured yearly. For example, "Discharges (1000s)" is the average number of patients a hospital receives in a year. The hospital sample used is the primary analytic hospital sample.

Figure 3.2: *Distribution of Research Outcomes*



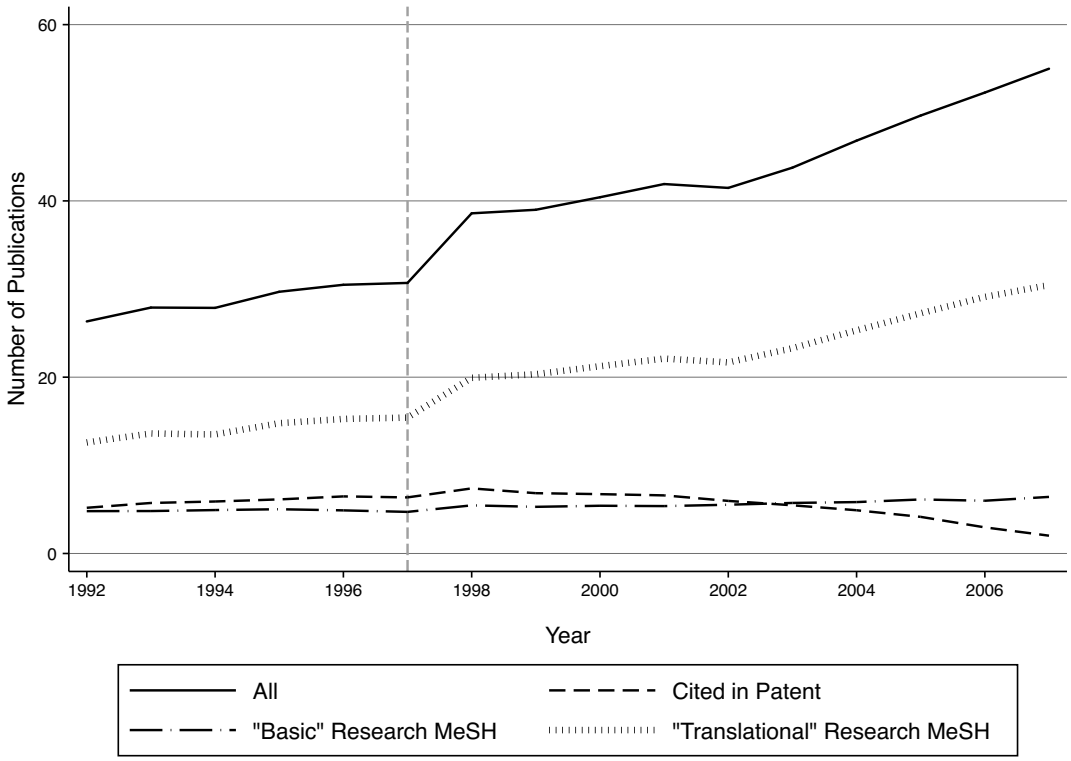
Notes: This figure shows histograms of the average number of research outcomes (grant applications, publications, industry-sponsored trials) across hospitals. For clarity, the sample of hospitals is restricted to those with at least five outcomes.

Figure 3.3: *Change in NIH Grant Applications, 1992-2007*



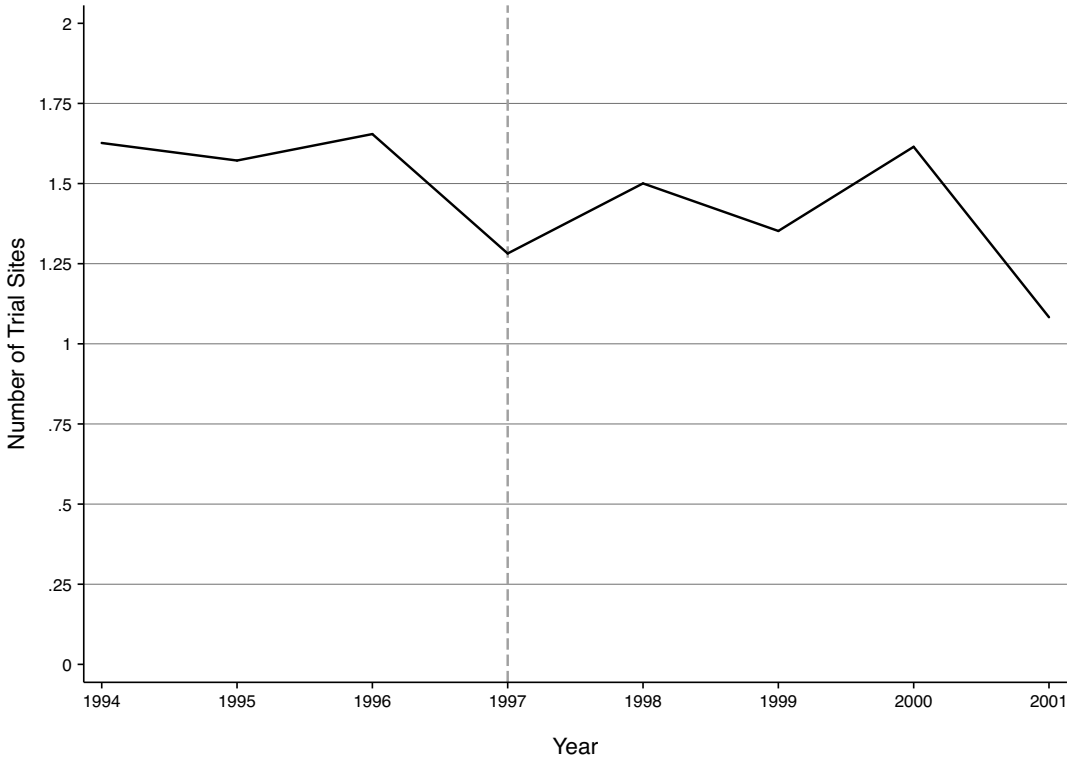
Notes: This figure plots the annual number of grant applications, averaged across all hospitals. The dashed line indicates the year in which the BBA came into effect.

Figure 3.4: *Change in Publications, 1992-2007*



Notes: This figure plots the annual number of publications, averaged across all hospitals. The dashed vertical line indicates the year in which the BBA came into effect.

Figure 3.5: *Change in Industry-Sponsored Trials, 1994-2007*



Notes: This figure plots the annual number of industry-sponsored trial sites, averaged across all hospitals. Data is industry-sponsored trials is only available from 1994 to 2001. The dashed vertical line indicates the year in which the BBA came into effect. See the text for details on variable construction.

Appendix Table C.1 provides summary statistics for hospitals in the AMC sample. A comparison of the primary analytic hospital sample and the AMC sample can be found in Appendix Table C.2. Relative to hospitals in the primary analytic hospital file, AMC sample hospitals produce significantly more grant applications and publications, and are more likely to engage in facilitating clinical trials.

3.4 Empirical Strategy and Results

To examine the impact of the BBA on subsequent research activity, we exploit the fact that some hospitals were more exposed to the reform than others. In particular, there are substantial differences in the importance of Medicare as a source of revenue, generating heterogeneity in the potential impact of the BBA across hospitals. That is, hospitals with a greater share of patients that are funded by Medicare (or "Medicare Share") in the pre-BBA period experienced a greater loss in revenues per discharge.

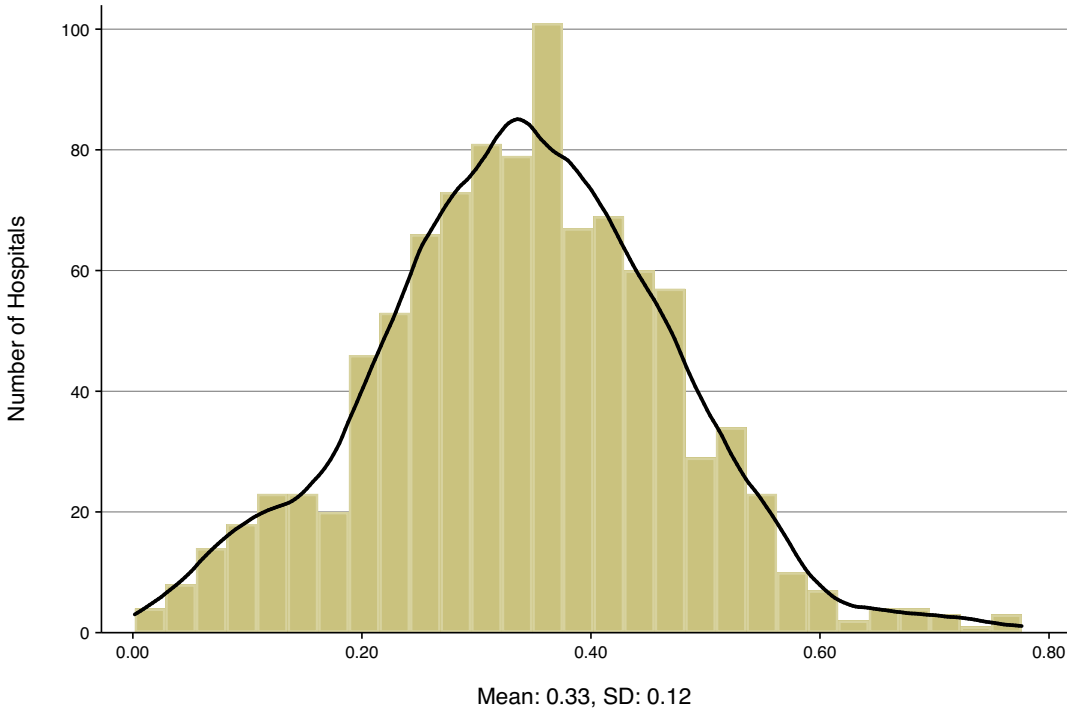
Our empirical strategy thus consists of comparing research outputs, before and after the implementation of the BBA, between hospitals that faced a potentially large decrease in the level of Medicare reimbursements (and have a relatively higher Medicare Share) to those that were minimally impacted by the reform (with a smaller Medicare Share). This follows a large literature which measures the impact of changes in Medicare reimbursements by using the share of a hospital's discharges that are reimbursed by Medicare in the period prior to the change (Acemoglu and Finkelstein, 2008; Kaestner and Guardado, 2008; Wu and Shen, 2015).

Our measure of Medicare Share comes from identifying the pre-BBA share of patient discharges that are funded by Medicare for each hospital h:

$$MedicareShare_h = \overline{\left(\frac{MedicareDischarges_{h,t}}{TotalDischarges_{h,t}}\right)}_{t=1992 \text{ to } 1995} \tag{3.1}$$

Figure 3.6 shows hospital-level variation in the Medicare Share.

Figure 3.6: *Distribution of Medicare Shares*



Notes: This figure shows a histogram of the annual Medicare Share, averaged over 1992-1995. The hospital sample used is the primary analytic hospital sample.

Appendix Figures C.1 and C.2 report the relationship between pre-BBA Medicare Share and hospital subsidy payments in the data. The x-axis is the Medicare Share. The y-axis measures the inverse hyperbolic sine of DSH and IME payments in the two years that immediately follow the BBA. Taking an inverse hyperbolic sine of a variable is similar to a natural logarithm transformation, but the inverse hyperbolic transformation is defined at 0 (Burbidge *et al.*, 1988). All standard errors are robust and clustered by hospital. Binned scatterplots are shown, with subsidy payments residualized on hospital fixed effects and year fixed effects. As expected, hospitals with a higher Medicare Share experience an unambiguous decrease in DSH payments (Figure A.1). The fact that hospitals with a higher Medicare Share do not experience a significant decrease in IME payments may seem surprising (Figure A.2). However, as discussed in Section 3.2, hospitals may have attempted to offset the potential loss in IME payments through increasing the number of residents (Cromwell *et al.*, 2006) or inpatient admissions (Bazzoli *et al.*, 2004). Efforts to offset the loss of clinical care revenues may impact the rate and direction of subsequent research activity. The goal of this empirical analysis, therefore, is to examine the impact of the BBA *net* of hospitals' off-setting behaviors.

With h indexing hospitals and t indexing years, we estimate regressions of the form:

$$y_{h,t} = \beta MedicareShare_h \times After_t + \delta_h + \zeta_t + \epsilon_{h,t} \tag{3.2}$$

where $MedicareShare_h$ is the Medicare Share in hospital h, $After_t$ is an indicator is equal to 1 after 1997, δ_h are hospital fixed effects, and ζ_t are year fixed effects. Given the skewness of the outcome variables, we use an inverse hyperbolic sine transformation on our outcome variable in the majority of our regressions. The coefficient of interest is β, which is the difference-in-differences estimate of the effect of the BBA on subsequent research outcomes.

One potential threat to estimation is that the NIH doubled its budget between 1998 and 2003 (Korn *et al.*, 2002). If the NIH budget increase disproportionately decreased (or increased) funding for the type of research conducted by high Medicare Share hospitals, the resulting estimates will be negatively (or positively) biased. However, in a separate analysis

we find that the doubling of the NIH budget did not disproportionately affect funding for research conducted by hospitals with high Medicare Shares.

3.4.1 Impact on the Rate of Innovation

Table 3.2 present the estimates of the impact of the BBA on the level of grant applications. To ease interpretation of the results, we present in the third row the $MedicareShare_h \times After_t$ coefficient multiplied by 1 standard error of the Medicare share (0.12).

The first column of Table 3.2 describes the impact of the BBA on the total number of grant applications. Column 1 shows that a 1 SD increase in a hospital's Medicare Share translates into a 4 percent decrease in the total number of subsequent grant applications. In subsequent columns, we examine the impact of cuts in Medicare reimbursement on different types of grants (by grant cycle and principal investigator type). The estimates presented in Column 2 suggest that this effect is largely driven by a decline in new grant applications. In contrast, there is not a statistically significant effect on the number of renewals, which is likely due to the fact that the process of seeking a novel grant application is more resource intensive (and may be more dependent on cross-subsidies) than the process of renewing a grant. Estimates in Columns 4-6 show that financing cuts affect research conducted by MDs, PhDs, and MD/PhDs similarly.

Table 3.2: *Impact on the Number of Grant Applications*

		Grant Cycle		Principal Investigator		
	Total	New	Renewal	MD	PhD	MD-PhD
	(1)	(2)	(3)	(4)	(5)	(6)
Medicare Share × After	-0.352***	-0.422***	-0.000506	-0.244**	-0.381***	-0.430***
	(0.106)	(0.110)	(0.0735)	(0.0981)	(0.0921)	(0.0906)
Mean of Dep. Var	0.78	0.73	0.38	0.53	0.55	0.34
% Change in 1 SD Share	-4.146	-4.97	-0.01	-2.87	-4.49	-5.06
Observations	11,513	11,513	11,513	11,513	11,513	11,513
Test for Diff. in Percent Change		0.00		0.19		

Notes: This table displays the effect on the number of grant applications. The hospital sample used is the primary analytic hospital sample. Outcome variables have been transformed with inverse hyperbolic sine transformations. Column 1 presents results from OLS regressions. Columns 2-6 present estimates from seemingly unrelated regressions. Hospital fixed effects and year fixed effects are included in all regressions, and standard errors are clustered at the hospital level. The third row shows the coefficient on Medicare Share × After multiplied by 1 SD in Medicare Share. The fourth row shows t-test results from comparing estimates in different columns (Column 2 vs. Column 3; Column 4 vs. Column 5).
*p<0.10, p**<0.05, ***p<0.001.

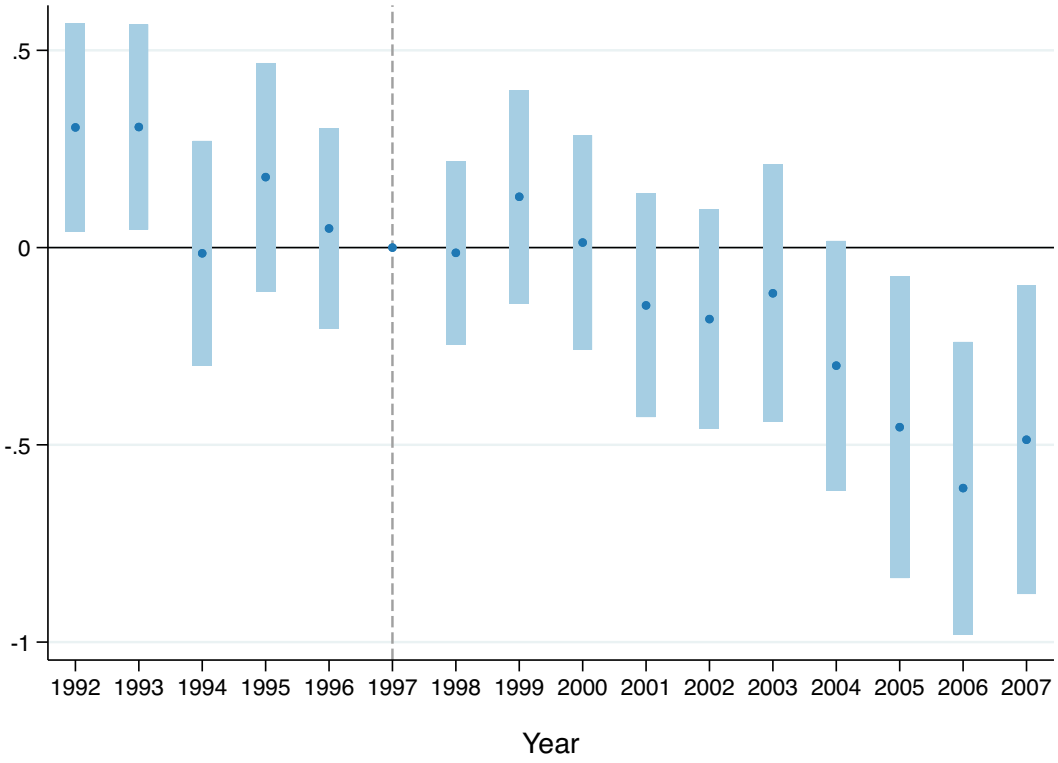
To explore the timing of these estimated effects, we estimate:

$$Y_{h,t} = \alpha + \sum_{z} \beta_z \times 1(z) \times MedicareShare_h + \delta_h + \tau_t + \epsilon_{h,t} \tag{3.3}$$

where δ_h and τ_t represent hospital and year fixed effects, respectively, for hospital t and year t. z represents the "lag," or the years relative 1997, which is the year in which the BBA is implemented.

Figure 3.7 presents estimates of β_z from performing the regression on the primary analytic hospital sample and corresponds to a dynamic version of Table 3.2, Column 1. The light blue colored lines represent the 95-percent confidence intervals and the dashed gray line represents the year in which the BBA was enacted. There are two takeaways from this figure: first, there is no significant evidence of pre-treatment trends in the years right before the BBA. Second, the impact of the BBA takes nearly 8 years to occur, suggesting that any immediate impact of the BBA may have been mitigated by AMCs' reliance on alternative sources of funding (e.g., clinical trial contracts with pharmaceutical firms), which we explore in subsequent sections.

Figure 3.7: *Impact on the Number of Grant Applications*



Notes: The dark blue dots in the figure above correspond to coefficient estimates stemming from OLS specifications in which the number of grant applications is regressed onto year effects, hospital effects, as well as interaction terms between a hospital's Medicare Share interacted with the number of years before/after the BBA (the Medicare share interacted with the year of the BBA, 1997, is omitted). The 95% confidence interval (corresponding to robust standard errors, clustered around hospital) around these estimates is plotted with light blue bars. The dashed grey line corresponds to the year of the BBA. The hospital sample used is the primary analytic hospital sample.

Table 3.3, Column 1 presents the estimates of the impact of health care financing cuts on the total number of publications. We find results that are similar to the effect on the number of grant applications: a 1 SD increase in a hospital's Medicare Share is associated with a 4 percent decrease in the total number of subsequent publications. Figure 3.8 provides the corresponding event study results for the primary analytic hospital sample. The figure shows that the BBA effect takes place 10 years after the reform is enacted, likely due to the reasons discussed above, as well as the time lag between research and eventual publication.

Table 3.3: *Impact on the Number of Publications, Quality*

	Total	Citation Ranking						
		≤25	26-50	51-75	76-90	91-95	96-99	>99
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Medicare Share × After	-0.304*	-0.303**	-0.300**	-0.390**	-0.497***	-0.431***	-0.339**	-0.256***
	(0.163)	(0.137)	(0.127)	(0.129)	(0.126)	(0.105)	(0.107)	(0.072)
Mean of Dep. Var	2.06	1.39	1.19	1.12	0.88	0.53	0.48	0.24
% Change in 1 SD Share	-3.58	-3.57	-3.54	-4.59	-5.85	-5.08	-3.99	-3.02
Observations	11,513	11,513	11,513	11,513	11,513	11,513	11,513	11,513
Test for Diff. in Percent Change	-	0.99	0.97	0.42	0.12	0.37	0.82	0.77

Notes: This table displays the effect on total publications and publication quality in hospitals. The hospital sample used is the primary analytic hospital sample. Outcome variables have been transformed with inverse hyperbolic sine transformations. Estimates are from seemingly unrelated regressions. Hospital fixed effects and year fixed effects are included in all regressions, and standard errors are clustered at the hospital level. The third row shows the coefficient on Medicare Share × After multiplied by 1 SD in Medicare Share. The fourth row shows t-test results from comparing estimates in Column 1 vs. estimates in other columns (Column 1 vs. Column 2; Column 1 vs. Column 2, etc.).
*p<0.10, p**<0.05, ***p<0.001.

Figure 3.8: *Impact on the Number of Grant Application*

Notes: The dark blue dots in the figure above correspond to coefficient estimates stemming from OLS specifications in which the number of publications is regressed onto year effects, hospital effects, as well as interaction terms between a hospital's Medicare Share interacted with the number of years before/after the BBA (the Medicare share interacted with the year of the BBA, 1997, is omitted). The 95% confidence interval (corresponding to robust standard errors, clustered around hospital) around these estimates is plotted with light blue bars. The dashed grey line corresponds to the year of the BBA. The hospital sample used is the primary analytic hospital sample.

3.4.2 Impact on the Quality of Innovation

Next, we examine how the BBA shapes the quality of subsequent innovation. Cuts to health care financing have ambiguous effects on the quality of subsequent innovation. On the one hand, researchers with limited resources may direct research efforts towards less resource-intensive research activities. To the extent that less resource-intensive activities are generally lower quality, negative financial shocks may lead researchers to disproportionately decrease high quality research investments. On the other hand, researchers with limited resources may focus their efforts only on research projects that they believe to be high quality and high impact ex-ante. This suggest that cuts to financing may not have a disproportionate impact on the quality of subsequent research.

Following a long line of literature that uses citations as a proxy for quality and impact (Gittelman and Kogut, 2003; Hall *et al.*, 2005), we use a publication's citation ranking as a proxy for publication quality. We report estimates across citation rankings in Table 3.8. Successive columns show the BBA's impact on the rate of publications with citations rankings around a series of percentile thresholds, starting from below the 25^{th} to above the 99^{th}. We see that low and high quality innovation responded similarly to the BBA: the response for each publication quality type (e.g., those with low or high levels of citations) are similar in magnitude and not statistically significantly different from the impact on total publications (shown in Column 1).

3.4.3 Impact on the Direction of Innovation

Finally, we explore how negative financial shocks influence the direction of subsequent innovation. The discussion in Section 3.2 suggests that in the presence of negative financial shocks, AMC researchers should be less likely to invest in "translational" research.

Using the definitions outlined in Section 3.3, Table 3.4 shows that negative financial shocks leads to "hollowing out" of subsequent research: in response to the BBA, first redirect research efforts away from "translational" research, towards "basic" and "applied" research.

To start, Column 1 reveals that the BBA causes an increase in subsequent "basic" research based on the patent definition: for each SD increase in Medicare share, the number of bench research publications increases by 2 percent. In contrast, Column 2 shows that there is no statistically significant relationship between the BBA and bench research as measured by the MeSH definition. Turning to our measures of "translational" research, we estimate that the BBA leads to a nearly 4 percent decrease in subsequent research activity (Column 3). Columns 4 and 5 show that this dampening effect also extends to research that "builds on translational research" and "inspires translational research."

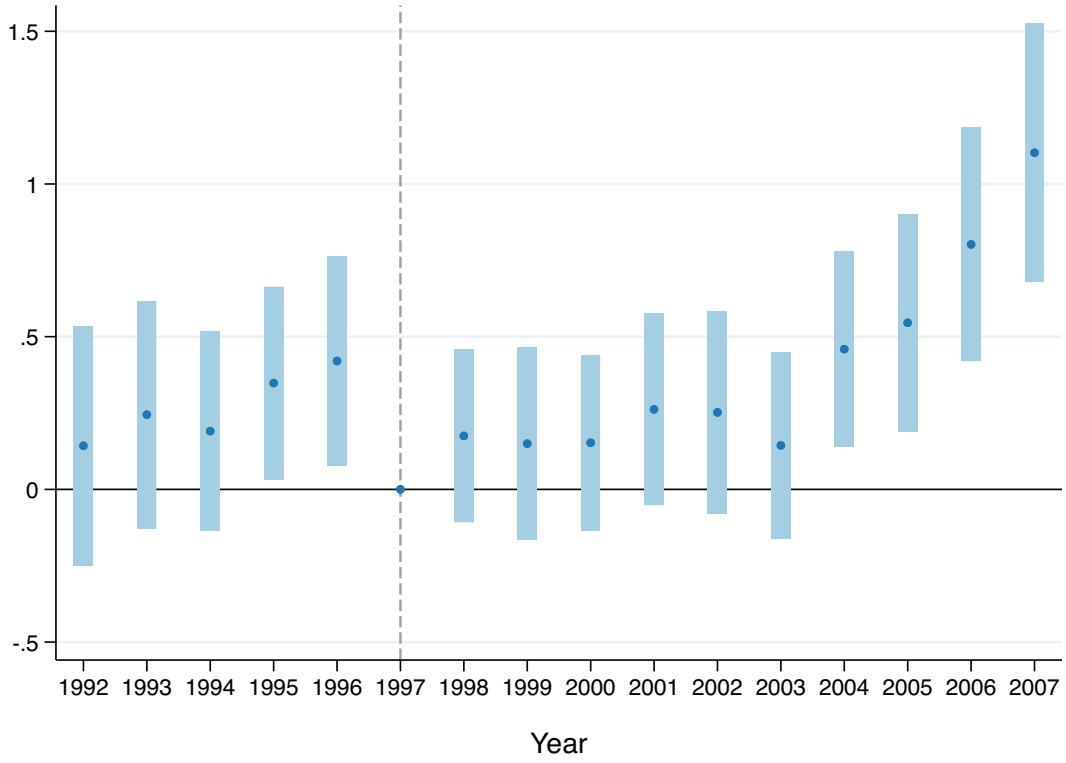
Finally, Columns 6 and 7 suggest that hospitals more exposed to the reform are more likely to increase their "applied" research activity through obtaining more industry-sponsored contracts to conduct clinical trials. Figures 3.10, 3.10, and 3.11 provide event study graphs for "bench" research (Figure 3.9), "translational" research (Figure 3.10), and "bedside" research (Figure 3.11). Taken together, the "hollowing out" of "translational" research suggests that one mechanism through which the BBA affects subsequent research efforts is to shift hospital-level research activity along the extensive margin (i.e., through hiring more "bench" and "bedside" researchers, and fewer physician-scientists conducting "translational" research). Future versions of this paper will explore this workforce shift in more detail.

Table 3.4: *Impact on Research Direction*

	"Bench" Research		"Translational" Research			"Bedside" Research	
	Publications, Cited in Patent	Publications, Basic MeSH	Publications, "Translational" MeSH	Publications, Builds on "Translational" MeSH	Publications, Inspiring "Translational" MeSH	Number of Industry-Sponsored Trial Sites	Industry-Sponsored Trial Dollars ($)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Medicare Share × After	0.179*	-0.143	-0.352**	-0.414***	-0.394***	0.169**	0.645*
	(0.106)	(0.120)	(0.130)	(0.118)	(0.115)	(0.0567)	(0.370)
Mean of Dep. Var.	0.80	0.64	0.95	0.69	0.67	0.37	1.79
% Change in 1 SD Share	2.126	-1.665	-4.057	-4.764	-4.539	1.995	7.609
Observations	11,513	11,513	11,513	11,513	11,513	5,770	5,770

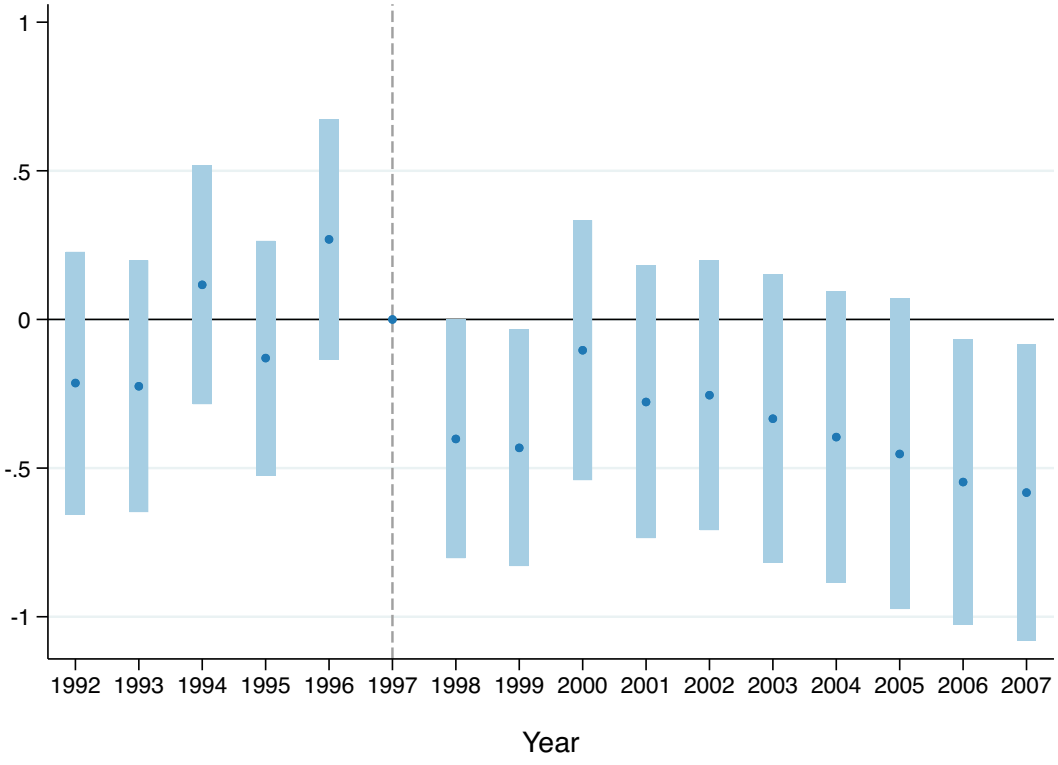
Notes: This tables displays the effect on "bench," "translational," and "bedside" research in hospitals. The hospital sample used is the primary analytic hospital sample. Outcome variables have been transformed with inverse hyperbolic sine transformations. Estimates are from OLS regressions. Hospital fixed effects and year fixed effects are included in all regressions, and standard errors are clustered at the hospital level. The third row shows the coefficient on Medicare Share × After multiplied by 1 SD in Medicare Share. Column 2 refers to publications with MeSH terms affiliated with a molecular biology technique or that use a model organism. Column 3 refers to publications that are disease-oriented and relies either on a molecular biology technique or a model organism (based on MeSH terms). Column 4 refers to publications that report the results of a clinical trials, or are tagged by a human MeSH term and also cite a translational publication. Column 5 refers to publications that are translational and is cited by a clinical trial publication (or one that contains a human MeSH term). Finally, Columns 6 and 7 refer to clinical trial contracts.
*p<0.10, p**<0.05, ***p<0.001.

Figure 3.9: *Impact on Bench Research*
(Number of Publications Cited by a Patent)



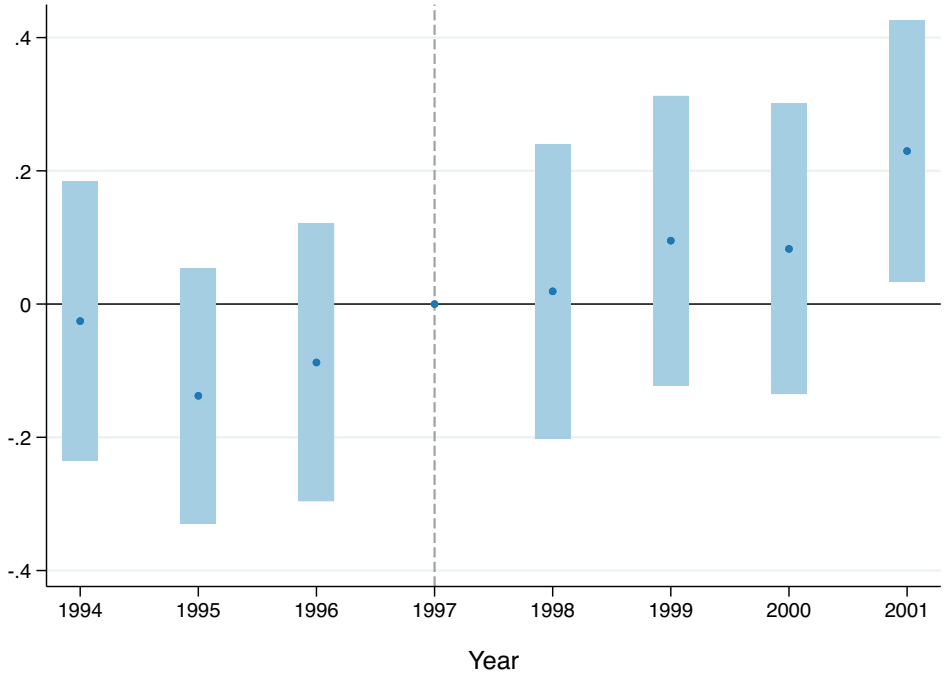
Notes: The dark blue dots in the figure above correspond to coefficient estimates stemming from OLS specifications in which the number of publications cited by a patent is regressed onto year effects, hospital effects, as well as interaction terms between a hospital's Medicare Share interacted with the number of years before/after the BBA (the Medicare share interacted with the year of the BBA, 1997, is omitted). The 95% confidence interval (corresponding to robust standard errors, clustered around hospital) around these estimates is plotted with light blue bars. The dashed grey line corresponds to the year of the BBA. The hospital sample used is the primary analytic hospital sample.

Figure 3.10: *Impact on Translational Research*
(Number of "Translational" Publications)



Notes: The dark blue dots in the figure above correspond to coefficient estimates stemming from OLS specifications in which the number of "translational" publications is regressed onto year effects, hospital effects, as well as interaction terms between a hospital's Medicare Share interacted with the number of years before/after the BBA (the Medicare share interacted with the year of the BBA, 1997, is omitted). "Translational" publications are those that are disease-oriented, relies either on a molecular biology technique or a model organism, and is not a clinical trial publication (based on MeSH terms). The 95% confidence interval (corresponding to robust standard errors, clustered around hospital) around these estimates is plotted with light blue bars. The dashed grey line corresponds to the year of the BBA. The hospital sample used is the primary analytic hospital sample.

Figure 3.11: *Impact on Bedside Research (Number of Industry-Sponsored Trial Sites)*



Notes: The dark blue dots in the figure above correspond to coefficient estimates stemming from OLS specifications in which the number of industry-sponsored clinical trial sites is regressed onto year effects, hospital effects, as well as interaction terms between a hospital's Medicare Share interacted with the number of years before/after the BBA (the Medicare share interacted with the year of the BBA, 1997, is omitted). The 95% confidence interval (corresponding to robust standard errors, clustered around hospital) around these estimates is plotted with light blue bars. The dashed grey line corresponds to the year of the BBA. The hospital sample used is the primary analytic hospital sample.

Performing the analysis on the AMC Sample (Appendix Tables C.3, C.4, and C.5) reveals similar results. Among research-intensive hospitals, a 1 SD increase in a hospital's Medicare Share translates into a 9 percent decrease in the total number of subsequent grant applications. We see a similarly negative effect on the subsequent number of publications and a similar "hollowing out" of subsequent research, though the results are not statistically significant, likely due to the smaller sample size.[14]

3.5 Discussion and Conclusion

3.5.1 Discussion

In previous sections, we provide evidence consistent with anecdotal and survey evidence suggesting that cross-subsidies of clinical care revenues play a key role in seeding research performed by physician-scientists within AMCs (Jones and Sanderson, 1996; Weissman *et al.*, 1999). Specifically, we show that cuts to health care financing lead to a decrease in the total number of grant applications and publications. The findings above are consistent with the view that physician-scientists within AMCs are allocating their time and effort away from research, towards other activities such as patient care. As a result, the potentially welfare-decreasing effects associated with a decrease in subsequent research activity may be countered by improvements in clinical care quality. To better understand the effects of the BBA on overall welfare, we next explore whether the BBA impacts patient outcomes within the hospitals in our sample.

Our main measures of clinical outcomes are from Chandra *et al.* (2016)'s analysis of the relationship between hospital quality and market size. In particular, Chandra *et al.* (2016) construct hospital-level measures of 30-day risk-adjusted survival for four conditions: heart attacks (called acute myocardial infarction, or AMI), congestive heart failure, pneumonia, and hip and knee replacements (a common pair of surgical procedures). The authors construct

[14]As a robustness check, we confirmed that our results in the primary analytic hospital sample are not being driven by hospitals *not* found in the AMC sample.

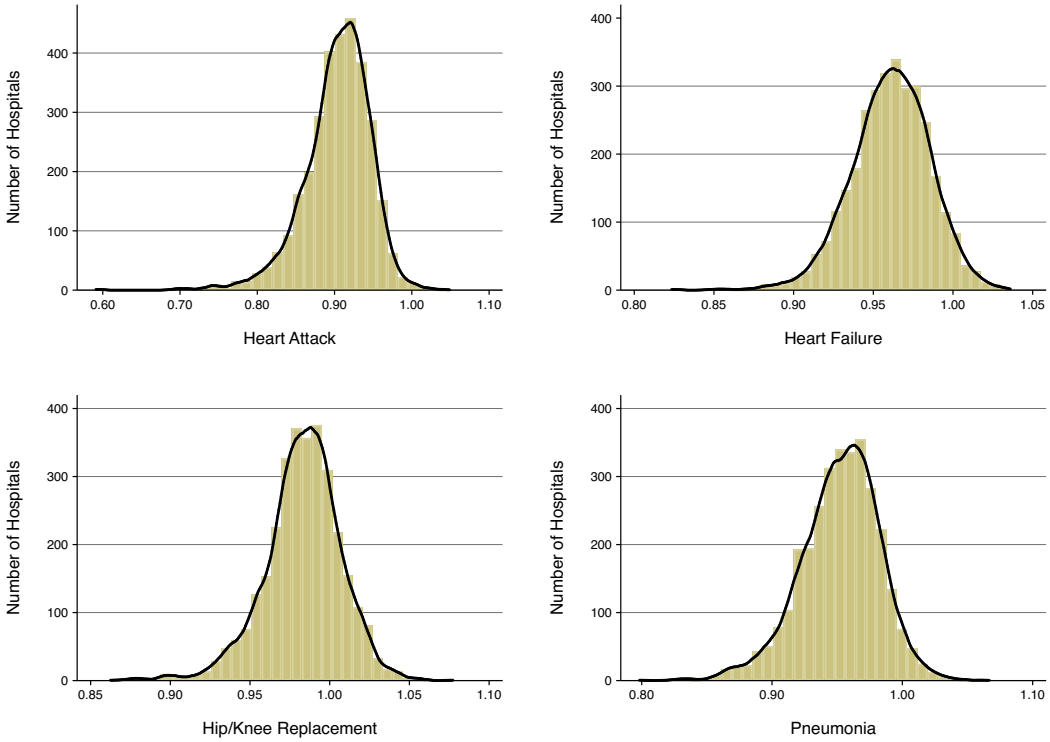
condition-specific measures over three-year periods. For example, 30-day risk-adjusted survival for 1996 is calculated over patients claims from 1994-1996. For each hospital, we use three-year bins for three years (1996, 2002, 2005). Table 3.5 provides an overview of the survival outcomes and Figure 3.12 presents a histogram of the survival rates for each of the conditions.

Table 3.5: *Summary of 30-Day Risk-adjusted Survival Rates*

	Observations	Mean	Median	Standard Deviation	Minimum	Maximum
Heart Attack	668	0.91	0.91	0.03	0.73	1.00
Heart failure	668	0.96	0.96	0.02	0.90	1.01
Pneumonia	668	0.98	0.98	0.02	0.90	1.04
Hip/knee	668	0.95	0.95	0.02	0.86	1.04

Notes: This table shows clinical outcomes summary statistics. The hospital sample used is the primary analytic hospital sample. Clinical outcomes are measured in three-year bins – e.g., hospital-level survival rates in 2005 are estimated over patient claims in 2003, 2004, and 2005. For each hospital, we use three-year bins for four years (1996, 2002, 2005, and 2008).

Figure 3.12: *Risk-Adjusted 30-Day Survival Rates*



Notes: These figure shows histograms of the average risk-adjusted 30 day survival rates across hospitals in the primary analytic hospital sample for patients following a heart attack, heart failure, hip/knee replacement, and pneumonia.

We next estimate the causal impact of the BBA on survival rates for the four conditions. Because our clinical outcome data is in 3-year bins, with h indexing hospitals and c indexing conditions, we estimate long difference regressions of the following form:

$$\frac{surv_{h,c,2005'} + surv_{h,c,2002'}}{2} - surv_{h,c,1996'} = \beta MedicareShare_h + \epsilon_h \tag{3.4}$$

where $surv_{h,c,t'}$ is the risk-adjusted 30-day mortality rate in hospital h, for condition c, estimated over patients claims from $t-2$ to t and $MedicareShare_h$ is the Medicare share in hospital h.

Table 3.6 reports the results. We find that, within our hospital sample, the BBA does not affect survival rates for the majority of the conditions. This suggest that to the extent that physician-investigators substituted away from research, towards patient care activities, there was no concurrent improvement in patient outcomes. This is consistent with Volpp *et al.* (2005) that find minimal or no positive impact of the BBA on patient outcomes (Seshamani *et al.*, 2006; Volpp *et al.*, 2005; Wu and Shen, 2015). Insofar as medical research can lead to welfare improvements, the sum of our results—that financing cuts decrease subsequent innovation, but do not lead to changes in clinical outcomes—suggest that health care financing cuts have the potential to be welfare-decreasing.

Table 3.6: *Impact on Clinical Outcomes*

	(1) Heart Attack	(2) Heart Failure	(3) Hip/Knee	(4) Pneumonia
Medicare Share	-0.0234	0.00464	0.0141	-0.0326**
	(0.0186)	(0.0105)	(0.0100)	(0.0126)
Mean of Dep. Var	0.03	0.01	-0.00	0.01
% Change in 1 SD Share	-0.253	0.0502	0.152	-0.352
Observations	668	668	668	668

Notes: This table displays the effect on risk-adjusted survival rates in hospitals in the primary analytic hospital sample. Outcomes are the difference in average survival rates between the post-BBA time period and the pre-BBA time period. Estimates are from OLS regressions. Robust standard errors are clustered at the hospital level.
*p<0.10, p**<0.05, ***p<0.001.

3.5.2 Conclusion

This paper examines how the impacts of financing cuts can spread throughout a key institution in the U.S. innovation system. We examine an institution that bridges the gap between the medical research and medical care, and find a substantial decrease in subsequent research activity. Using a differences-in-difference approach, we show that research falls by 4 percent in response to a decrease in clinical care payments. We also find that cuts to health care financing lead to a "hollowing out" of "translational" research, as physician-scientists redirect research efforts towards "bench" and "bedside" research, and away from research that bridges the two.

This paper has limitations and suggests avenue for future research. First, we study how changes in financing affect research and teaching-intensive hospitals, and the results may not be applicable to other settings. Second, our analysis focuses on the BBA, whose impacts were somewhat muted by subsequent policies (BBRA, BIPA). It would be useful for future work to explore how financing cuts influence research activity in contexts experiencing longer-term financing cuts. Third, we study research activity on one dimension: the hospital level. Future work could usefully provide a more nuanced understanding of how financing cuts influence allocation of effort (across research, patient care, and teaching) by examining responses at the level of the individual physician-researcher.

Growing concern about health care costs raises important questions about how public policies should respond. Bridging the health policy and innovation literatures, this paper presents evidence that cost-containment efforts may have spillover effects that can impact future health outcomes through shifting the rate of biomedical innovation. As governments seek to craft policies to make the health care system more efficient, understanding how innovative activity responds is essential to designing policies aimed at improving welfare.

References

Abbott, L. (1955). *Quality and Competition; an Essay in Economic Theory*. New York: Columbia University Press.

Acemoglu, D. and Finkelstein, A. (2008). Input and technology choices in regulated industries: Evidence from the health care sector. *Journal of Political Economy*, **116** (5), 837–880.

— and Linn, J. (2004). Market size in innovation: Theory and evidence from the pharmaceutical industry. *The Quarterly Journal of Economics*, **119** (3), 1049–1090.

Adjei, A. A., Christian, M. and Ivy, P. (2009). Novel designs and end points for phase ii clinical trials. *Clinical Cancer Research*, **15** (6), 1866–1872.

Ahmadpoor, M. and Jones, B. (2017). The dual frontier: Patented inventions and prior scientific advance. *Science*, **357** (6351), 583–587.

Ali, A. and Gittelman, M. (2016). Research paradigms and useful inventions in medicine: Patents and licensing by teams of clinical and basic scientists in academic medical centers. *Research Policy*, **45** (8), 1499–1511.

Allain, M.-L., Henry, E. and Kyle, M. (2016). Competition and the efficiency of markets for technology. *Management Science*, **62** (4), 1000–1019.

American Cancer Society (2018). Key statistics for ovarian cancer. `https://www.cancer.org/cancer/ovarian-cancer/about/key-statistics.html`, Last Accessed on 2018-10-01.

Anderson, K., Monique L. Chiswell, Peterson, E. D., Tasneen, A., Topping, J. and Califf, R. M. (2015). Compliance with results reporting at clinicaltrials.gov. *The New England Journal of Medicine*, **372** (11), 1031–1039.

Angrist, J. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist?s Companion*. Princeton, New Jersey: Princeton University Press.

Arora, A. and Gambardella, A. (1994). Evaluating technological information and utilizing it. *Journal of Economic Behavior and Organization*, **24** (1), 91–114.

AstraZeneca (2011). *AstraZeneca Annual Report and Form 20-F Information 2011*. Tech. rep., AstraZeneca.

— (2017). *Delivering the Next Wave of Scientific Innovation: 2017 - a Year in Review*. Tech. rep., AstraZeneca.

Avorn, J. (2015). The $2.6 billion pill—methodologic and policy considerations. *The New England Journal of Medicine*, **372** (20), 1877–1879.

Azoulay, P., Graff Zivin, J. S., Li, D. and Sampat, B. N. (2018). Public r&d investments and private-sector patenting: Evidence from nih funding rules. *The Review of Economic Studies*, **24** (1), 91–114, rdy034.

—, Greenblatt, W. H. and Heggeness, M. L. (2019). Long-term effects from early exposure to research: Evidence from the nih "yellow berets".

Bach, P. (2015). Indication-specific pricing for cancer drugs. *JAMA*, **312** (16), 1629–1630.

Barker, A. D. and Collins, F. S. (2008). Mapping the cancer genome. *Scientific American*.

Bazzoli, G. J., Lindrooth, R. C., Hasnain-Wynia, R. and Needleman, J. (2004). The balanced budget act of 1997 and u.s. hospital operation. *Inquiry*, **41** (4), 401–417.

Berger, V. W. and Alperson, S. Y. (2009). A general framework for the evaluation of clinical trial quality. *Reviews on Recent Clinical Trials*, **4** (2), 79–88.

Berndt, E. R., Cockburn, I. M. and Grepin, K. A. (2006). The impact of incremental innovation in biopharmaceuticals. *Pharmacoeconomics*, **24** (2), 69–86.

Bloom, N., Jones, C. I., Van Reenen, J. and Webb, M. (2018). Are ideas getting harder to find?, working Paper.

—, Propper, C., Seiler, S. and Van Reenen, J. (2015). The impact of competition on management quality: Evidence from public hospitals. *Review of Economic Studies*, **0**, 1–33.

Blumenthal, G. (2017). Primer on drug development. Presentation at Partners in Progress: Cancer Patient Advocates and FDA Public Workshop.

Board, O. (2009). Competition and disclosure. *Journal of Industrial Economics*, **57** (1), 197–213.

Bryant, H. E., Schultz, N., Thomas, H. D., Parker, K. M., Flower, D., Lopez, E., Kyle, S., Meuth, M., Curtin, N. J. and Helleday, T. (2005). Specific killing of brca2-deficient tumors with inhibitors of poly(adp-ribose) polymerase. *Nature*, **434**, 913–917.

Budish, E., Roin, B. and Williams, H. (2015). Do firms underinvest in long-term research? evidence from cancer clinical trials. *American Economic Review*, **105** (7), 2044–2085.

Bujar, M., McAuslane, N., Walker, S. R. and Salek, S. (2017). Evaluating quality of decision-making processes in medicines' development, regulatory review, and health technology assessment: A systematic review of the literature. *Frontiers in Pharmacology*, **8**, 189.

Burbidge, J. B., Magee, L. and Robb, A. L. (1988). Alternative transformations to handle extreme values of the dependent variable. *Journal of the American Statistical Association*, **83** (401), 123–127.

Burke, L. G., Frakt, A. B., Khullar, D., Orav, E. J. and Jha, A. K. (2017). Association between teaching status and mortality in u.s. hospitals. *JAMA*, **317** (2), 2105–2113.

Campbell, J. D., Alexandrov, A., Kim, J., Wala, J., Berger, A. H., Pedamallu, C. S., Shukla, S. A., Guo, G., Brooks, A. N. and Meyerson, M. (2016). Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nature Genetics*, **48**, 607–616.

Centers for Disease Control and Prevention (2018). Deaths: Leading causes for 2016. In *National Vita Statistics Reports*, vol. 67.

Chandra, A., Finkelstein, A. F., Sacarny, A. S. and Syverson, C. (2016). Health care exceptionalism? performance and allocation in the us health care sector. *American Economic Review*, **106** (8), 2110–2144.

—, Garthwaithe, C. and Stern, A. D. (2018). Characterizing the drug development pipeline for precision medicines, nBER Working Paper No. 24026.

Cockburn, I. (2007). Is the pharmaceutical industry in productivity crisis? In J. Lerner and S. Stern (eds.), *Innovation Policy and the Economy, Volume 7*, Cambridge: MIT Press, pp. 1–32.

— and Henderson, R. (1997). Public-private interaction and the productivity of pharmaceutical research, nBER Working Paper No. 6018.

Cohen, W. M. and Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administration Science Quarterly*, **35** (1), 128–152.

Collins, F. S. (2011). Mining for therapeutic gold. *Nature Reviews Drug Discovery*, **10** (6), 397.

Conti, R. M., Bernstein, A. C. B., Villaflor, V. M., Schilsky, R. L., Rosenthal, M. B. and Bach, P. B. (2013). Prevalence of off-label use and spending in 2010 among patent-protected chemotherapies in a population-based cohort of medical oncologists. *Journal of Clinical Oncology*, **31** (9), 1134–1139.

Cook, D., Brown, D., Alexander, R., March, R., Morgan, P., Satterwaite, G. and Pangalos, M. N. (2014). Lessons learned from the fate of astrazeneca's drug pipeline: a five-dimensional framework. *Nature Reviews Drug Discovery*, **13** (6), 419–431.

Cromwell, J., Adamache, W. and Drozd, E. (2006). Bba impacts on hospital residents, finances, and medicare subsidies. *Health Care Financing Review*, **28** (1), 117–129.

Cutler, D. M. (1998). Cost shifting or cost cutting? the incidence of reductions in medicare payments. *Tax Policy and the Economy*, **12**, 1–27.

— and McClellan, M. (2001). Is technological change in medicine worth it. *Nature Reviews Drug Discovery*, **20** (5), 11–29.

Danzon, P. M. and Keuffel, E. L. (2014). Regulation of the pharmaceutical-biotechnology industry. In N. L. Rose (ed.), *Economic Regulation and Its Reform: What Have We Learned?*, Chicago: University of Chicago Press, pp. 407–484.

Dasgupta, P. and David, P. A. (1994). Towards a new economics of science. *Research Policy*, **23** (5), 487–521.

David, G., Lindrooth, R. C., Helmchen, L. A. and Burns, L. R. (2014). Do hospitals cross-subsidize? *Journal of Health Economics*, **37**, 198–218.

David, P. A., Hall, B. H. and Toole, A. A. (2000). Is public r&d a complement or substitute for private r&d? a review of the econometric evidence. *Research Policy*, **29** (4-5), 497–529.

Dees, N. D., Zhang, Q., Kandoth, C., Wendl, M. C., Schierding, W., Koboldt, D. C., Mooney, T. B., Callaway, M. B., Dooling, D. and Mardis, E. R. (2012). Music: Identifying mutational significance in cancer genomes. *Genome Research*, **22** (8), 1589–1598.

Delong, J. B. D. and Lang, K. (1992). Are all economic hypotheses false. *Journal of Political Economy*, **100**, 1257–1272.

DeMonaco, H., Ayfer, A. and von Hippel, E. (2012). The major role of clinicians in the discovery of off-label drug therapies. *Pharmacotherapy*, **26**, 323–332.

Department of Health and Human Services (2016a). Clinical trials registration and results information submission; final rule. 42 CFR Part 11.

Department of Health and Human Services (2016b). Investigational new drug application. 21 CFR Part 312.

Dhani, N., Tu, D., Sargent, D. J., Seymour, L. and Moore, M. J. (2009). Alternate endpoints for screening phase ii studies. *Clinical Cancer Research*, **15** (6), 1873–1882.

Dickler, R. and Shaw, G. (2000). The balanced budget act of 1997: Its impact on u.s. teaching hospitals. *Annals of Internal Medicine*, **132**, 820–824.

DiMasi, J. A. (2001). New drug development in the united states from 1963 to 1999. *Clinical Pharmacology and Therapeutics*, **69** (5), 286–296.

— (2013). Innovating by developing new uses of already-approved drugs: Trends in the marketing approval of supplemental indications. *Clinical Therapeutics*, **35** (6), 808–818.

—, Feldman, L., Seckler, A. and Wilson, A. (2010). Trends in risks associated with new drug development: Success rates for investigational drugs. *Clinical Pharmacology and Therapeutics*, **87** (3), 272–277.

—, Hansen, R. W. and Grabowski, H. G. (2003). The price of innovation: New estimates of drug development costs. *Journal of Health Economics*, **22** (2), 151–185.

Donelan, R., Walker, S. and Salek, S. (2015). Factors influencing quality decision-making: Regulatory and pharmaceutical industry perspectives. *Pharmacoepidemiology and Drug Safety*, **24**, 319–328.

Dougherty, B. A., Lai, Z., Hodgson, D. R., Orr, M. C., Hawryluk, M., Sun, J., Yelensky, R., Spencer, S. K., Robertson, J. D. and Barrett, J. C. (2012). Biological and clinical evidence for somatic mutations in brca1 and brca2 as predictive markers for olaparib response in high-grade serous ovarian cancers in the maintenance setting. *Genome Research*, **22** (8), 1589–1598.

Dranove, D. and Jin, G. Z. (2010). Quality disclosure and certification: Theory and practice. *Journal of Economic Literature*, **48** (4), 953–963.

Dubois, P., de Mouzon, O., Scott-Morton, F. and Seabright, P. (2015). Market size and pharmaceutical innovation. *The RAND Journal of Economics*, **46** (4), 844–871.

Dulbecco, R. (1986). A turning point in cancer research: Sequencing the human genome. *Science*, **231** (4742), 1055–1056.

Eisenberg, R. S. (2005). The problem of new uses. *Yale Journal of Health Policy, Law, and Ethics*, **5** (2), 717–740.

Farmer, H., McCabe, N., Lord, C. J., Tutt, A. N., Johnson, D. A., Richardson, T. B., Santarosa, M., Dillon, K. J., Hickson, I. and Ashworth, A. (2005). Targeting the dna repair defect in brca mutant cells as a therapeutic strategy. *Nature*, **434**, 917–921.

Fishman, L. E. (1992). Medicare payment issues of interest to teaching hospitals.

Fleming, L. and Sorenson, O. (2003). Navigating the technological landscape of innovation. *MIT Sloan Management Review*, **44** (2), 15–23.

— and — (2004). Science as a map in technological search. *Strategic Management Journal*, **25**, 909–928.

Fleming, T. R. (2005). Surrogate endpoints and fda's accelerated approval process. *Health Affairs*, **24** (1), 67–78.

— and Powers, J. H. (2012). Biomarkers and surrogate endpoints in clinical trials. *Statistics in Medicine*, **31** (25), 2973–2984.

Fontanarosa, P. B. and DeAngelis, C. D. (2002). Basic science and translational research. *JAMA*, **132**, 820–824.

Food and Drug Administration Act of 2007 (2007). Public Law No. 110-185 §801.

Furman, J. L., Murray, F. and Stern, S. (2012). Growing stem cells: the impact of federal funding policy on the u.s. scientific frontier. *Journal of Policy Analysis and Management*, **31** (3), 661–705.

— and Stern, S. (2011). Climbing atop the shoulders of giants: The impact of institutions on cumulative research. *The American Economic Review*, **101** (5), 1933–1963.

Futureal, P. A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N. and Stratton, M. R. (2004). A census of human cancer genes. *Nature Reviews Cancer*, **4** (3), 177–183.

Gaynor, M. (2011). What do we know about competition and quality in health care markets? *Foundations and Trends in Microeconomics*, **2**, 441–508.

—, Moreno-Serra, R. and Propper, C. (2013). Death by market power: Reform, competition, and patient outcomes in the national health service. *American Economic Journal: Economic Policy*, **5**, 134–166.

Gelijns, A. and Rosenberg, N. (1995). The changing nature of medical technological development. In N. Rosenberg, A. Gelijns and H. Dawkins (eds.), *Sources of Medical Technology: University and Industry*, Washington, D.C.: National Academies Press, pp. 3–14.

GeneEd: Genetics, Education, Discovery (2018). Dna, genes, chromosomes. `https://geneed.nlm.nih.gov/topic_subtopic.php?tid=15`, Last Accessed on 2018-09-07.

Gershon, E. S. (1998). Making progress: Does clinical research lead to breakthroughs in basic biomedical sciences. *Academic Medicine*, **42**, 95–102.

Giroud, X. and Mueller, H. M. (2017). Firms? internal networks and local economic shocks. NBER Working Paper 23176.

Gittelman, M. and Kogut, B. (2003). Does good science lead to valuable knowledge? biotechnology firms and the evolutionary logic of citation patterns. *Management Science*, **49** (4), 366–382.

Grossman, S. A., Schreck, K. C., Ballman, K. and Alexander, B. (2017). Point/counterpoint: Randomized versus single arm phase ii trials for patients with newly diagnosed glioblastoma. *Neuro-Oncology*, **19** (4), 469–474.

Guedj, I. and Scharfstein, D. (2004). Organizational scope and investment: Evidence from drug development strategies and performance of biopharmaceutical firms.

Hall, B. and Van Reenen, J. (2000). How effective are fiscal incentives for r&d? a review of the evidence. *Research Policy*, **29**, 449–469.

Hall, B. H., Jaffe, A. and Trajtenberg, M. (2005). Market value and patent citations. *The RAND Journal of Economics*, **36** (1), 16–38.

Harris, G. (2010). Diabetes drug maker hid test data, files indicate. *The New York Times*, **344**, d7373.

Hay, M., Thomas, D. W., Craighead, J. L., Economides, C. and Rosenthal, J. (2014). Clinical development success rates for investigational drugs. *Nature Biotechnology*, **32** (1), 40–51.

Hellerstein, J. (1998). Public funds, private funds, and medical innovation: How managed care affects public funds for clinical research. *The American Economic Review*, **88** (2), 112–116.

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

Howard, D. H., Bach, P. B., Berndt, E. and Conti, R. (2015). Pricing in the market for anticancer drugs. *Journal of Economic Perspectives*, **29** (1), 139–162.

Iglehart, J. K. (1999). Support for academic medical centers: Revisiting the 1997 balanced budget act. *New England Journal of Medicine*, **341**, 299–304.

Ioannidis, J., Stanley, T. and Doucouliagos, H. (2017). The power of bias in economics research. *The Economic Journal*, **127**, F236–F265.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, **2** (8), e124.

IQIVIA Institute (2018). Global oncology trends 2018.

Jacobson, M., Earle, C. C. E., Price, M. P. and Newhouse, J. P. (2010). How medicare?s payment cuts for cancer chemotherapy drugs changed patterns of treatment. *Health Affairs*, **29**, 1391–1399.

Jayaraj, S. (2018). Scientific maps and innovation: Impact of the human genome on drug discovery, doctoral dissertation, Rutgers University.

Jin, G. Z. (2005). Competition and disclosure incentives: An empirical study of hmos. *RAND Journal of Economics*, **36** (1), 93–112.

Jones, B. (2009). The burden of knowledge and the "death of the renaissance man": Is innovation getting harder? *Review of Economic Studies*, **76**, 283–317.

Jones, R. and Sanderson, S. (1996). Clinical revenues used to support the academic mission of medical schools, 1992-93. *Academic Medicine*, **71**, 300–307.

Jovanovic, B. (1982). Truthful disclosure of information. *Bell Journal of Economics*, **13**, 36–44.

Kaestner, R. and Guardado, J. (2008). Medicare reimbursement, nurse staffing and patient outcomes. *Journal of Health Economics*, **27**, 339–361.

Kao, J. (2019). Charted territory: Evidence from mapping the cancer genome and r&d decisions in the pharmaceutical industry.

Keeler, E. B., Carter, G. M. C. and Trude, S. (1998). Insurance aspect of drg outlier payments. *Journal of Health Economics*, **7**, 193–214.

Kemp, R. and Prasad, V. (2017). Surrogate endpoints in oncology: When are they acceptable for regulatory and clinical decisions, and are they currently overused. *BMC Medicine*, **15** (1), 134.

Kesselheim, A. and Mello, M. (2007). Confidentiality laws and secrecy in medical research: Improving public access to data on drug safety. *Health Affairs*, **26** (2).

Kolata, G. (2013). Cancers share gene patterns, studies affirm. *New York Times*, may 1. https://www.nytimes.com/2013/05/02/health/dna-research-points-to-new-insight-into-cancers.html.

Korn, D., Rich, R. R., Garrison, H. H., Golub, S. H., Hendrix, M. J., Heinig, S. J., Masters, B. S. and Turman, R. (2002). Science policy : The nih budget in the 'postdoubling' era. *Science*, **296**, 1401–1402.

Krieger, J. L. (2017). Trials and terminations: Learning from competitors' r&d failures.

Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Sivanchenko, A., Carter, S. L., Stewart, C., Mermel, C. H. and Roberts, S. A. (2014). Mutational heterogeneity in cancer and the search for new cancer genes. *Nature*, **499** (7457), 214–218.

Lendrem, D. W., Lendrem, C., Peck, R. W., Senn, S. J., Day, S. and Isaacs, J. D. (2015). Progression-seeking bias and rational optimism in research and development. *Nature Reviews Drug Discovery*, **14**, 219–221.

Lexchin, J., Bero, L., Djulbegovic, B. and Clark, O. (2008). Pharmaceutical industry sponsorship and research outcome and quality: Systematic review. *BMJ*, **326**, 1167–1170.

Lijima, M., Banno, K., Okawa, R., Yanokura, M., Lida, M., Takeda, T., Kunitomi-Irie, H., Adachi, M., Nakamura, K. and Aoki, D. (2017). Genome-wide analysis of gynecologic cancer: The cancer genome atlas in ovarian and endometrial cancer. *Oncology Letters*, **13** (3), 1063–1070.

Lindrooth, R. C., Bazzoli, G. B., Needleman, J. and Hasnain-Wynia, R. (2005a). The effect of changes in hospital reimbursement on staffing decisions at safety net and non-safety net hospitals. *Health Services Research*, **41**, 701–720.

—, Bazzoli, G. J. and Clement, J. (2005b). The effect of reimbursement on the intensity of hospital services. *Southern Economic Journal*, **73** (3), 577–587.

Mardis, E. R. (2018). Insights from large-scale cancer genome sequencing. *Annual Reviews of Cancer Biology*, **2**, 429–444.

Mathieu, S., Boutron, I., Moher, D., Altman, D. and Ravaud, P. (2009). Comparison of registered and published primary outcomes in randomized controlled trials. *JAMA*, **302** (9).

Matsa, D. A. (2011). Competition and product quality in the supermarket industry. *Quarterly Journal of Economics*, **126**, 1539–1591.

McGuire, T. G. and Pauly, M. V. (1991). Physician response to fee changes with multiple payers. *Journal of Health Economics*, **10** (4), 385–410.

Medicare Payment Advisory Commission (2003). Medicare payment policy.

Milgrom, P. R. (1981). Good news and bad news: Representation theorems and applications. *Bell Journal of Economics*, **12**, 380–391.

Mokyr, J. (2006). The market for ideas and the origins of economic growth in the eighteenth century europe, working Paper.

Molitor, D. and Agha, L. (2012). Technology adoption under uncertainty: Off-label prescribing in cancer care, dissertation Chapter.

Morgan, P., Brown, D. G., Lennard, S., Anderton, M. J., Barrett, J. C., Eriksson, U., Fidock, M., Hamren, B., Johnson, A. and Pangalos, M. (2018). Impact of a five-dimensional framework on r&d productivity at astrazeneca. *Nature Reviews Drug Discovery*, **17** (3), 167–181.

Mostaghim, S., Gagne, J. and Kesselheim, A. (2017). Safety related label changes for new drugs after approval in the us through expedited regulatory pathways: Retrospective cohort study. *BMJ*, **358**, j3837.

Murray, F., Aghion, P., Detratripont, M., Kolev, J. and Stern, S. (2016). Of mice and academics: Examining the effect of openness on innovation. *American Economic Journal: Economic Policy*, **8** (1), 212–252.

Nagaraj, A. (2017). The private impact of public maps: Landsat satellite imagery and gold exploration.

NCI Center for Cancer Research (n.d.). Clinical trial design. `http://docplayer.net/15224109-Clinical-trial-design-sponsored-by-center-for-cancer-research-national-cance` `html`, Last Accessed on 2018-10-30.

Nelson, R. R. (1982). The role of knowledge in r&d efficiency. *The Quarterly Journal of Economics*, **97** (3), 453–470.

Olson, M. (2008). The risk we bear: the effects of review speed and industry user fees on new drug safety. *Journal of Health Economics*, **27** (2).

Peck, R. W., Lendrem, D. W., Grant, I., Lendrem, B. C. and Isaacs, J. D. (2015). Why is it hard to terminate failing projects in pharmaceutical r&d? *Nature Reviews Drug Discovery*, **14**, 663–664.

Pfister, D. G. (2012). Off-label use of oncology drugs: The need for more data and then some. *Journal of Clinical Oncology*, **30**, 584–586.

Prasad, V. and Berger, V. W. (2015). Hard-wired bias: How even double-blind, randomized controlled trials can be skewed from the start. *Mayo Clinic Proceedings*, **90** (9), 1171–1175.

—, Kim, C., Burotto, M. and Vandross, A. (2015). The strength of association between surrogate endpoints and survival in oncology. *JAMA Internal Medicine*, **175** (8), 1389–1398.

Prayle, A. P., Hurley, M. N. and Smythe, A. R. (2012). Compliance with mandatory reporting of clinical trial results on clinicaltrials.gov: Cross sectional study. *BMJ*, **344**.

Reinhardt, U. E. (2006). The pricing of u.s. hospital services: Chaos behind a veil of secrecy. *Health Affairs*, **25** (1), 57–69.

Robertson, A. G., Kim, J., Al-Ahmadie, H., Bellmunt, J., Guo, G., Cherniack, A. D., Hinoue, T., Laird, P. W., Hoadley, K. A. and Lerner, S. P. (2017). Comprehensive molecular characterization of muscle-invasive bladder cancer. *Cell*, **171** (3), 540–556.

Roin, B. (2014). Solving the problem of new uses. *Michigan State Law Review*.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, **98** (5), S71–S102.

Rosenberg, N. (1963). Technological chance in the machine tool industry, 1840-1910. *Journal of Economic History*, **23**, 414–416.

— (1979). Technological independence in the american economy. *Technology and Culture*, **20** (1), 25–50.

— (1990). Why do firms do basic research (with their own money)? *Research Policy*, **19** (2), 165–174.

— (1994). *Exploring the Black Box*. New York: Cambridge University Press.

— (2008). Some critical episodes in the progress of medical innovation: An anglo-american perspective. Stanford Institute for Economic Policy Discussion Paper.

Sampat, B. N. (2012). Mission-oriented biomedical research at the nih. *Research Policy*, **41** (10), 1729–1741.

Samuel, N. and Hudson, T. J. (2013). Translating genomics to the clinic: Implications of cancer heterogeneity. *Clinical Chemistry*, **59** (1), 127–137.

Scott Morton, F. and Kyle, M. (2011). Markets for pharmaceutical products. In M. V. Pauly, T. G. Mcguire and P. P. Barros (eds.), *Handbook of Health Economics, Volume 2*, Elsevier Science, pp. 763–823.

Seshamani, M., Schwartz, J. S. and Volpp, K. (2006). The effect of cuts in medicare reimbursement on hospital mortality. *Health Services Research*, **41** (3 Pt 1), 683–700.

Seymour, L., Ivy, S. P., Sargent, D., Spriggs, D., Baker, L., Rubinstein, L., Ratain, M. J., Le Blanc, M., Stewart, D. and Berry, D. (2010). The design of phase ii clinical trials testing cancer therapeutics: Consensus recommendations from the clinical trial design task force of the national cancer institute investigational drug steering committee. *Clinical Cancer Research*, **16** (6), 1764–1769.

Shaked, A. and Sutton, J. (1982). Relaxing price competition through price differentiation. *Review of Economic Studies*, **49** (1), 3–13.

Sharpe, P. and Keelin, J. (1998). How smithkline beecham makes better resource-allocation decisions. *Harvard Business Review*, **76** (2).

Shen, Y.-C. and Wu, V. Y. (2013). Reductions in medicare payments and patient outcomes: An analysis of 5 leading medicare conditions. *Medical Care*, **51** (11), 970–977.

Spetzler, C., Winter, H. and Meyer, J. (2016). *Decision Quality: Value Creation from Better Business Decisions*. Hoboken: John Wiley and Sons.

Steinbrook, R. (2004). Public registration of clinical trials. *New England Journal of Medicine*, **351**, 315–317.

Stratton, M. R., Campbell, P. J. and Futreal, P. A. (2009). The cancer genome. *Nature*, **458**, 719–724.

Tabakovic, H. and Wollmann, T. (2018). The impact of money on science: Evidence from unexpected ncaa football outcomes.

Task Force on Academic Medical Centers (). From bench to bedside: Preserving the research mission of academic medical health centers.

The Cancer Genome Atlas Research Network (2011a). Integrated genomic analyses of ovarian carcinoma. *Nature*, **474**, 609–615.

The Cancer Genome Atlas Research Network (2011b). News release: The cancer genome atlas completes detailed ovarian cancer analysis. `https://cancergenome.nih.gov/newsevents/newsannouncements/ovarianpaper`, Last Accessed on 2018-10-01.

The Cancer Genome Atlas Research Network (2012). Comprehensive genomic characterization of squamous cell lung cancers. *Nature*, **489**, 519–525.

The Cancer Genome Atlas Research Network (2018). What is cancer genomics? `https://cancergenome.nih.gov/cancergenomics/whatisgenomics/whatis`, Last Accessed on 2018-10-01.

Thomas, D. W., Burns, J., Audette, J., Carroll, A., Dow-Hygelund, C. and Hay, M. (2016). *Clinical Development Success Rates: 2006-2015*. Tech. rep., BIO Industry Analysis.

Trajtenberg, M., Henderson, R. and Jaffe, A. (1992). Ivory tower versus corporate lab: An empirical study of basic research and appropriability, nBER Working Paper No. 4146.

Tse, T., Williams, R. and Zarin, D. (2009). Reporting "basic results" in clinicaltrials.gov. *CHEST*, **2009**, 259–303.

Turner, E., Matthews, A., Linardatos, E., Tell, R. and Rosenthal, R. (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. *New England Journal of Medicine*, **358** (1), 252–260.

Tversky, A. and Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, **185** (4157), 1124–1131.

U.S. Food and Drug Administration (1998a). *Guidance for Industry: FDA Approval of New Cancer Treatment Uses for Marketed and Biological Products*. Tech. rep.

U.S. Food and Drug Administration (1998b). *Guidance for Industry: Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products*. Tech. rep.

U.S. Food and Drug Administration (2004). *Guidance for Industry: IND Exemptions for Studies of Lawfully Marketed Drug or Biological Products for the Treatment of Cancer*. Tech. rep.

U.S. Food and Drug Administration (2007). *Guidance for Industry: Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics*. Tech. rep.

Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz Jr., L. A. and Kinzler, K. W. (2013). Cancer genome landscapes. *Science*, **339** (6127), 1546–1558.

Volpp, K. G., Konetzka, R., Zhu, J., Parsons, L. and Peterson, E. (2005). Effect of cuts in medicare reimbursement on process and outcome of care for acute myocardial infarction patients? *Circulation*, **112** (15), 2268–2275.

Ward, M. and Dranove, D. (1995). The vertical chain of research and development in the pharmaceutical industry. *Economic Inquiry*, **33** (1), 70–87.

Weissman, J. S., Saglam, D., Campbell, E. G., Causino, N. and Blumenthal, D. (1999). Market forces and unsponsored research in academic medical centers. *JAMA*, **281** (12), 1093–1098.

Williams, H. (2013). Intellectual property rights and innovation: Evidence from the human genome. *Journal of Political Economy*, **121** (1), 1–27.

Woolf, S. H. (2008). The meaning of translational research and why it matters. *JAMA*, **299** (2), 211–213.

Wu, V. Y. and Shen, Y.-C. (2015). Long-term impact of medicare payment reductions on patient outcomes. *Health Services Research*, **49** (5), 1596–1615.

Wurtman, R. J. and Bettiker, R. L. (1995). The slowing of treatment discovery. *JAMA*, **1** (11), 1122–1125.

Yang, Y., Dong, X., Xie, B., Ding, N., Chen, J., Li, Y., Zhang, Q., Qu, H. and Fang, X. (2015). Databases and web tools for cancer genomics study. *Genomics Proteomics Bioinformatics*, **13**, 46–50.

Yip, W. C. (1998). Physician response to medicare fee reductions: Changes in the volume of cabg surgeries in the medicare and private sectors. *Journal of Health Economics*, **17** (6), 675–699.

Zarin, D. and Tse, T. (2008). Moving towards transparency of clinical trials. *Science*, **319** (5868), 1340–1342.

Zarin, D. A., Tse, T., Williams, R., Califf, R. and Ide, N. C. (2011). The clinicaltrials.gov results database–update and key issues. *New England Journal of Medicine*, **369** (9), 852–860.

Appendix A

Appendix to Chapter 1

A.1 Data Description

This appendix describes additional detail on the datasets used in this analysis.

A.1.1 Cancer Mapping Data

Mapping Studies

Cancer mapping data comes from cBioPortal for Cancer Genomes[1] and the Catalogue of Somatic Mutations in Cancer.[2] These two publicly available databases contain datasets from many published cancer mapping studies. I focus on the set of cancer mapping studies that are high impact and large (in terms of the number of tumors mapped).

To identify high impact cancer mapping studies, I isolate the list of cancer mapping studies published top genetics journals. A top genetics journal is one that is ranked highly under the Scimago Journal & Country Rank (SJR) system, a yearly ranking scheme that ranks journals using a citation-based algorithm.[3] The SJR measures a journal's influence by looking at the

[1]For more details, see `http://www.cbioportal.org/`

[2]For more details, see `https://cancer.sanger.ac.uk/cosmic`

[3]For more details, see: `https://www.scimagojr.com/`

number of citations received by a journal during the past three years (Gonzalez-Pereira et al., 2009). I define a genetics journal to be highly ranked if it is ranked in the top 25 of the "Genetics" SJR ranking at least once between 1999 (the earliest year the SJR rankings are publicly available) and 2004 (the last year in which a mapping study published in a particular journal cannot influence that same journal's ranking.[4]

A mapping study is defined as "large" if is published in cBioPortal, which focuses on 'large-scale cancer genomics projects" (Cerami et al., 2012) or in COSMIC's "Whole Genome & Large-scale Systematic Screens" sequencing study database.[5] This results in gene-cancer level mutation information from 168 high quality and large cancer mapping studies.

Mutation Data

There are two key facts to note about the mutations analyzed in this paper: first, I focus on mutations that occur in the protein-coding region of the DNA and are non-inherited: non-silent somatic mutations. Somatic mutations are DNA aberrations that occur after conception. According to Stratton *et al.* (2009), "all cancers arise as a result of somatically acquired changes in the DNA of cancer cells." Somatic mutations differ from germline mutations, which are inherited.[6] I exclude silent mutations which are mutations that occur in the non-protein coding region of the DNA. The final list of included mutations include: missense, nonsense, insertions, deletions, frameshift, nonstop extension.[7]

Second, this paper focuses primarily on mutations, but other types of genetic alterations may contribute to the progression and growth of cancer. These genetic alterations include: DNA rearrangements, where DNA is broken and then added to a DNA segment from another

[4]Results using journals ranked in the top 25 using the 2017 "Genetics" SJR ranking, the 1999 to 2004 "Medicine" SJR ranking, or the 2017 "Medicine" SJR ranking produce similar results.

[5]For more details, see `https://cancer.sanger.ac.uk/cosmic/papers`

[6]For more details, see: `https://www.cancer.gov/publications/dictionaries/cancer-terms/def/somatic-mutation`

[7]For more details, see `https://ghr.nlm.nih.gov/primer/mutationsanddisorders/possiblemutations`

part of the genome; deletions of small or large parts of the DNA; amplifications or excess copies of a gene. For more details, see Stratton *et al.* (2009) and Vogelstein *et al.* (2013).

A.1.2 Identifying Well-Designed and Non-Well-designed Trials

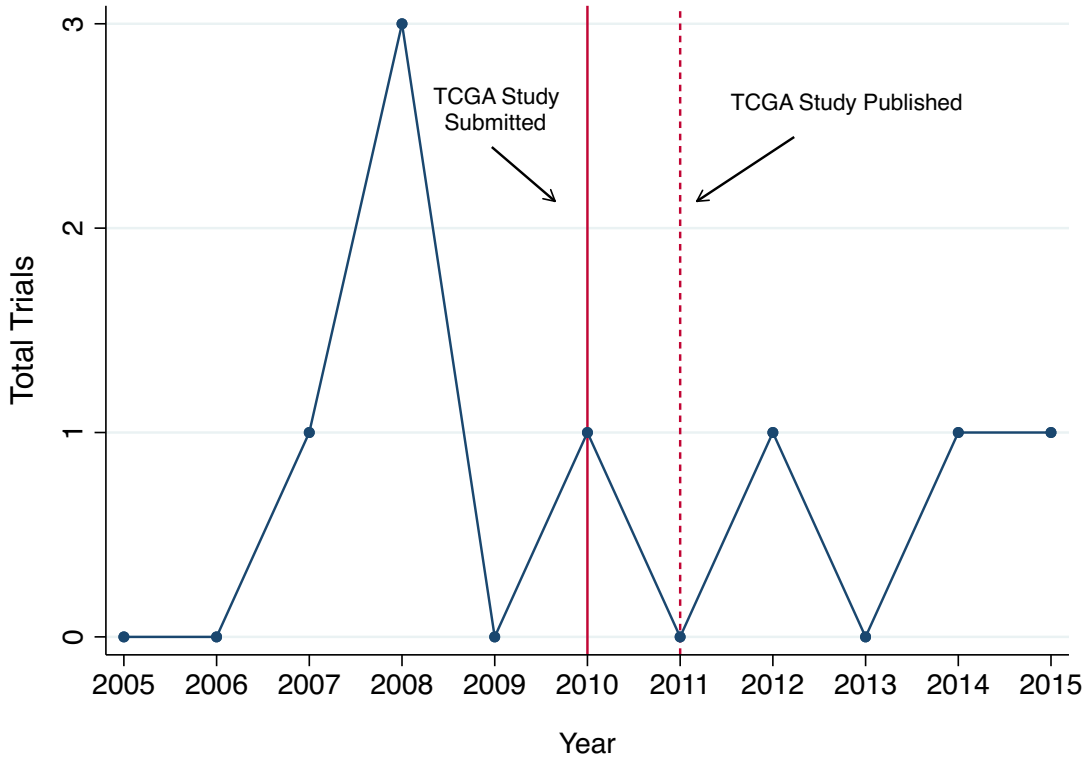
This section describes how clinical trials were classified into well-designed and non-well-designed trials. Using recommended standards outlined in the scientific literature (Adjei *et al.*, 2009; Berger and Alperson, 2009; Blumenthal, 2017; Dhani *et al.*, 2009; U.S. Food and Drug Administration, 2007; Grossman *et al.*, 2017; Kemp and Prasad, 2017; NCI Center for Cancer Research, n.d.; Prasad *et al.*, 2015; Seymour *et al.*, 2010), I classify phase II trials as well-designed if they satisfied one of the following three criteria:

1. Randomized, controlled, overall survival endpoint
2. Randomized, controlled, validated surrogated endpoint
3. Non-randomized, controlled, validated surrogate endpoint

Information on validated surrogate endpoints comes from Prasad *et al.* (2015). Trials that are not classified as well-designed are considered non-well-designed.

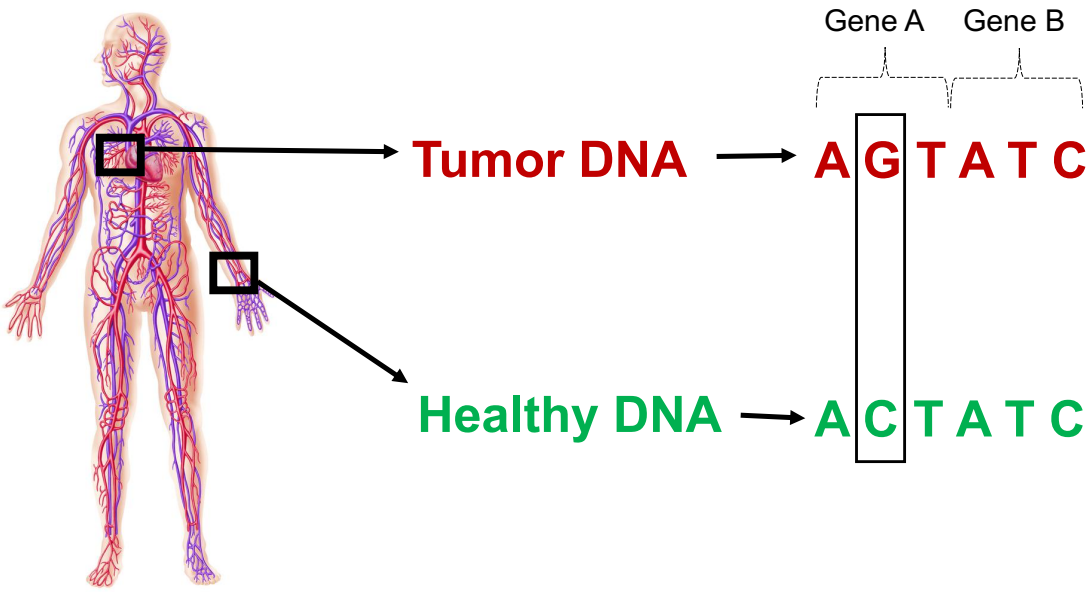
A.2 Additional Figures and Tables

Figure A.1: *Trials Enrolling Patients with Ovarian Cancer and BRCA2 Gene Mutations, Olaparib Only*



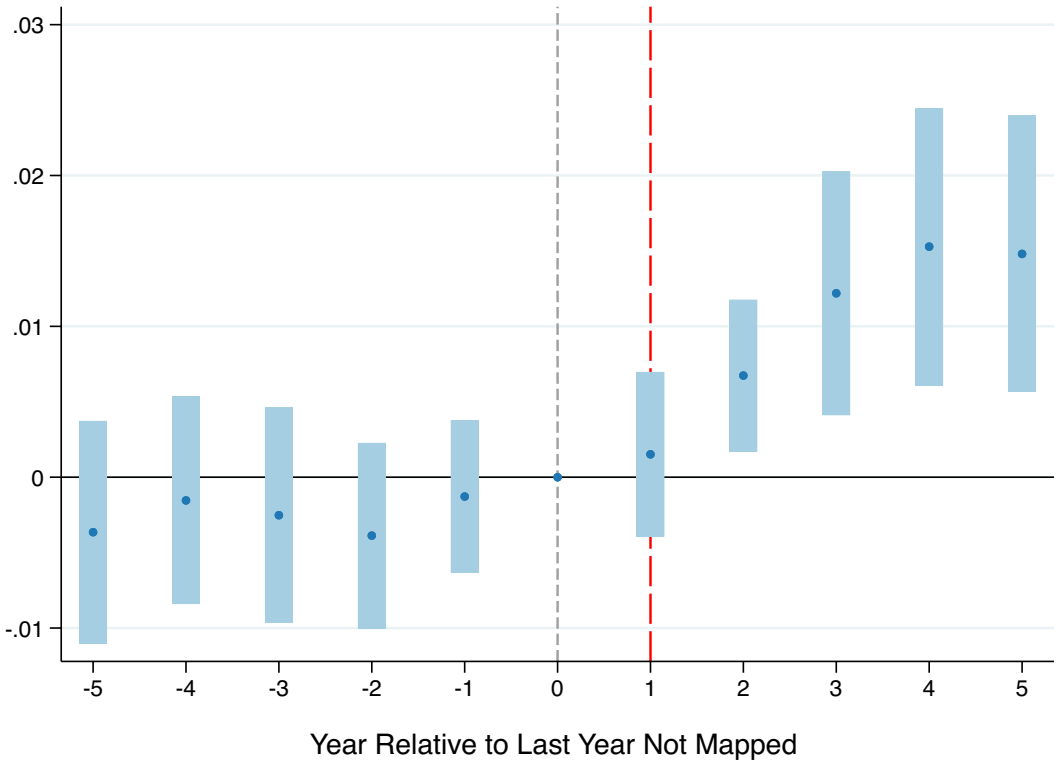
Notes: This figure shows the total number of clinical trials (privately-funded, phase II only) enrolling patients with BRCA2-mutated ovarian cancer and testing Olaparib in each year from 2005 to 2015. The vertical lines indicated the years in which the TCGA's ovarian cancer study (TCGA, 2011a) was submitted to (solid line) and published in (dashed line) the journal *Nature*.

Figure A.2: *Overview of Scientific Background on Cancer Genome Sequencing*



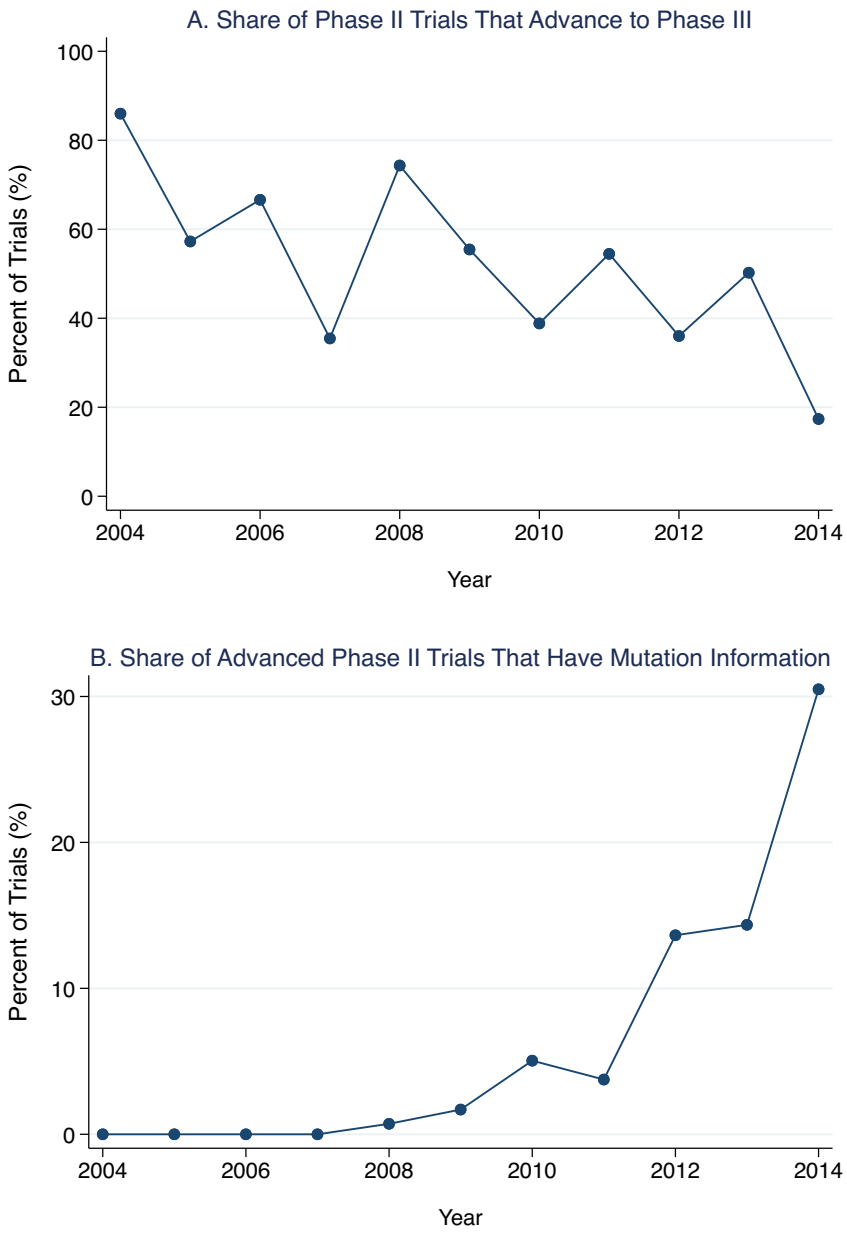
Notes: This figure graphically summarizes the scientific background described in Section 1.3.2. A individual's genome is all the DNA contained in a particular a cell. DNA is comprised of four bases : adenine (A), cytosine (C), guanine (G), and thymine (T). The unique sequence of these four DNA bases—A, C, G, and T—provides a "blueprint" for the human body (GeneEd: Genetics, Education, Discovery, 2018). A gene is a segment of DNA that provides instructions for unique traits. Cancer can be caused by a mutation, or a change in the sequence of DNA bases. Cancer genome researchers aim to identify the mutations that drive the development and growth of cancer by comparing the DNA sequences of cancer cells (in red) to those of normal tissue (in green). This figure is a modified version of Figure 1 found in Samuel and Hudson (2013).

Figure A.3: *Event Study Estimates—Impact of Mapping on Trial Quantity, Trials with Non-Missing Intervention*



Notes: This figure shows the relationship between cancer mapping and the subsequent quantity of clinical trials using the subset of trials with non-missing intervention data. The figure plots coefficients (and 95 percent confidence intervals) from the event study specification described in Equation 1.2. On the x-axis are years z relative to a "zero" relative year that marks the last year the gene-cancer was not known to be mutated based on the cancer mapping studies (i.e., year 1 marks the first year a mutation in a gene-cancer was publicly disclosed by a cancer mapping study). As in the specifications in Table 1.3, this specification is based on gene-cancer-year level observations, the coefficients are estimates from OLS models, the sample includes all gene-cancer-years (excluding gene-cancer pairs known in 2004) from 2004-2016, and the standard errors are robust and clustered at the gene and cancer level. Gene-cancer fixed effects, year fixed effects, and cancer-year linear time trends are included. All trials are privately-funded phase II trials. Cancer mapping studies are large and published in a top 25 Genetic journal, based on rankings between 1999-2004.

Figure A.4: *Trial Advancement Rates, by Year*



Notes: Panel A plots the percent of privately-funded phase II clinical trials that successfully advance to phase III. Panel B plots the percent of privately-funded phase II clinical trials that are initiated in gene-cancer pairs with mutation information, as a share of the total number of privately-funded phase II trials that successfully advance to phase III. In this figure, trials are classified as having successfully advanced to phase III if they transition to phase III within 4 years of the phase II trial start date. Sample includes all phase II trials that are completed or terminated as of July 14, 2017. Observations are at the trial-gene-cancer level.

Table A.1: *Impact of Mapping Information, Excluding Genes Affected by Patent Regulation*

	Dependent Variable: 1(Any Clinical Trials)		
	(1)	(2)	(3)
1(PostDisclGeneCancer)	0.00786**	0.00802**	0.00543**
	(0.00247)	(0.00291)	(0.00170)
Mean of Dep. Var.	0.017	0.017	0.017
Percent Gain	46.25%	47.17%	31.96%
Gene-cancer FEs	X	X	X
Year FEs	X	X	X
Cancer × Linear Year Trend	X		
Cancer × Year FEs		X	
Observations	642,018	642,018	642,018

Notes: In the 1990s, the firm Myriad received a patent on the sequenced BRCA1 and BRCA2 genes and associated mutations (Gold and Carbone, 2010). Concerned that such patent protection could limit the detection of such mutations, in 2013 the Supreme Court rule that genes and their mutations could not be patented. This table examines how the relationship between cancer mapping information and quantity of subsequent trials, excluding BRCA1 and BRCA2—genes that are most likely to be affected by changing intellectual property regulation. Gene-cancer-year level observations. All estimates are from OLS models. The sample includes all gene-cancer-years (excluding gene-cancer pairs known in 2004) from 2004-2016 (N = 644,046 gene-cancer-year observations). Controls include gene-cancer fixed effects, year fixed effects, and cancer-year linear time trends. Robust standard errors, clustered at the gene and cancer level, are shown in parenthesis. Cancer mapping studies are large and published in a top 25 Genetic journal, based on rankings between 1999-2004. Outcomes: 0/1 = 1 if a privately-funded phase II clinical trial is reported in a gene-cancer-year. PostDisclGeneCancer: 0/1 = 1 for the year after the mutation was disclosed. Mean of Dep. Var. is the mean number of trials in a gene-cancer before the first disclosure of a mutation. Percent Gain is calculated using the pre-mutation information trial mean.
*p<0.10, p**<0.05, ***p<0.01.

Table A.2: *Impact of Mapping Information, Excluding Genes Affected by Patent Regulation*

	Dependent Variable: 1(Any Clinical Trials)		
	(1)	(2)	(3)
1(PostDisclGeneCancer)	0.00786**	0.00802**	0.00543**
	(0.00247)	(0.00291)	(0.00170)
Mean of Dep. Var.	0.017	0.017	0.017
Percent Gain	46.25%	47.17%	31.96%
Gene-cancer FEs	X	X	X
Year FEs	X	X	X
Cancer × Linear Year Trend	X		
Cancer × Year FEs		X	
Observations	642,018	642,018	642,018

Notes: In the 1990s, the firm Myriad received a patent on the sequenced BRCA1 and BRCA2 genes and associated mutations (Gold and Carbone, 2010). Concerned that such patent protection could limit the detection of such mutations, in 2013 the Supreme Court rule that genes and their mutations could not be patented. This table examines how the relationship between cancer mapping information and quantity of subsequent trials, excluding BRCA1 and BRCA2—genes that are most likely to be affected by changing intellectual property regulation. Gene-cancer-year level observations. All estimates are from OLS models. The sample includes all gene-cancer-years (excluding gene-cancer pairs known in 2004) from 2004-2016 (N = 644,046 gene-cancer-year observations). Controls include gene-cancer fixed effects, year fixed effects, and cancer-year linear time trends. Robust standard errors, clustered at the gene and cancer level, are shown in parenthesis. Cancer mapping studies are large and published in a top 25 Genetic journal, based on rankings between 1999-2004. Outcomes: 0/1 = 1 if a privately-funded phase II clinical trial is reported in a gene-cancer-year. PostDisclGeneCancer: 0/1 = 1 for the year after the mutation was disclosed. Mean of Dep. Var. is the mean number of trials in a gene-cancer before the first disclosure of a mutation. Percent Gain is calculated using the pre-mutation information trial mean.
*p<0.10, p**<0.05, ***p<0.01.

Table A.3: *Impact of Mapping Information in Same Gene, Related Cancer*

	Dependent Variable: 1(Any Clinical Trials)	
	(1)	(2)
1(PostDisclGeneCancer)		0.00817***
		(0.00230)
1(PostDisclGeneCancerGroup)	0.00502**	0.00129
	(0.00233)	(0.00197)
Mean of Dep. Var.		
Same Gene-Same Cancer	–	0.017
Same Gene-Related Cancer	0.014	0.014
Percent Gain		
PostDisclGeneCancer	–	57.16%
PostDisclGeneCancerGroup	35.10%	9.03%
Gene-cancer FEs	X	X
Year FEs	X	X
Cancer × Linear Year Trend	X	X
Observations	644,046	644,046

Notes: This table examines how the relationship between cancer mapping information and quantity of subsequent trials varies across mapping information with differing levels of clinical relevance. Gene-cancer-year level observations. All estimates are from OLS models. The sample includes all gene-cancer-years (excluding gene-cancer pairs known in 2004) from 2004-2016 (N = 644,046 gene-cancer-year observations). Controls include gene-cancer fixed effects, year fixed effects, and cancer-year linear time trends. Robust standard errors, clustered at the gene and cancer level, are shown in parenthesis. Cancer mapping studies are large studies and published in a top 25 Genetic journal, based on rankings between 1999-2004. Outcomes: 0/1 = 1 if a privately-funded phase II clinical trial is reported in a gene-cancer-year. PostDisclGeneCancer: 0/1 = 1 for the year after the mutation in a same gene-same cancer was disclosed. PostDisclGeneCancerGroup: 0/1 = 1 for the year after the mutation in a same gene-related cancer was disclosed. Cancers are classified as related if they are in the same cancer site group, based on the Surveillance, Epidemiology, and End Results (SEER) classification. Mean of Dep. Var. is the mean number of trials in a gene-cancer before the first disclosure of a mutation in the same-gene, same-cancer or the same-gene, related-cancer. The same-gene, same-cancer trial mean is used to calculate percentage gain for PostDisclGeneCancer. The same-gene, related-cancer trial mean is used to calculate the percent gain for PostDisclGeneCancerGroup.
*p<0.10, p**<0.05, ***p<0.01.

Table A.4: *Impact of Mapping Information for Different Types of Firms*

	Dependent Variable: 1(Any Clinical Trials)	
	Large Firms	Small Firms
	(1)	(2)
1(PostDisclGeneCancer)	0.00376	0.00706***
	(0.00235)	(0.00191)
Mean of Dep. Var.	0.0120	0.0079
Percent Gain	31.24%	89.00%
Gene-cancer FEs	X	X
Year FEs	X	X
Cancer × Linear Year Trend	X	X
Observations	644,046	644,046
Test for Diff. in Percent Gain		
P[($\beta_1 = \beta_2$)] =	0.00	

Notes: This table examines how the relationship between mapping information and quantity of subsequent trials for different types of firms. Gene-cancer-year level observations. All estimates are from seemingly unrelated models. The sample includes all gene-cancer-years (excluding gene-cancer pairs known in 2004) from 2004-2016 (N = 644,046 gene-cancer-year observations). Controls include gene-cancer fixed effects, year fixed effects, and cancer-year linear time trends. Seemingly unrelated models do not permit two-way clustering, so standard errors are clustered at the most conservative (gene) level and are shown in the parenthesis. Cancer mapping studies are large studies and published in a top 25 Genetic journal, based on rankings between 1999-2004. Mean of Dep. Var. is the mean number of trials in a gene-cancer before the first disclosure of a mutation. Percent Gain is calculated using the pre-mutation information trial mean. Outcomes: 0/1 = 1 if a phase II clinical trial conducted by a large firm (column 1) or small firm (column 2) is reported in a gene-cancer-year. Large firms are those with more than 100 patents prior to 2004. All remaining firms are classified as small firms. PostDisclGeneCancer: 0/1 = 1 for the year after the mutation was disclosed.
p<0.10, p**<0.05, ***p<0.01.

Table A.5: *Impact of Mapping Information for Different Types of Diseases*

	Dependent Variable: 1(Any Clinical Trials)			
	Pre-2004 Clinical Trials		Pre-2004 Market Size	
	≤ Median	> Median	≤ Median	> Median
	(1)	(2)	(3)	(4)
PostDisclGeneCancer	0.00675***	0.0873***	0.00828***	0.00885***
	(0.000786)	(0.0165)	(0.00123)	(0.00115)
Mean of Dep. Var.	0.012	0.276	0.015	0.018
Percent Gain	56.27%	31.63%	55.20%	49.16%
Gene-cancer FEs	X	X	X	X
Year FEs	X	X	X	X
Cancer × Linear Year Trend	X	X	X	X
Observations	644046	644046	644046	644046
Tests for Percent Gain				
P[($\beta_{below} = \beta_{above}$)]	0.01		0.56	

Notes: This table examines how the relationship between mapping information and quantity of subsequent trials varies across types of disease. Gene-cancer-year level observations. All estimates are from seemingly unrelated models. The sample includes all gene-cancer-years (excluding gene-cancer pairs known in 2004) from 2004-2016 (N = 644,046 gene-cancer-year observations). Controls include gene-cancer fixed effects, year fixed effects, and cancer-year linear time trends. Robust standard errors, clustered at the gene-cancer level, are shown in parenthesis. Cancer mapping studies are large studies and published in a top 25 Genetic journal, based on rankings between 1999-2004. Mean of Dep. Var. is the mean number of trials in a gene-cancer before the first disclosure of a mutation. Percent Gain is calculated using the pre-mutation information trial mean. Outcomes: 0/1 = 1 if a privately-funded phase II clinical trial is reported in a gene-cancer-year. PostDisclGeneCancer: 0/1 = 1 for the year after the mutation was disclosed. Columns 1 and 2 shows how the effect of mapping information varies across gene-cancers with low and high levels of clinical trial investment (calculated based on pre-2004 clinical trial levels). Columns 3 and 4 shows how the effect of mapping information varies across gene-cancers with low and high levels of market size (measured at the cancer level, calculated based on pre-2004 diagnoses levels). P-values are from 2-sided t-tests. *p<0.10, p**<0.05, ***p<0.01.

Table A.6: *Impact of Mapping Information for Different Trial Design Types*

	Dependent Variable: 1(Any Clinical Trials)	
	Well-Designed	Non-Well-Designed
	(1)	(2)
1(PostDisclGeneCancer)	0.00226***	0.00692***
	(0.000410)	(0.000835)
Mean of Dep. Var.	0.004	0.015
Percent Gain	63.63%	46.04%
Gene-cancer FEs	X	X
Year FEs	X	X
Cancer × Linear Year Trend	X	X
Observations	644,046	644,046
Test for Diff. in Percent Gain		
P[($\beta_1 = \beta_2$)] =	0.14	

Notes: This table examines how the relationship between mapping information and quantity of subsequent trials varies across well-designed and non-well-designed trials. Gene-cancer-year level observations. All estimates are from seemingly unrelated models. The sample includes all gene-cancer-years (excluding gene-cancer pairs known in 2004) from 2004-2016 (N = 644,046 gene-cancer-year observations). Controls include gene-cancer fixed effects, year fixed effects, and cancer-year linear time trends. Robust standard errors, clustered at the gene-cancer level, are shown in parenthesis. Cancer mapping studies are large studies and published in a top 25 Genetic journal, based on rankings between 1999-2004. Mean of Dep. Var. is the mean number of trials in a gene-cancer before the first disclosure of a mutation. Percent Gain is calculated using the pre-mutation information trial mean. Outcomes: 0/1 = 1 if a phase II clinical trial is reported in a gene-cancer-year and is a well-designed trial (column 1) or non-well-designed trial (column 2). See Appendix A.1.2 for a description of how well-designed and non-well-designed trials are identified. PostDisclGeneCancer: 0/1 = 1 for the year after the mutation was disclosed.
p<0.10, p**<0.05, ***p<0.01.

Table A.7: *Phase II Outcomes and Phase II to Phase III Transitions*

	Dependent Variable: Time to Phase III	
	(1)	(2)
Phase II Outcome (IHS(Response Rate))	0.245***	0.232***
	(0.0813)	(0.0831)
Firm Experience (IHS(# Clinical Trials))		-0.198
		(0.137)
Percent Change	27.784	26.169
Linear Year Trend	X	X
Nb. Observations	2,354	2,354
Nb. Trials	164	164
Nb. Genes	92	92
Nb. Cancers	80	80
Log Likelihood	-3455	-3443

Notes: This table shows the relationship between phase II outcomes and phase II transition rates. Trial-gene-cancer level observations. Estimates are from Cox proportional hazard models, stratified by cancer and large firm status. The sample includes all phase II trial-gene-cancers that are between from 2004-2016 (excluding gene-cancer pairs known in 2004), have available clinical outcomes data, and are completed or terminated as of July 14, 2017. Controls include a linear year time trend. Robust standard errors, clustered at the gene and cancer level, are shown in parenthesis. Cancer mapping studies are large studies and published in a top 25 Genetic journal, based on rankings between 1999-2004. DisclGeneCancer: 0/1 = 1 for whether a driver (clinically-relevant) mutation information was disclosed for the gene-cancer by the start of the clinical trial. FirmExperience is the inverse hyperbolic sine of the total number of clinical trials the trial sponsor has conducted in the focal cancer, one month prior to the trial start date. Response Rate refers to the trial's objective response rate, or the share of patients who respond to treatment.
*p<0.10, p**<0.05, ***p<0.01.

Appendix B

Appendix to Chapter 2

B.1 Identifying Clinical Trials

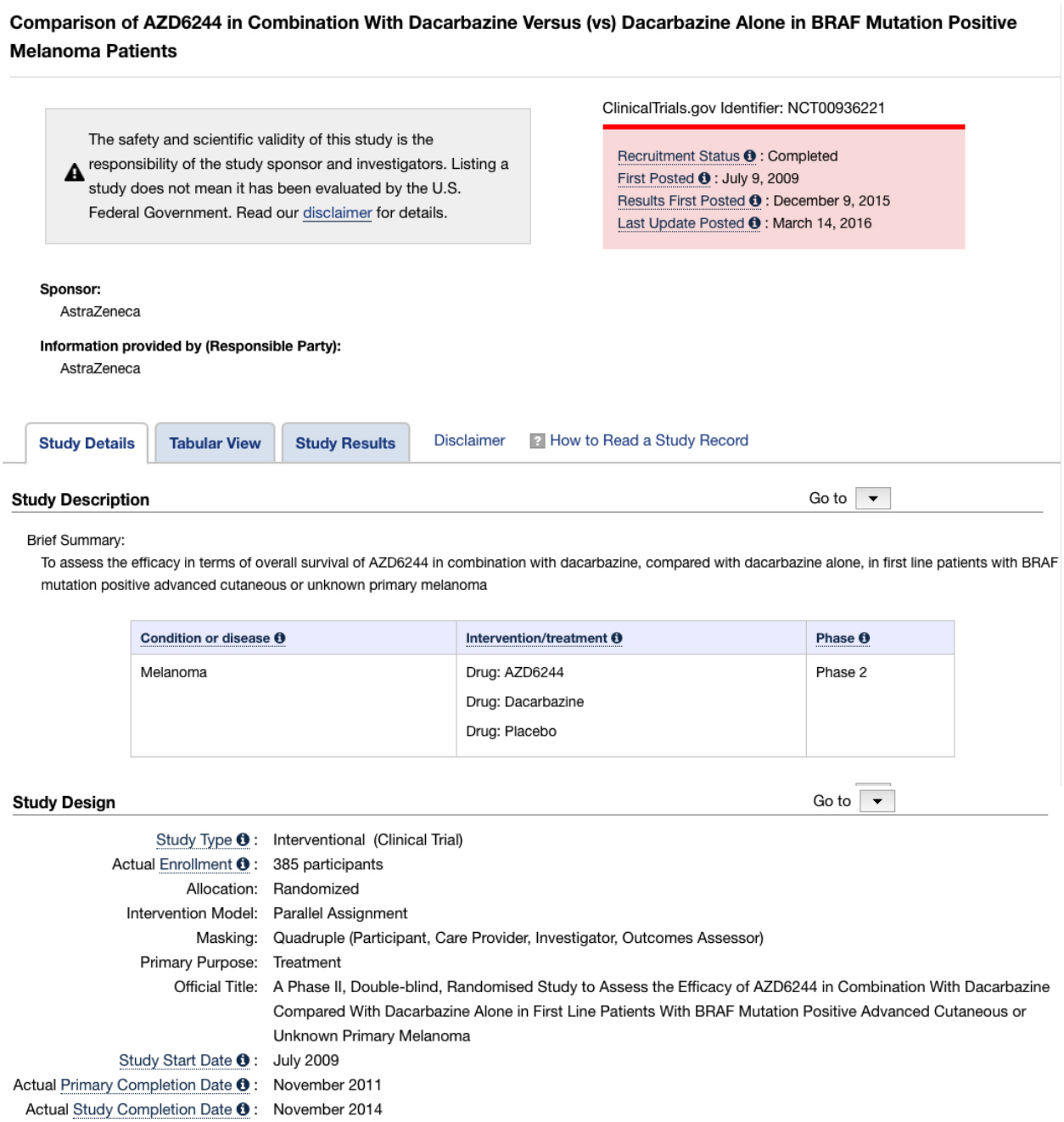
This section describes how the analytic trial sample is constructed. To identify the set of trials whose results are most important for influencing clinical decision making, I restrict the set of trials (called "Applicable Clinical Trials") to those whose results are required by the FDAAA. Figure B.1 demonstrates how I adapt the algorithm outlined in Anderson *et al.* (2015) to identify the set of "Highly Likely Applicable Clinical Trials."

Table B.1: *Identifying Analytic Trial Sample*

Exclusion Criterion	**Number Trials Remaining**
Studies downloaded from ClinicalTrials.gov (7/14/2018)	277,765
Studies matched to trials in Cortellis Competitive Intelligence (which provides standardized drug, firm names)	153,708
Exclude: Overall recruitment status == WITHDRAWN	149,307
Exclude: Primary completion date $\leq$ 12/2007, or if missing, completion date $\leq$ 12/2007	129,092
Exclude: Study type not INTERVENTIONAL	110,970
Exclude: Phase 1	92,561
Exclude: No US FDA oversight or only non-US sites	47,075
Exclude: Overall recruitment status not COMPLETED or TERMINATED	35,727
Exclude: Primary completion date $\geq$ 7/2017, or if missing, completion date $\geq$ 7/2017	33924
Exclude: Primary completion and completion dates missing, and verification date $\geq$ 7/2017	33,117
Exclude: Non-cancer trials	7,812
Exclude: Trials with start date $<$ 1/2000 or start date $>$ 12/2016	7,583
Exclude: Intervention not Drug or Biologic	7,008
Exclude: Missing phase or funder type	6,692 trials **31,125 trial-cancer observations**

B.2 Example ClinicalTrials.gov Trial Record

Figure B.1: *Example Clinical Trial Record*



Comparison of AZD6244 in Combination With Dacarbazine Versus (vs) Dacarbazine Alone in BRAF Mutation Positive Melanoma Patients

The safety and scientific validity of this study is the responsibility of the study sponsor and investigators. Listing a study does not mean it has been evaluated by the U.S. Federal Government. Read our disclaimer for details.

ClinicalTrials.gov Identifier: NCT00936221

Recruitment Status : Completed
First Posted : July 9, 2009
Results First Posted : December 9, 2015
Last Update Posted : March 14, 2016

Sponsor:
AstraZeneca

Information provided by (Responsible Party):
AstraZeneca

Study Details | Tabular View | Study Results | Disclaimer | How to Read a Study Record

Study Description Go to

Brief Summary:
To assess the efficacy in terms of overall survival of AZD6244 in combination with dacarbazine, compared with dacarbazine alone, in first line patients with BRAF mutation positive advanced cutaneous or unknown primary melanoma

Condition or disease	Intervention/treatment	Phase
Melanoma	Drug: AZD6244 Drug: Dacarbazine Drug: Placebo	Phase 2

Study Design Go to

Study Type :	Interventional (Clinical Trial)
Actual Enrollment :	385 participants
Allocation:	Randomized
Intervention Model:	Parallel Assignment
Masking:	Quadruple (Participant, Care Provider, Investigator, Outcomes Assessor)
Primary Purpose:	Treatment
Official Title:	A Phase II, Double-blind, Randomised Study to Assess the Efficacy of AZD6244 in Combination With Dacarbazine Compared With Dacarbazine Alone in First Line Patients With BRAF Mutation Positive Advanced Cutaneous or Unknown Primary Melanoma
Study Start Date :	July 2009
Actual Primary Completion Date :	November 2011
Actual Study Completion Date :	November 2014

Figure B.2: *Example Clinical Trial Results Record*

1. Primary Outcome

Title	Overall Survival
▼ Description	Following progression survival data was collected until documentation of death, withdrawal of consent, loss to follow-up or the final data cut-off, whichever occurred first.
Time Frame	From date of randomization until death, withdrawal of consent or the end of the study. The end of the study was defined as the date all AZD6244 patients had been followed for a minimum of 12 months, or the date of final analysis, whichever was later

▼ Outcome Measure Data

▼ Analysis Population Description
Intention to Treat (ITT)

Arm/Group Title	Selumetinib 75mg BD + Dacarbazine	Placebo + Dacarbazine
▼ Arm/Group Description:	selumetinib 75mg twice daily + dacarbazine	Matched Placebo + dacarbazine
Overall Number of Participants Analyzed	45	46
Median (Full Range) Unit of Measure: Days		
	424 (63 to 760)	321 (66 to 739)

▼ Statistical Analysis 1

Statistical Analysis Overview	Comparison Group Selection	Selumetinib 75mg BD + Dacarbazine, Placebo + Dacarbazine
	Comments	If the true hazard ratio (HR) is 0.57, 58 deaths provides at least 80% power to demonstrate a statistically significant difference for OS, assuming a 1-sided 10% significance level.
	Type of Statistical Test	Superiority or Other
	Comments	[Not Specified]
Statistical Test of Hypothesis	P-Value	0.3873
	Comments	1-sided p-value
	Method	Regression, Cox
	Comments	Cox model adjusting for treatment, WHO performance status, LDH, M status and tumour sub-type.
Method of Estimation	Estimation Parameter	Hazard Ratio (HR)
	Estimated Value	0.93
	Confidence Interval	(2-Sided) 80% 0.67 to 1.28
	Estimation Comments	[Not Specified]

B.3 ClinicalTrials.gov Results Template

Figure B.3: *ClinicalTrials.gov Results Template*

More details available in the Results Data Element Definitions. April 2017

Outcome Measure Template ***ClinicalTrials.gov***

* Outcome Measure Type	(Select One)	Primary	Secondary	Other Pre-specified	Post-Hoc
* Outcome Measure Title					
[*] Outcome Measure Description					
* Outcome Measure Time Frame					

* Arm/Group Title						
*§ Arm/Group Description ①						
* Overall Number of Participants Analyzed ②						
[*] Analysis Population Description						
* Measure Type	* Measure of Dispersion/Precision					
(Select One) Count of Participants ③ Mean Median Least Squares Mean (LSM) Geometric Mean Geometric LSM Number Count of Units ③	(Select One) Not Applicable ④ Standard Deviation Standard Error Inter-Quartile Range Full Range _____ % Confidence Interval Geometric Coefficient of Variation					
[*] Row/Category Title ⑤			③④		③④	③④
[*] Row/Category Title ⑤			③④		③④	③④
* Unit of Measure						

* Required *§ Required if Primary Completion Date is on or after January 18, 2017 [*] Conditionally required

① Arm/Group Description describes details about the intervention strategy (e.g., dose, dosage form, frequency, duration) or groups evaluated.
② Overall Number of Units Analyzed and Type of Units Analyzed may also be specified.
③ If Measure Type is a "count," percentage of participants/units is automatically calculated from Overall Number of Participants/Units Analyzed. The percentage can be hidden (display is optional).
④ Not Applicable should be used only if Measure Type is Number, Count of Participants, or Count of Units. No dispersion/precision value is needed if Measure of Dispersion is Not Applicable.
⑤ [Optional] Add as many Rows/Categories as needed. If more than one is entered, a Row/Category Title and Outcome Measure Data are required for each row. Row/Category Titles are only required if more than one row.

Notes: This figure shows a template trial sponsors use to submit clinical trial results to ClinicalTrials.gov. Source: `https://prsinfo.clinicaltrials.gov/results_table_layout/DataEntryTable_StatAnalysisForm.pdf`

B.4 Additional Figures and Tables

Table B.2: *Summary Stats: Trial-Cancer Level*

	Observations	Mean	Standard Deviation	Minimum	Maximum
Phase					
0/1: Phase 2	31,125	0.89	0.31	0	1
0/1: Phase 3	31,125	0.09	0.29	0	1
0/1: Phase 4	31,125	0.02	0.13	0	1
Funding Institution					
0/1: Industry	31,125	0.64	0.48	0	1
0/1: Government	31,125	0.21	0.41	0	1
0/1: Academic, Non-Government	31,125	0.15	0.36	0	1
Results Disclosure					
0/1: Disclosed Results Ever	31,125	0.55	0.50	0	1
0/1: Disclosed Results Within 12 Months	31,125	0.17	0.37	0	1
Clinical Results: Objective Response Rate (Phase 2 Only)					
Primary Endpoints	2,874	24.87	28.09	0	100
All Endpoints	4,881	21.94	25.71	0	100
Design					
0/1: Validated Endpoint	31,125	0.22	0.41	0	1
0/1: Randomized	31,125	0.26	0.44	0	1
0/1: Doubled Blinded	31,125	0.05	0.21	0	1
0/1: Control (Placebo)	28,636	0.09	0.28	0	1
0/1: Control (Active Comparator)	28,636	0.13	0.34	0	1
0/1: Control (Sham Comparator)	28,636	0.00	0.01	0	1
Specialization					
# Diseases Examined	31,125	32.75	29.51	1	80
# Patients Enrolled	30,907	104.21	365.01	1	35,533
Competition					
# Drugs (Prior 5 Years)	31,125	252.40	100.07	45	709
# Drugs (Since 2000)	31,125	384.58	236.71	0	1,572
Other Characteristics					
# Trial Sites	29,317	17.14	46.34	1	932
0/1: Small Molecule Intervention	31,125	0.85	0.35	0	1
Patient Age (Mean)	30,500	24.00	14.64	0	120
Trial Sponsor Experience (# Patents)	31,125	1037.32	1295.14	0	8,602

Notes: This table shows summary statistics at the trial-cancer level in the Full Sample. There are 48,406 trial-cancer observations in this sample. The period of analysis is 2000-2016. "Disclosed Results Ever" = 1 if a trial's result is disclosed by July 2018. See text and the appendix for more detailed data and variable descriptions.

Table B.3: *Competition and Trial Design, Competition Since 2000*

	0/1: Validated Endpoint		0/1: Randomized		0/1: Double Blinded		0/1: Controlled (Placebo)		0/1: Controlled (Active Comparator)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
ln(Competition, # drugs + 1)	0.279***	0.189**	0.092	0.090	0.069*	0.057	0.115**	0.114*	-0.045	-0.021
	(0.067)	(0.068)	(0.096)	(0.101)	(0.039)	(0.043)	(0.054)	(0.059)	(0.047)	(0.050)
Mean of Dep. Var.	0.23	0.23	0.22	0.22	0.02	0.02	0.05	0.05	0.12	0.12
Controls	X	X	X	X	X	X	X	X	X	X
Full Sample	X		X		X		X		X	
Restricted Sample		X		X		X		X		X
Observations	16,210	15,027	16,210	15,027	16,210	15,027	16,210	15,027	16,210	15,027
R^2	0.09	0.11	0.06	0.06	0.02	0.02	0.03	0.04	0.04	0.04

Notes: This table examines the relationship between competition and trial design. Trial-cancer level observations. Estimates are from OLS models. Competition is measured by the number of drugs tested in the same disease, since 2000. Odd-numbered columns show estimates using the Full Sample and even-numbered columns show estimates using the Restricted Sample. Controls include cancer fixed effects, year of trial start fixed effects, firm experience, and drug type. Robust standard errors, clustered at the cancer level, are shown in parenthesis.
*p<0.10, p**<0.05, ***p<0.01.

Table B.4: *Competition and Trial Specialization, Competition Since 2000*

	ln(# Diseases Examined)		ln(# Patients Enrolled)		Primary Endpoints		All Endpoints	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(Competition, # drugs + 1)	-0.604*	-0.686*	0.460**	0.308	84.378***	57.974***	83.731***	55.523***
	(0.358)	(0.366)	(0.223)	(0.211)	(15.666)	(16.130)	(13.676)	(13.177)
Mean of Dep. Var.	2.95	2.95	3.63	3.63	24.11	24.11	21.78	21.78
Controls	X	X	X	X	X	X	X	X
Full Sample	X		X					
Restricted Sample		X		X				
Disclosed, Full Sample					X		X	
Disclosed, Restricted Sample						X		X
Observations	16,210	15,027	16,210	15,027	1,814	1,583	3,528	3,084
R^2	0.42	0.42	0.14	0.15	0.33	0.35	0.28	0.29

Notes: This table examines the relationship between competition and trial specialization. Trial-cancer level observations. Estimates are from OLS models. Competition is measured by the number of drugs tested in the same disease, since 2000. Columns 1 and 3 show estimates using the Full Sample and Columns 2 and 4 show estimates using the Restricted Sample. Columns 5 and 7 show estimates using the subset of Full Sample trials with disclosed trial results. Columns 6 and 8 show estimates using the subset of Restricted Sample trials with disclosed trial results. Controls include cancer fixed effects, year of trial start fixed effects, firm experience, and drug type. Robust standard errors, clustered at the cancer level, are shown in parenthesis.
*p<0.10, p**<0.05, ***p<0.01.

Table B.5: *Competition and Disclosure, Competition Since 2000*

	(1)	(2)	(3)	(4)
ln(Competition, # drugs + 1)	0.540***	0.423***	0.584***	0.489***
	(0.080)	(0.079)	(0.083)	(0.082)
0/1: Validated Endpoint		0.190***		0.219***
		(0.008)		(0.009)
0/1: Randomized		0.026***		0.030***
		(0.007)		(0.007)
0/1: Double Blinded		0.133***		0.137***
		(0.026)		(0.025)
0/1: Control (Placebo)		-0.048**		-0.050**
		(0.018)		(0.017)
0/1: Control (Active Comparator)		-0.023*		-0.019
		(0.013)		(0.013)
ln(Diseases Tested (#))		-0.008***		-0.006**
		(0.002)		(0.002)
ln(# Patients Enrolled)		0.009***		0.007**
		(0.002)		(0.003)
0/1: Intervention: Small Molecule		-0.027**		-0.019*
		(0.010)		(0.010)
ln(Trial Sponsor Experience, # patents + 1)		0.152***		0.141***
		(0.006)		(0.006)
Mean of Dep. Var.	0.15	0.15	0.15	0.15
Controls	X	X	X	X
Full Sample	X	X		
Restricted Sample			X	X
Observations	16,210	15,642	15,027	14,461
R^2	0.05	0.14	0.05	0.15

Notes: This table examines the relationship between competition and trial design. Trial-cancer level observations. Estimates are from OLS models. Competition is measured by the number of drugs tested in the same disease, since 2000. The first two columns show estimates using the Full Sample and the last two columns show estimates using the Restricted Sample. Controls include cancer fixed effects and year of trial start fixed effects. Robust standard errors, clustered at the cancer level, are shown in parenthesis.
*p<0.10, p**<0.05, ***p<0.01.

Appendix C

Appendix to Chapter 3

C.1 Medicare Inpatient Reimbursement

Overview of Medicare Payments. This section provides an overview of how Medicare reimburses hospitals for care provided to beneficiaries, which will be useful to understand the impact of the reform analyzed in this paper.

Inpatient hospital care is covered under Medicare Part A. Since 1984, payments have been under the Prospective Payment System (PPS). The total PPS payment received by hospital h in year t can be expressed as:

$$PPS_{h,t} = f(patient_{h,t}, p_{h,t}, drg, outlier_{h,t}, ime_{h,t}, dsh_{h,t}) \tag{C.1}$$

Under the PPS, each Medicare patient ($patient_{h,t}$) in a given Diagnosis Related Group (DRG) is given a fixed payment ($p_{h,t}$). There approximately 1000 DRGs total and each is assigned a weight (drg) according to hospitals' aggregate historical costs of treating patients in each DRG. The PPS payment also includes outlier payments—$outlier_{h,t}$—reimbursements made to compensate providers for patients with exceptionally costly stays Keeler *et al.* (1998). Each of these payments are multiplied by several subsidies. The largest subsidies are indirect medical education subsidies ($ime_{h,t}$) and disproportionate share ($dsh_{h,t}$) subsidies. These adjustments correspond to payments received by hospitals for training physicians and treating

poor patients. The shock to hospital financing comes from cuts to the $ime_{h,t}$ and $dsh_{h,t}$ subsidies, which we now describe in further detail.

Indirect Medical Education Subsidies. Teaching hospitals receive two supplemental payments from Medicare: direct medical education (DME) and indirect medical education (IME) payments, which account for 38% and 62% respectively of total GME payments in 1998. DME payments reimburse a teaching hospital for Medicare's share of the direct costs of training residents. The IME adjustment is a percentage add-on payment to the hospital's basic DRG payment (Fishman, 1992). IME payments are meant to compensate teaching hospitals for indirect expenses stemming for example from use of diagnostic services by clinically inexperienced residents or decreased productivity of nurses and support staff involved in teaching of residents. Since 1989, the DRG payment a hospital receives for admitting a Medicare patient increases non-linearly with the hospital's resident-to-bed ratio and a multiplier.

$$ime_{h,t} = \alpha_t \times \left[\left(1 + \frac{residents_{h,t}}{beds_{h,t}} \right)^{.405} - 1 \right] \tag{C.2}$$

where α is a multiplier set at 1.89 in the pre-reform period. This correspond to a price increase of approximately 7.65% for every 10% increase in a hospital's resident-to-bed ratio.

Disproportionate Share Subsidies The Medicare DSH adjustment was enacted by the Consolidated Omnibus Budget Reconciliation Act of 1985 and became effective in 1986. Like the IME adjustment, the DSH adjustment is a percentage add-on to the hospital's basic DRG payment. The key determinant of whether a hospital is eligible for this subsidy is the fraction of total patient-days allocated to poor patients. Above a certain threshold, hospitals become eligible for a DSH payment adjustment, which varies according to whether the hospital is urban or rural, is a sole community hospital, and the number of beds.[1]

Balanced Budget Act. The BBA and subsequent reforms (BBRA and BIPA) modified IME and DSH subsidies. Specifically, the net effect of the BBA, BBRA, and BIPA on the

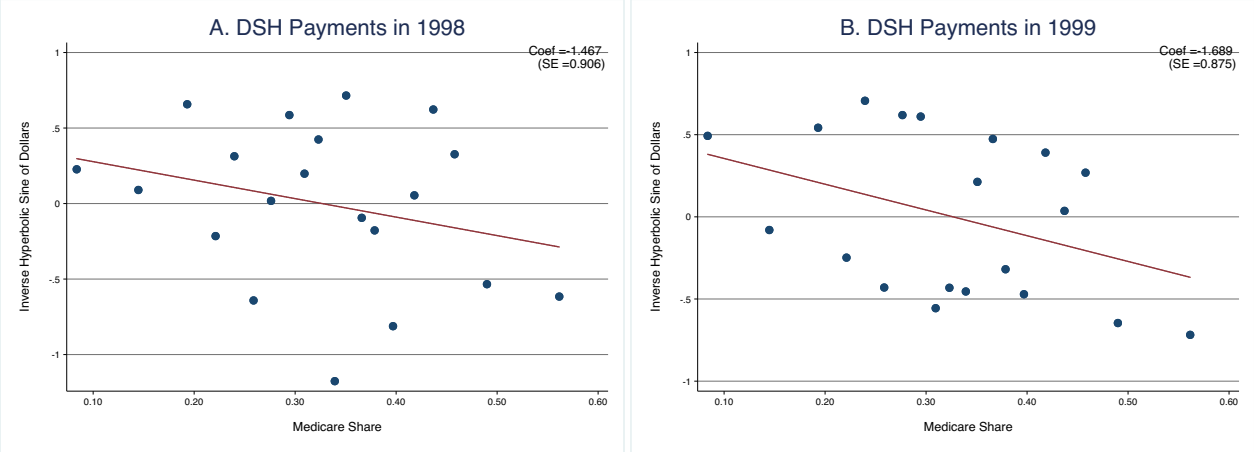
[1]For more details, see `https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/dsh`

multiplier α_t in Equation C.2 is:

$$\alpha_t = \begin{cases} 1.72 \text{ for discharges occurring in fiscal year 1998,} \\ 1.60 \text{ for discharges occurring between fiscal year 1999 and 2000,} \\ 1.54 \text{ for discharges occurring between October 2000 and March 2001,} \\ 1.66 \text{ for discharges occurring between April 2001 and September 2001,} \\ 1.60 \text{ for discharges occurring in fiscal year 2002,} \\ 1.35 \text{ for discharges occurring in fiscal year 2003 and March 2004,} \\ 1.47 \text{ for discharges occurring in between April 2004 and September 2004,} \\ 1.42 \text{ for discharges occurring in fiscal year 2005,} \\ 1.37 \text{ for discharges occurring in fiscal year 2006,} \\ 1.32 \text{ for discharges occurring in fiscal year 2007 and thereafter.} \end{cases}$$

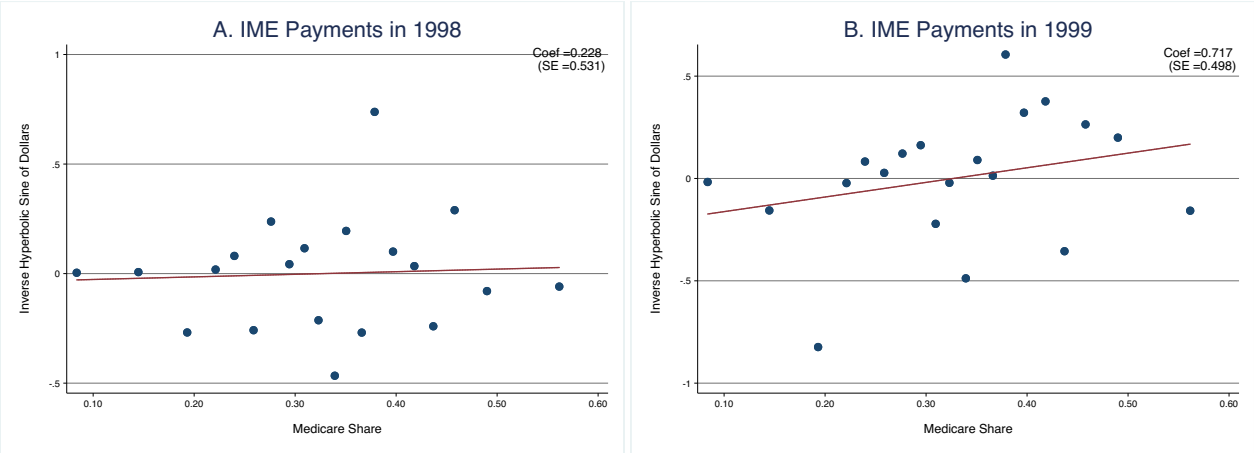
Relationship Between Medicare Share and Change in Hospital Payments. The bin scatter plots in Figures C.1 and C.2 describe how a hospital's pre-BBA Medicare Share relates to the percent change in average IME (Figure C.1) and DSH (Figure A2) payments. We plot each hospital's Medicare Share against the inverse hyperbolic sine of the total subsidy payment in 1998 (panel A) or 1999 (panel B). Plotted values are residuals after controlling for hospital fixed effects and year fixed effects. The figures show that an increase in Medicare Share is associated with a decrease in DSH payments, but not IME payments.

Figure C.1: *Medicare Share and Disproportionate Medical Share Payments*



Notes: Each dot represents 36 hospitals, on average. The x-axis measures pre-BBA Medicare Share. The y-axis measures the inverse hyperbolic sine of DSH payments in 1998 (Panel A) and 1999 (Panel B). Plotted values are residuals after controlling for hospital fixed effects and year fixed effects. Observations are "binned" according to their x-axis values.

Figure C.2: *Medicare Share and Indirect Medical Education Payments*

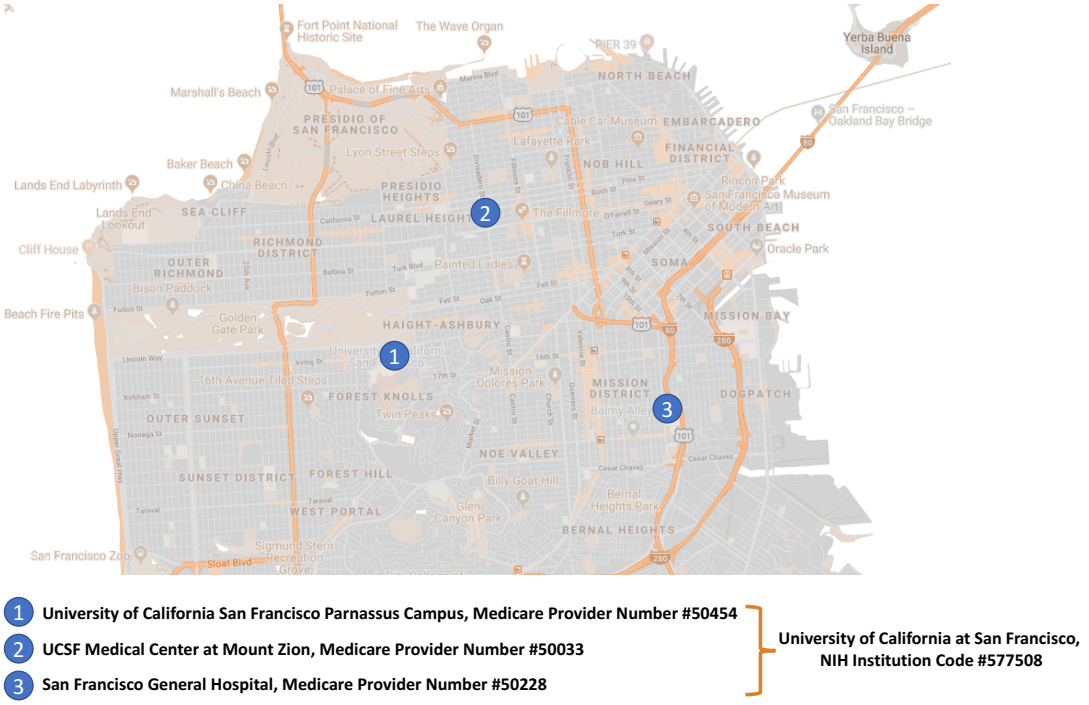


Notes: Each dot represents 36 hospitals, on average. The x-axis measures pre-BBA Medicare Share. The y-axis measures the inverse hyperbolic sine of IME payments in 1998 (Panel A) and 1999 (Panel B). Plotted values are residuals after controlling for hospital fixed effects and year fixed effects. Observations are "binned" according to their x-axis values.

C.2 Mapping Outcomes to Hospitals

Figure C.3 provides an example of how NIH grant IDs and Medicare provider IDs are allocated to hospitals and medical centers. The University of California, San Francisco medical center includes the Parnassus Heights Campus, Mt. Zion Hospital and Medical Center, and San Francisco General Hospital. Each of these locations has a unique Medicare provider number and therefore receive an independent Medicare payment. However, the three campuses share a single, common NIH institutional code. Our strategy consists of looking at each of the PI addresses affiliated with the UCSF NIH institutional code and allocating each PI (and grant) to one of the three hospitals.

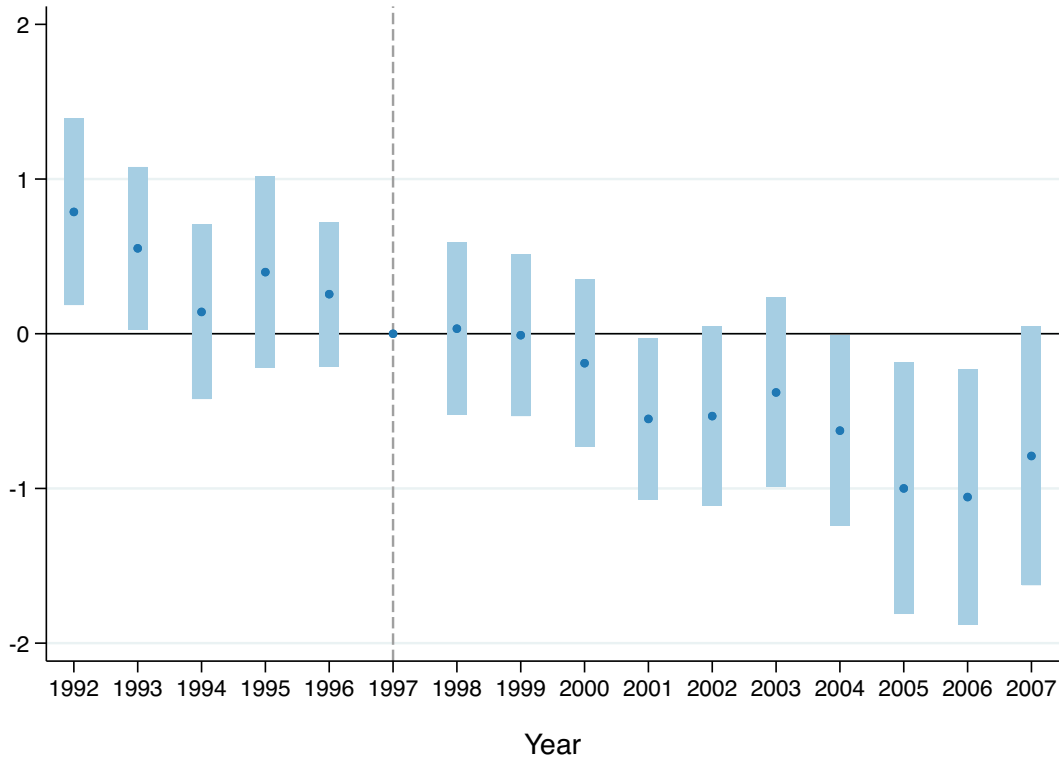
Figure C.3: *Mapping Research Activity to Hospitals: An Example*



C.3 Academic Medical Center Sample Results

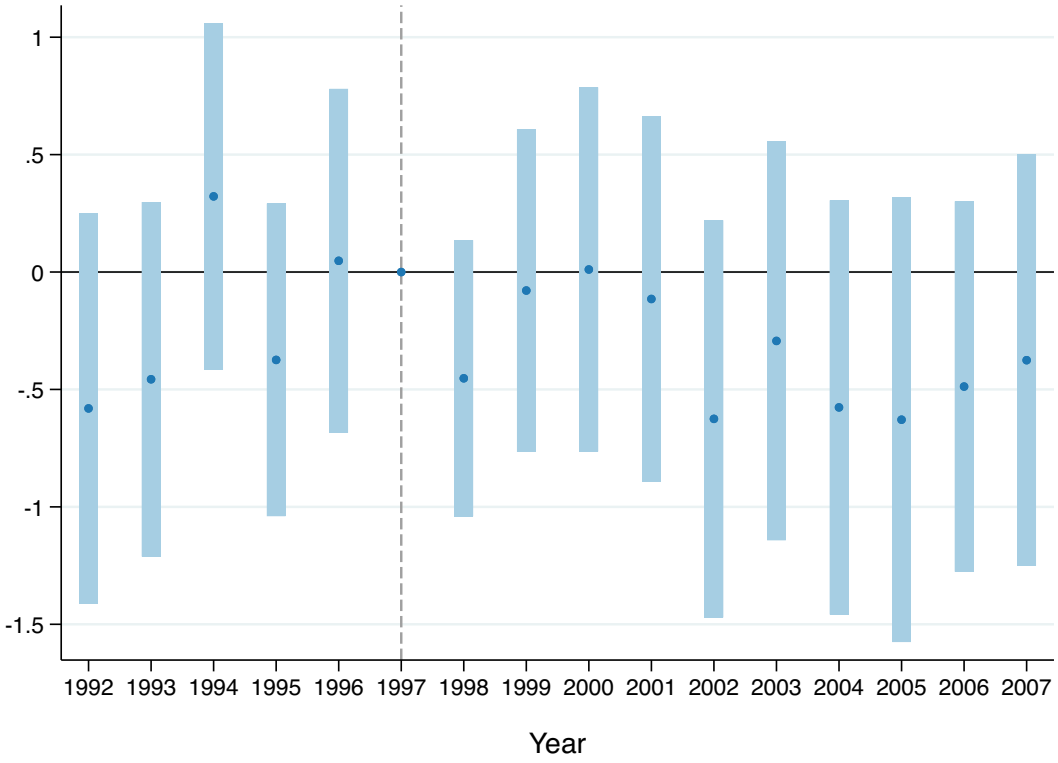
This section provides summary statistics and regression results for the Academic Medical Center sample of hospitals. See Section 3.3 for a description of how hospitals in this sample were identified.

Figure C.4: *Impact on the Number of Grant Applications*



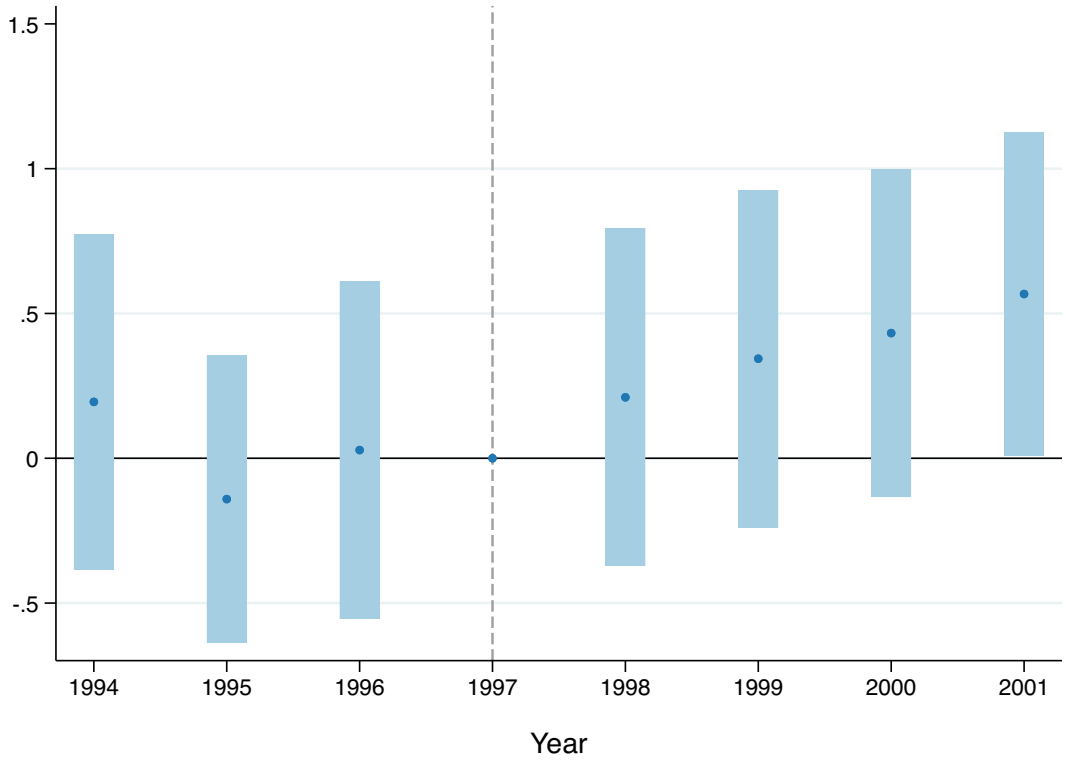
Notes: The dark blue dots in the figure above correspond to coefficient estimates stemming from OLS specifications in which the number of grant applications is regressed onto year effects, hospital effects, as well as interaction terms between a hospital's Medicare Share interacted with the number of years before/after the BBA (the Medicare share interacted with the year of the BBA, 1997, is omitted). The 95% confidence interval (corresponding to robust standard errors, clustered around hospital) around these estimates is plotted with light blue bars. The dashed grey line corresponds to the year of the BBA. The hospital sample used is the AMC sample.

Figure C.5: *Impact on the Number of Publications*



Notes: The dark blue dots in the figure above correspond to coefficient estimates stemming from OLS specifications in which the number of publications is regressed onto year effects, hospital effects, as well as interaction terms between a hospital's Medicare Share interacted with the number of years before/after the BBA (the Medicare share interacted with the year of the BBA, 1997, is omitted). The 95% confidence interval (corresponding to robust standard errors, clustered around hospital) around these estimates is plotted with light blue bars. The dashed grey line corresponds to the year of the BBA. The hospital sample used is the AMC sample.

Figure C.6: *Impact on Bedside Research*
(Number of Industry-Sponsored Trial Sites)



Notes: The dark blue dots in the figure above correspond to coefficient estimates stemming from OLS specifications in which the number of industry-sponsored clinical trial sites is regressed onto year effects, hospital effects, as well as interaction terms between a hospital's Medicare Share interacted with the number of years before/after the BBA (the Medicare share interacted with the year of the BBA, 1997, is omitted). The 95% confidence interval (corresponding to robust standard errors, clustered around hospital) around these estimates is plotted with light blue bars. The dashed grey line corresponds to the year of the BBA. The hospital sample used is the AMC sample.

Table C.1: *Summary Statistics*

	Observations	Mean	Median	Standard Deviation	Minimum	Maximum
Panel A. Hospital Characteristics						
Discharges (1,000s)	282	25.32	23.52	11.35	5.12	65.82
Medicare Share of Discharges	282	0.30	0.31	0.10	0.02	0.54
Medicare Teaching Payment ($1,000,000s)	282	11.52	8.34	9.98	0.13	59.76
Medicare Disproportionate Share Payment ($1,000,000s)	282	6.99	6.17	5.72	0.00	36.93
Panel B. Grants						
Number of Grant Applications						
Total	282	22.37	0.94	50.21	0.00	444.00
Renewal	282	4.25	0.12	10.10	0.00	88.25
New	282	18.12	0.78	40.22	0.00	355.75
MD	282	7.70	0.52	17.98	0.00	158.62
PhD	282	10.94	0.29	24.99	0.00	193.38
MD-PhD	282	3.45	0.07	8.95	0.00	87.75
Funding Amount ($ Mill. Dollars)						
Total	282	33.15	0.92	79.27	0.00	709.37
Renewal	282	9.12	0.16	23.19	0.00	211.80
New	282	24.03	0.74	56.55	0.00	497.57
MD	282	13.61	0.54	34.24	0.00	297.43
PhD	282	13.51	0.16	31.84	0.00	256.55
MD-PhD	282	5.83	0.06	16.71	0.00	153.27
Panel C. Publications						
Total	282	114.40	19.00	225.62	0.00	1,683.62
Citation Ranking: $\leq$25	282	27.75	7.41	45.71	0.00	306.12
Citation Ranking: 26-50	282	26.19	4.75	48.12	0.00	333.38
Citation Ranking: 51-75	282	27.95	4.06	56.79	0.00	413.19
Citation Ranking: 76-90	282	18.32	2.38	41.24	0.00	325.19
Citation Ranking: 91-95	282	6.68	0.71	16.50	0.00	131.44
Citation Ranking: 96-99	282	5.78	0.62	15.67	0.00	130.56
Citation Ranking: >99	282	1.72	0.19	5.06	0.00	45.12
"Basic Research": Cited in Patent	282	16.80	1.62	39.11	0.00	316.44
"Basic Research": MeSH	282	17.47	0.75	40.61	0.00	269.44
"Translational" Research	282	24.30	3.09	50.92	0.00	403.94
Builds on "Translational" Research	282	6.76	1.66	12.89	0.00	85.75
Inspires "Translational" Research	282	11.10	1.31	24.97	0.00	208.94
Panel D. Trials						
Number of Trials	282	4.55	0.00	8.96	0.00	41.62
Trials Sites (1,000s)	282	414.19	0.00	851.64	0.00	4522.10
Panel E. Clinical Outcomes						
Risk-adjusted Survival Rates (30 days), Heart Attack	271	0.91	0.91	0.03	0.74	1.00
Risk-adjusted Survival Rates (30 days), Heart failure	271	0.97	0.97	0.02	0.90	1.00
Risk-adjusted Survival Rates (30 days), Pneumonia	271	0.98	0.99	0.02	0.90	1.03
Risk-adjusted Survival Rates (30 days), Hip/knee	271	0.95	0.95	0.02	0.86	1.04

Notes: This table shows summary statistics for the primary analytic hospital sample between 1992-2007. All variables are measured yearly. For example, "Discharges (1000s)" is the average number of patients a hospital receives in a year. Clinical outcomes are an exception and are measured in three-year bins—e.g., hospital-level survival rates in 2005 are estimated over patient claims in 2003, 2004, and 2005. For each hospital, we use three-year bins for four years (1996, 2002, 2005, and 2008). The hospital sample used is the AMC sample.

Table C.2: *Comparing the Primary Analytic and AMC Sample*

	Primary Analytic Sample			AMC Sample			
	Observations	Mean	Standard Deviation	Observations	Mean	Deviation Deviation	P-value from T-Test Diff. of Means
Panel A. Hospital Characteristics							
Discharges (1,000s)	725	18.57	10.78	282	25.32	11.35	0.00***
Medicare Share of Discharges	725	0.33	0.11	282	0.30	0.10	0.00***
Medicare Teaching Payment ($1,000,000s)	725	5.67	7.97	282	11.52	9.98	0.00***
Medicare Disproportionate Share Payment ($1,000,000s)	725	4.37	4.83	282	6.99	5.72	0.00***
Panel B. Grants Applications							
Total	725	8.87	33.1	282	22.37	50.21	0.00***
Renewal	725	1.68	6.62	282	4.25	10.1	0.00***
New	725	7.19	26.54	282	18.12	40.22	0.00***
MD	725	3.06	11.8	282	7.7	17.98	0.00***
PhD	725	4.33	16.45	282	10.94	24.99	0.00***
MD-PhD	725	1.36	5.82	282	3.45	8.95	0.00***
Panel C. Publications							
Total	725	46.44	150.84	282	114.4	225.62	0.00***
Citation Ranking: ≤25	725	11.44	31.37	282	27.75	45.71	0.00***
Citation Ranking: 26-50	725	10.65	32.49	282	26.19	48.12	0.00***
Citation Ranking: 51-75	725	11.3	37.84	282	27.95	56.79	0.00***
Citation Ranking: 76-90	725	7.37	27.16	282	18.32	41.24	0.00***
Citation Ranking: 91-95	725	2.68	10.77	282	6.68	16.5	0.00***
Citation Ranking: 96-99	725	2.32	10.15	282	5.78	15.67	0.00***
Citation Ranking: >99	725	.69	3.26	282	1.72	5.06	0.00***
"Basic Research": Cited in Patent	725	6.73	25.67	282	16.8	39.11	0.00***
"Basic Research": MeSH	725	6.89	26.68	282	17.47	40.61	0.00***
"Translational" Research	725	9.73	33.82	282	24.3	50.92	0.00***
Builds on "Translational" Research	725	2.79	8.65	282	6.76	12.89	0.00***
Inspires "Translational" Research	725	4.44	16.45	282	11.1	24.97	0.00***
Panel D. Trials							
Number of Trials	725	1.8	6.01	282	4.55	8.96	0.00***
Trial Sites (1,000s)	725	163.27	567.82	282	414.19	851.64	0.00***
Panel E. Clinical Outcomes							
Survival Rates (30 days), Heart Attack	668	0.91	0.03	271	0.91	0.03	0.26
Survival Rates (30 days), Heart Failure	668	0.96	0.02	271	0.97	0.02	0.00**
Survival Rates (30 days), Pneumonia	668	0.98	0.02	271	0.98	0.02	0.17
Survival Rates (30 days), Hip/Knee	668	0.95	0.02	271	0.95	0.02	0.27

Notes: This table compares the primary analytic hospital sample and the AMC sample between 1992-2007. All variables are measured yearly. For example, "Discharges (1000s)" is the average number of patients a hospital receives in a year. Clinical outcomes are an exception and are measured in three-year bins—e.g., hospital-level survival rates in 2005 are estimated over patient claims in 2003, 2004, and 2005. For each hospital, we use three-year bins for four years (1996, 2002, 2005, and 2008).

Table C.3: *Impact on the Number of Grant Applications*

		Grant Cycle		Principal Investigator		
	Total	New	Renewal	MD	PhD	MD-PhD
	(1)	(2)	(3)	(4)	(5)	(6)
Medicare Share × After	-0.863***	-0.910***	-0.0641	-0.436**	-0.804***	-0.659**
	(0.232)	(0.231)	(0.185)	(0.218)	(0.212)	(0.216)
Mean of Dep. Var	1.80	1.69	0.94	1.26	1.32	0.82
% Change in 1 SD Share	-9.472	-9.98	-0.70	-4.78	-8.82	-7.23
Observations	4,475	4,475	4,475	4,475	4,475	4,475
Test for Diff. in Percent Change		0.00		0.10		

Notes: This table displays the effect on the number of grant applications. The hospital sample used is the AMC sample. Outcome variables have been transformed with inverse hyperbolic sine transformations. Column 1 presents results from OLS regressions. Columns 2-6 present estimates from seemingly unrelated regressions. Hospital fixed effects and year fixed effects are included in all regressions, and standard errors are clustered at the hospital level. The third row shows the coefficient on Medicare Share × After multiplied by 1 SD in Medicare Share. The fourth row shows t-test results from comparing estimates in different columns (Column 2 vs. Column 3; Column 4 vs. Column 5).
*p<0.10, p**<0.05, ***p<0.001.

Table C.4: *Impact on the Number of Publications, Quality*

	Total	Citation Ranking						
		≤25	26-50	51-75	76-90	91-95	96-99	>99
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Medicare Share × After	-0.188	-0.287	-0.249	-0.456*	-0.458*	-0.605**	-0.573**	-0.366**
	(0.314)	(0.287)	(0.281)	(0.268)	(0.270)	(0.232)	(0.214)	(0.167)
Mean of Dep. Var	3.79	2.76	2.47	2.36	1.93	1.23	1.11	0.57
% Change in 1 SD Share	-2.06	-3.15	-2.73	-5.00	-5.02	-6.63	-6.28	-4.02
Observations	4,475	4,475	4,475	4,475	4,475	4,475	4,475	4,475
T-test P-Value	-	0.46	0.71	0.14	0.24	0.11	0.18	0.58

Notes: This table displays the effect on total publications and publication quality in hospitals. The hospital sample used is the AMC sample. Outcome variables have been transformed with inverse hyperbolic sine transformations. Estimates are from seemingly unrelated regressions. Hospital fixed effects and year fixed effects are included in all regressions, and standard errors are clustered at the hospital level. The third row shows the coefficient on Medicare Share × After multiplied by 1 SD in Medicare Share. The fourth row shows t-test results from comparing estimates in Column 1 vs. estimates in other columns (Column 1 vs. Column 2; Column 1 vs. Column 2, etc.).
*p<0.10, p**<0.05, ***p<0.001.

Table C.5: *Impact on Research Direction*

	"Bench" Research		"Translational" Research			"Bedside" Research	
	Publications, Cited in Patent	Publications, Basic MeSH	Publications, "Translational" MeSH	Publications, Builds on "Translational" MeSH	Publications, Inspiring "Translational" MeSH	Number of Industry-Sponsored Trial Sites	Industry-Sponsored Trial Dollars ($)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Medicare Share × After	0.250	-0.299	-0.325	-0.309	-0.533**	0.367**	1.417
	(0.251)	(0.299)	(0.309)	(0.243)	(0.245)	(0.143)	(0.976)
Mean of Dep. Var.	1.78	1.54	2.11	1.51	1.55	0.93	4.37
% Change in 1 SD Share	2.781	-3.225	-3.505	-3.336	-5.684	4.027	15.54
Observations	4,475	4,475	4,475	4,475	4,475	2,243	2,243

Notes: This tables displays the effect on "bench," "translational," and "bedside" research in hospitals. The hospital sample used is the AMC sample. Outcome variables have been transformed with inverse hyperbolic sine transformations. Estimates are from OLS regressions. Hospital fixed effects and year fixed effects are included in all regressions, and standard errors are clustered at the hospital level. The third row shows the coefficient on Medicare Share × After multiplied by 1 SD in Medicare Share. Column 2 refers to publications with MeSH terms affiliated with a molecular biology technique or that use a model organism. Column 3 refers to publications that are disease-oriented and relies either on a molecular biology technique or a model organism (based on MeSH terms). Column 4 refers to publications that report the results of a clinical trials, or are tagged by a human MeSH term and also cite a translational publication. Column 5 refers to publications that are translational and is cited by a clinical trial publication (or one that contains a human MeSH term). Finally, Columns 6 and 7 refer to clinical trial contracts.
*p<0.10, p**<0.05, ***p<0.001.

Table C.6: *Impact on Clinical Outcomes*

	(1) Heart Attack	(2) Heart Failure	(3) Hip/Knee	(4) Pneumonia
Medicare Share	-0.0323	0.00327	0.0335**	-0.0410*
	(0.0329)	(0.0191)	(0.0162)	(0.0228)
Mean of Dep. Var	0.03	0.01	0.00	0.02
% Change in 1 SD Share	-0.338	0.0342	0.350	-0.429
Observations	271	271	271	271

Notes: This table displays the effect on risk-adjusted survival rates in hospitals in the AMC sample. Outcomes are the difference in average survival rates between the post-BBA time period and the pre-BBA time period. Estimates are from OLS regressions. Robust standard errors are clustered at the hospital level.
*p<0.10, p**<0.05, ***p<0.001.