

November 2023

ESSAYS ON MARKET INEFFICIENCIES ARISING FROM INFORMATION ASYMMETRY AND MARKET POWER

Ming Ge
University of Massachusetts Amherst

Follow this and additional works at: https://scholarworks.umass.edu/dissertations_2



Part of the [Behavioral Economics Commons](#), [Health Economics Commons](#), [Industrial Organization Commons](#), and the [Policy Design, Analysis, and Evaluation Commons](#)

Recommended Citation

Ge, Ming, "ESSAYS ON MARKET INEFFICIENCIES ARISING FROM INFORMATION ASYMMETRY AND MARKET POWER" (2023). *Doctoral Dissertations*. 2887.
<https://doi.org/10.7275/35997449> https://scholarworks.umass.edu/dissertations_2/2887

This Open Access Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

**ESSAYS ON MARKET INEFFICIENCIES ARISING FROM
INFORMATION ASYMMETRY AND MARKET POWER**

A Dissertation Presented

by

MING GE

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2023

Department of Resource Economics

© Copyright by Ming Ge 2023

All Rights Reserved

ESSAYS ON MARKET INEFFICIENCIES ARISING FROM INFORMATION ASYMMETRY AND MARKET POWER

A Dissertation Presented

by

MING GE

Approved as to style and content by:

DocuSigned by:



4FFAE2FB63314E0...

Christian Rojas, Co-chair

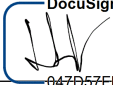
DocuSigned by:



FCE0DEEC0881410...

Rong Rong, Co-chair

DocuSigned by:



047D57FE2F4B4AA...

Juan-Camilo Cárdenas, Member

DocuSigned by:



A4FB90C8840C4E6...

Matt Woerman, Member

DocuSigned by:



6CA29BE0C6B249A...

John Spraggon, Department Chair
Department of Resource Economics

DEDICATION

To those unforgettable and splendid memories of my Ph.D. life;

To those stumbling blocks that failed to trip me;

To those lights that split the darkest night and illuminated my path;

Thanks for letting me embark on this adventure to the end!

ACKNOWLEDGMENTS

First and foremost, I would like to express my profound appreciation to my esteemed doctoral dissertation committee chairs, Dr. Christian Rojas and Dr. Rong Rong, whose invaluable guidance supported me throughout the entire process of researching and writing this dissertation. Without their mentoring and encouragement, this accomplishment would not have been possible.

I would like to thank Dr. Juan-Camilo Cárdenas and Dr. Matt Woerman, who served on my dissertation committee, for providing insightful comments and feedback on my research. I am also grateful to the other faculty members and graduate fellows at the University of Massachusetts Amherst who offered me precious assistance and support during my Ph.D. journey.

I am also indebted to the department chairs, Dr. Daniel Lass and Dr. John Spraggon, as well as the graduate program directors, Dr. Nathan Chan and Dr. Christian Rojas, for their assistance and coordination in securing funding during the academic semesters and summer breaks. Thanks to their diligent efforts, I was able to focus on my studies and research without any financial concerns. In addition, the research grants from the graduate school and the Department of Resource Economics at the University of Massachusetts Amherst are gratefully acknowledged.

Last but not least, I dedicate this accomplishment to my honored parents, Cunde Ge and Zunqiu Pang, and to my beloved wife, Yiming Dai. They have been my pillars of strength, helping me through challenging times and celebrating my achievements. Their constant companionship and unwavering support have made my Ph.D. journey even more enriching.

ABSTRACT

ESSAYS ON MARKET INEFFICIENCIES ARISING FROM INFORMATION ASYMMETRY AND MARKET POWER

SEPTEMBER 2023

MING GE

B.S., WEBER STATE UNIVERSITY

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Christian Rojas and Professor Rong Rong

This dissertation comprises three chapters that empirically investigate various kinds of market inefficiencies arising from seller misbehaviors. The areas of focus include physicians' overtreatment in healthcare markets and the exercise of market power by energy suppliers in wholesale electricity markets. Furthermore, I seek potential remedies to effectively address these sellers' opportunistic behaviors and improve market performance.

The first two chapters of my dissertation focus on the role of reputation in healthcare services, especially its potential to combat physician overtreatment. Healthcare is a salient example of a credence good due to the fact that physicians have an informational advantage over their patients regarding illnesses and appropriate treatments. Asymmetric information between physicians and patients leads to rampant overtreatment and low market efficiency. Thus, exploring effective measures to address overtreatment is of particular importance. Unfortunately, a standard reputation

system falls short of being effective because patients cannot tell whether a high-cost treatment recommendation (versus a less costly and complex treatment) is necessary even after the service is completed.

In the first chapter, titled “**It Takes Two Hands to Clap: The Effects of Reputation and Search in Healthcare Markets**”, I propose a solution to reinstate the function of reputation by combining a reputation mechanism with patient search. The key insight is that patient search can act as a channel to build a meaningful record of physicians’ honesty. I test this new mechanism through a controlled laboratory experiment and demonstrate the effectiveness of reputation using both non-parametric and random-effect panel regression analyses. This study complements the existing body of literature, wherein the authors have shown that enabling patients to seek second opinions effectively reduces overtreatment in a one-time transaction. I provide additional empirical evidence suggesting that the opportunity for reputation-building can further restrain overtreatment and reduce patients’ need for second opinions. As a result of fewer overtreatments and searches, market efficiency is dramatically improved when reputation is at play. Additionally, I manipulate search costs in the experiment and investigate how the level of search costs influences the effectiveness of reputation. I find weak evidence indicating that a prohibitively high search cost discourages patients from seeking second opinions, subsequently mitigating the disciplining effect of reputational concerns. The findings of this study not only illuminate how to design a meaningful reputation mechanism to combat overtreatment but also emphasize the importance of reducing search costs to avoid exacerbating the problem of health inequality.

In the second chapter of my dissertation, “**The Effect of Aggregate Treatment Information on Physician Overtreatment**”, I continue to explore the impact of reputation on the provision of healthcare services. In the first chapter, monitoring (detecting) physician honesty is through patients’ repeated searches. The accumu-

lation of such reputational information is costly both in terms of time and money. In this study, motivated by newly-emerged public information, the aggregate count of each type of treatment provided by a physician, I investigate whether such information could lead to fewer overtreatments and replace patients' reliance on costly searches in repeated interactions. I argue that physicians' aggregate records offer patients a more efficient and cost-saving way to monitor physicians' honesty compared to seeking second opinions. Contrary to my expectations, the results of a laboratory experiment indicate that disclosing physicians' aggregate records provides similar levels of physician overtreatment and patient search when compared to situations where patients can rely only on their own experiences. Furthermore, I observe that increased information transparency leads physicians to behave more strategically: they adjust their overtreatment rate to be close(r) to the rate offered by their competitors. This finding raises the concern that disclosing physicians' aggregate records may facilitate easier coordination among physicians and help create localized norms that may not benefit patients.

The third chapter of my dissertation, “**Market Power in the New England Electricity Market: Evidence from Nuclear Refueling Outages**”, co-authored with Xiaolin Zhou, investigates how nuclear refueling outages affect market-clearing prices and suppliers' bidding behaviors in the New England electricity market. Nuclear power is known for its comparatively low marginal cost among non-renewable energy sources, consistently positioning it at the bottom of the supply curve during reactor operations. However, each reactor needs to undergo periodic refueling approximately every 18 months, and refueling typically leads to month-long outages. The temporary baseload supply shortage provides other energy suppliers with more incentives to exercise market power to increase their markup. Using day-ahead hourly market data from 2016 to 2018, we observe that the market-cleared price increased by \$4.9/MWh on average when at least one of the four reactors in the region is offline.

Next, through a series of analyses of suppliers' bidding behaviors, we find that approximately 36% of the price effect could be explained by the exercise of market power. This raises the concern that increased market power during nuclear refueling outages may result in a sizable welfare transfer from consumers to suppliers. The findings of this study call for stricter regulatory oversight of suppliers' bidding behaviors during temporary baseload supply shocks, particularly in electricity markets where the share of intermittent energy generation is significant or experiencing substantial growth.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	v
ABSTRACT	vi
LIST OF TABLES	xiii
LIST OF FIGURES	xv
CHAPTER	
1. IT TAKES TWO HANDS TO CLAP: EFFECTS OF REPUTATION AND SEARCH IN HEALTHCARE MARKETS	1
1.1 Introduction	1
1.2 Literature Review	7
1.3 Experimental Design	10
1.3.1 Market Procedure	10
1.3.2 Treatment Design	13
1.3.3 Experimental Protocol	15
1.4 Hypotheses	15
1.5 Results	17
1.5.1 Overtreatment	17
1.5.1.1 Overtreatment in Strategy	18
1.5.1.2 Actual Overtreatment	20
1.5.2 Patient Search	22
1.5.2.1 Conditional Search	23
1.5.2.2 Unconditional Search	24
1.5.3 Market Efficiency	26

1.6	Conclusion	28
2.	THE EFFECT OF AGGREGATE TREATMENT INFORMATION ON PHYSICIAN OVERTREATMENT	30
2.1	Introduction	30
2.2	Literature Review	35
2.3	Experimental Design	39
	2.3.1 Market Procedure	40
	2.3.2 Treatment Design	42
	2.3.3 Experimental Protocol	43
2.4	Hypotheses	44
	2.4.1 Overtreatment Level	45
	2.4.2 Search Rate	46
	2.4.3 Market Efficiency	47
2.5	Results	48
	2.5.1 Overtreatment	49
	2.5.1.1 Overtreatment in Strategy	49
	2.5.2 Actual Overtreatment	52
	2.5.3 Patient Search	53
	2.5.3.1 Conditional Search Rate	54
	2.5.3.2 Unconditional Search Rate	56
	2.5.4 Market Efficiency	58
	2.5.5 Variance in Physicians' Decisions within Markets	60
2.6	Conclusion	62
3.	MARKET POWER IN THE NEW ENGLAND ELECTRICITY MARKET: EVIDENCE FROM NUCLEAR REFUELING OUTAGES	65
3.1	Introduction	65
3.2	New England Wholesale Electricity Market	70
	3.2.1 Day-Ahead, Real-Time, and Ancillary Services Markets	71
	3.2.2 Resource Mix	72
3.3	Data	74

3.4	Empirical Analysis	78
3.5	Results	81
3.6	Conclusion	85

APPENDICES

A.	APPENDIX OF CHAPTER 1	88
B.	APPENDIX OF CHAPTER 2	101

BIBLIOGRAPHY	116
---------------------------	------------

LIST OF TABLES

Table	Page
1.1 Experimental Set-up: Conditions	14
1.2 Random effects panel OLS: level of overtreatment in strategy	20
1.3 Random effects panel Probit: level of actual overtreatment	22
1.4 Random effects panel Probit: patient search	24
1.5 Random effects panel OLS: relative market efficiency	27
2.1 Experimental Set-up: Conditions	43
2.2 Random effects panel OLS: level of overtreatment in strategy	50
2.3 Random effects panel Probit: level of actual overtreatment	53
2.4 Random effects panel Probit: conditional patient search	55
2.5 Random effects panel Probit: unconditional patient search	57
2.6 Random effects panel OLS: relative market efficiency	59
2.7 Random effects panel OLS: market-level SD of physicians’ overtreatment in strategy	60
3.1 Fuel Mix (% of Native New England Generation)	72
3.2 Major Five Firms Gas Generation Capacity in 2016	74
3.3 Summary Statistics	79
3.4 Nuclear refueling outages & utilization rates – Effect on market prices	82
3.5 Nuclear refueling outages – Effect on suppliers’ bidding price	83

3.6 Nuclear utilization rate – Effect on suppliers’ bidding price 84

LIST OF FIGURES

Figure	Page
1.1 Level of Overtreatment in Strategy	19
1.2 Level of Actual Overtreatment	21
1.3 Conditional Search Rate	23
1.4 Unconditional Search Rate	25
1.5 Relative Market Efficiency	26
2.1 Level of Overtreatment in Strategy	49
2.2 Level of Actual Overtreatment	52
2.3 Conditional Search Rate	54
2.4 Unconditional Search Rate	56
2.5 Relative Market Efficiency	58
3.1 Welfare impact of market power	67
3.2 Market-wide aggregated offer curves. 2018/09/27: nuclear utilization at 93%; 2018/10/11: nuclear utilization at 30%.	77
A.1 Example seller's decision screen	91
A.2 Example buyer's first decision screen.....	92
A.3 Example buyer's second decision screen	93
A.4 Example buyer's third decision screen.....	94
A.5 Example buyer's first decision screen in Period 4 in R_7	95

B.1	Example seller's decision screen	104
B.2	Example buyer's first decision screen.....	105
B.3	Example buyer's second decision screen	106
B.4	Example buyer's third decision screen.....	107
B.5	Example buyer's first decision screen in Period 4 in PRI	108
B.6	Example buyer's first decision screen in Period 4 in AGG_P and AGG_P&PH	109
B.7	Example seller's decision screen in Period 4 in AGG_P&PH.....	110

CHAPTER 1

IT TAKES TWO HANDS TO CLAP: EFFECTS OF REPUTATION AND SEARCH IN HEALTHCARE MARKETS¹

1.1 Introduction

Healthcare, a prominent example of a credence good², is characterized by several factors that dampen market efficiency. Early work by Darby and Karni (1973) highlighted that a fundamental reason for such inefficiencies is the lack of information symmetry in credence goods markets; patients are less informed than physicians regarding the optimal treatment for their health problems. Therefore, uninformed patients must rely on physicians to diagnose their problems and provide appropriate treatment recommendations. While patients can observe the treatment outcome (e.g., their recovery from illness), they may never know whether a less expensive treatment could have achieved the same results. This leaves little room for patients to accumulate and share information about their experiences regarding physicians' conduct. Furthermore, due to this information asymmetry, physicians are often inclined

¹I thank Christian Rojas and Rong Rong for their invaluable guidance and advice. I also thank Juan-Camilo Cárdenas, Matt Woerman, John Spraggon, and Xiaolu Wang, as well as conference participants at ESA 2022, New England Experimental Economics Workshop (NEEEW) 2022, and UMass ResEcon Graduate Conference 2021 for thoughtful comments and feedback. Financial support from the Graduate School and the Department of Resource Economics at the University of Massachusetts Amherst is gratefully acknowledged.

²In addition to medical services, many other markets for professional services, such as automobile repair services (Schneider, 2012), taxi rides (Balafoutas et al., 2013), and management consulting (Craig, 2005), also exhibit properties of credence goods and, to some extent, suffer from market inefficiencies discussed in this study. Hence, the findings of this study are also insightful for these markets.

to recommend and provide services that exceed what is truly necessary for treating the disease, commonly known as the “overtreatment problem”. This issue is highly prevalent in the healthcare market. According to an American Medical Association survey, in the U.S. healthcare market alone, an average of 20.6% of medical care was deemed unnecessary, including 22.0% of prescription medications, 24.9% of tests, and 11.1% of procedures (Lyu et al., 2017).

Extensive studies have found that physicians’ decisions to overtreat are usually associated with greater marginal profits than providing appropriate treatment.³ A salient example can be found in Johnson and Rehavi (2016). Their study suggests that compared with physician mothers, nonphysician mothers are more likely to receive a C-section with a reimbursement rate higher than normal deliveries. In another instance, evidence from prescription drugs in China and Japan shows that physicians tend to prescribe and dispense costly drugs when they can personally benefit from such decisions. More broadly, evidence of overtreatment driven by financial incentives among physicians in a variety of medical settings can be found in Domenighetti et al. (1993), Delattre and Dormont (2003), Brownlee (2010), and Gottschalk, Mimra and Waibel (2020). In all these cases, physicians’ overtreatment not only leaves more costly medical bills to patients but also generates large amounts of wasted resources as more complex tests and treatments are typically costlier.⁴ In 2010, the Institute of Medicine reported that “unnecessary services” have become the primary cause of waste in the U.S. healthcare market (McGinnis et al., 2013). Therefore, the main objective of this study is to identify an efficient approach to restraining overtreatment fueled by pecuniary motivations.

³Physicians’ fear of malpractice (Lyu et al., 2017) and patients’ medical insurance coverage (Huck et al., 2016) are two additional factors that can lead to the overtreatment problem.

⁴Brownlee (2010) notes that overtreatment may make medical errors more likely, because the higher the volume of care a patient receives, the greater the odds are that somebody, somewhere, will make a mistake. As this study concentrates on the cost inefficiency of overtreatment, I assume that there are no adverse health effects from being overtreated.

There is an intuitive and widely accepted solution to mitigate physicians' overtreatment behavior, allowing patients to search for a second opinion. That is, patients are permitted to elicit multiple recommendations from different physicians and choose the one that is most cost-effective in treating their problem.⁵ Health insurers and government legislation in the U.S. and many European countries recommend this strategy in cases where patients are skeptical about the necessity of the treatment recommendation they received (Hu, 2017; Pieper, Hess and Mathes, 2017). There are two potential reasons why patient search can alleviate overtreatment and generate higher market efficiency. First, patients' ability to search can increase the intensity of competition among physicians. When a physician recommends expensive treatment, the patient is more likely to visit a different physician. This mechanism should have its intended benefits, whether the patient seeks treatment only once (what I refer to as "one-shot" interactions) or in cases when patients might seek treatment multiple times (what I refer to as "repeated market interactions"). A key difference between one-shot and repeated interactions is that a physician's reputation building does not exist (or matter) in the former but does in the latter. In repeated interactions, patients can avoid future visits once a dishonest physician is identified. The concern for future business may provide physicians with an additional incentive to provide truthful treatment recommendations. In light of this, I argue that reputation building has the potential to increase the effectiveness of patient search for curbing overtreatment (and, thereby, promoting market efficiency even further).

As the real-world healthcare market often involves repeated interactions, both competition intensity within a market period (facilitated by the role of patient search) and physicians' reputation building (facilitated by patients' ability to access health-

⁵Another advantage of searching for multiple opinions is to avoid incorrect diagnoses. In this study, I focus on its merit of restraining overtreatment only by excluding the possibility of misdiagnosis from the physician's choice set.

care providers' prior treatment recommendations) can be important factors in determining a market's level of observed overtreatment. This makes a laboratory experiment an ideal tool for identifying and quantifying the effectiveness of reputation by comparing its combined effects with patient search with the effect of patient search alone. To achieve this goal, this study designs a multi-seller-multi-buyer credence goods market in which informed sellers have the incentive to overtreat and uninformed buyers can always search for an additional treatment recommendation from a different seller. Under the baseline condition, I turn off the possibility of a reputation mechanism by randomly reshuffling seller IDs after each market period. The random ID design effectively turns a repeated game into many one-shot interactions where reputation-building is impossible.⁶ This allows me to measure the effectiveness of the market competition mechanism (i.e., as facilitated by free choice of physician and patient search) alone in reducing overtreatment. In the "reputation" condition, I fix the sellers' IDs and make the history of buyers' past transactions exclusively visible to them.⁷ Through this, a seller can build a reputation (or risk ruining it) by choosing proper treatment (or overtreatment) for buyers. Note that the aforementioned market competition mechanism also exists in this treatment. The study experiments with general credence goods framing to avoid a framing effect that may confound the main results. For the purpose of this study, I will refer to sellers as physicians and buyers as patients.

In the experiment, I ensure that patient search is an essential part of market design. It is because reputational incentives alone are weak in disciplining sellers in credence goods markets, as demonstrated by prior studies. This is not surprising since buyers in these markets can only observe a series of past recommendations with

⁶While it is possible to run one-shot games with different sets of buyers and sellers each time, the cost of such design could be prohibitive.

⁷This is different from displaying past history of all market transactions, which is harder to achieve in healthcare markets for privacy reasons.

no reputation value, and an honest high-cost treatment recommendation can never be distinguished from an overtreatment. The unique characteristic of credence goods determines that buyers need to have an additional channel to monitor sellers' honesty for the reputation mechanism to function effectively (Fong, Liu and Meng, 2022; Gerlach and Li, 2022). In the healthcare market, as patient search is a common practice, the detection of overtreatment becomes feasible. This occurs when patients receive contradictory treatment recommendations from various physicians. The availability of patient search is crucial for establishing a meaningful reputation mechanism.

In addition, the level of search costs is critical to my investigation of the effect of reputation building. In the real-world healthcare market, not all patients have their expenses for obtaining second opinions covered by insurance. Moreover, even if the diagnosis cost is covered, patients still incur additional expenses related to transportation and the time invested in seeking a second opinion. Undoubtedly, the willingness of patients to search for second opinions is constrained as search costs increase, which may diminish the disciplining effect of reputational concern. I hypothesize that prohibitively high search costs can practically convert a market with patient search and reputation discovery into one without. To empirically test this hypothesis, I introduce another treatment dimension in my experiment by varying the level of search costs. I manually create two conditions: one with a low search cost and another with a high search cost. The findings hold significant relevance for the ongoing discourse on healthcare inequality, as patients in lower-income brackets and with poor insurance coverage may have higher search costs. This may prevent them from receiving full benefits in terms of cost-effective medical treatments owing to prohibitive search costs.

The main conclusion of my study is that the reputation mechanism holds significance in bolstering the efficacy of patient search for the purpose of mitigating overtreatment. The average overtreatment level decreases from 38.54% to 21.67%

when reputation is at play. However, this reduction is only statistically detectable in the low search cost condition, which confirms my hypothesis that search costs need to be sufficiently low for overtreatment detection to be sizeable and, therefore, for reputation building to have a bite. Furthermore, I observe that patients search less frequently when physicians can build a reputation. The reason for this is that patients receive low-cost treatment recommendations more often on their first visit, which in turn reduces their need for costly searches. As a result of fewer overtreatments and searches, market efficiency significantly rises when reputation exists.

This study contributes to the literature by experimentally analyzing how reputation impacts overtreatment and market efficiency in healthcare markets. Contrary to all the earlier studies, which found a null effect of reputation in credence goods markets, this study highlights that patient search is necessary to establish a meaningful reputation mechanism in such markets. To the best of my knowledge, this study is the first to quantify the positive effect of reputation on reducing wasteful overtreatment in an experimental healthcare market. The comparison between low and high search costs in this environment also sheds light on a previously unexplored channel where the variation in search costs (e.g., insurance coverage, health care budget) can further exacerbate the problem of health inequality.

The remainder of this paper is organized as follows. Section 2 reviews related literature. Section 3 presents the experimental design of the study. Section 4 provides the hypotheses. Subsequently, Section 5 reports the experimental results. Finally, Section 6 concludes the paper and discusses the policy implications.

1.2 Literature Review

A large and growing body of literature has studied the different mechanisms that discipline seller behavior in credence goods markets.⁸ Two related strands of literature address the overtreatment problem in such markets from different angles. The first group of studies examines the effect of buyer search in a one-shot market, while the second focuses on the formation of reputation in repeated interactions.

Several studies have analyzed the impact of buyer search in a one-shot interaction. Wolinsky (1993) provides an earlier theoretical model describing the incentive to overtreat in credence goods markets. His model demonstrates that introducing costly buyer searches can prevent sellers from overtreating. Sellers in this model choose not only the treatment recommendations but also their corresponding prices. One of the model's predictions is a symmetric equilibrium, which is consistent with the experimental results of this study, where the amount of overtreatment is determined by the level of search costs⁹. Mimra, Rasch and Waibel (2016*b*) modify Wolinsky's model by focusing on cases in which the treatment prices are predetermined. They conduct the first laboratory study to examine the impact of costly buyer searches on the level of overtreatment and market efficiency in a one-shot market environment. They find that when the search cost is sufficiently low, introducing the possibility of searching for a second opinion significantly reduces the level of overtreatment, thereby improving market efficiency (high search costs significantly dampen this effect). This result is replicated by Agarwal, Liu and Prasad (2019) in a slightly different laboratory setting. In Agarwal, Liu and Prasad (2019), buyer search only acts as an information source, and buyers cannot undergo treatment with second-selected sellers. It is

⁸Dulleck and Kerschbamer (2006), Kerschbamer and Sutter (2017), and Balafoutas and Kerschbamer (2020) provide comprehensive reviews of the literature on when and how credence goods can be provided efficiently.

⁹An additional asymmetric equilibrium exists where some sellers specialize in being the ones who conduct only low-cost procedures, while the rest of the sellers can conduct either high-cost or low-cost procedures.

worth emphasizing that the market designs of Mimra, Rasch and Waibel (2016*b*) and Agarwal, Liu and Prasad (2019) use a random matching procedure between market periods. The absence of buyers' choice of which seller to visit removes the incentive for sellers to build reputation for future businesses. In this study, I complement their work by examining how buyer searches can improve credence goods market efficiency through the critical reputation channel.

A plethora of research suggests that reputation is crucial in facilitating trade in markets with asymmetric information. Most of these studies look at experience goods markets, where buyers know exactly what they need ex-ante and learn the value of the goods or services through consumption.¹⁰ In the case of experienced goods markets, past sales provide an objective metric for gauging a seller's product quality, thereby reducing fraud. This differs from credence goods, where information asymmetry persists even after consumption. This sharp distinction between experience goods and credence goods is captured in the results of the laboratory and field experiments. Reputation has been shown to reduce moral hazard and improve the market outcomes for experience goods (Dellarocas, 2006; Huck, Lünser and Tyran, 2012; Tadelis, 2016). However, in the context of credence goods, reputation alone appears to be ineffective in limiting sellers' overtreatment or improving market efficiency (Lab experiment: Dulleck, Kerschbamer and Sutter, 2011;¹¹ Huck et al., 2016; Field experiment: Schneider, 2012).¹² As mentioned earlier, this study highlights the unique characteristics of credence goods and explores the potential of reputation in

¹⁰See Bar-Isaac, Tadelis et al. (2008) for a literature review on reputation and trust in experience goods markets.

¹¹In their set-up, reputation-building incentivizes sellers to provide their buyers with sufficient services, but it could not restrain sellers from overtreating their buyers.

¹²The results are contradictory to the theoretical predictions in Wolinsky (1993) and Frankel and Schwarz (2014). In these two theoretical papers, the authors demonstrate that sellers' incentives to overtreat their buyers can be corrected if buyers give more business to sellers who previously provided low-cost treatments.

the more fitting institutional environment of such goods. The results of this study provide empirical evidence through a controlled laboratory experiment, indicating that it is necessary to offer buyers a way to monitor physicians' honesty for a reputation mechanism to operate effectively.

Two recent theoretical studies are closely related to this study. Fong, Liu and Meng (2022) model an infinitely repeated game in which long-lived sellers interact with short-lived buyers who can verify whether the selected seller makes an unnecessary treatment recommendation by searching for a costly second opinion. A seller loses all future businesses if his overtreatment is detected by a buyer. Moreover, new buyers observe the past transaction history of all buyer-seller pairs. Given that both of their design features are unlikely to be held in actual healthcare markets, I introduce two major changes in my experiment. First, punishment for dishonest physicians is endogenously determined by the patients. Second, the experimental design allows patients access to past transactions and search history involving themselves (and not other patients). This feature is consistent with the fact that other patients' experiences with a physician are protected by HIPAA authorization and are unlikely to be publicly available.

Gerlach and Li (2022) theoretically and experimentally study the level of overtreatment in both monopoly and duopoly markets. They also investigate two different scenarios in a duopoly market, with and without buyer search. They do not detect any significant impact of introducing buyer search from their experimental results, despite the efficiency-enhancing effect predicted by theory. A critical market feature that may contribute to this result is that buyers in all markets are allowed to purchase any treatment independent of the seller's recommendation. This "freedom of choice" design enables buyers to detect overtreatment whenever a high-cost recommendation is followed by a buyer's successful attempt to disregard such a recommendation. In other words, their experiment points out that buyer search can be redundant when

freedom of choice is possible. Results from Gerlach and Li (2022) apply directly to credence goods such as business consulting and repair services where “freedom of choice” for buyers is common. However, in healthcare markets, patients can rarely undergo procedures or tests that are not recommended by a physician. By comparison, the experimental design of this study provides a more suitable environment for studying the effect of reputation in the healthcare domain.

1.3 Experimental Design

1.3.1 Market Procedure

I construct a basic market structure with exogenous prices¹³, following the design of Mimra, Rasch and Waibel (2016b)¹⁴. Each market period consists of four patients and four physicians¹⁵. At the beginning of each session, the subjects are randomly assigned to be either a patient or a physician. Assignments are fixed for the entire experiment. Each experimental session has 20 market periods.

The decision sequence for patients and physicians in a particular market period is as follows. First, each patient independently receives a randomly determined type of problem. The problem can be either a major problem with probability $h=0.25$ or a minor problem with probability $(1-h=0.75)$. Patients have no information on the type of problem they have. Thus, in the next step, each patient must select one of the four physicians to provide a treatment recommendation. Simultaneously, each physician is asked to provide a treatment recommendation for each patient after observing the

¹³Fixed prices are common in the U.S. healthcare market where prices are set as a result of a centralized bargaining process (Sülzle and Wambach, 2005)

¹⁴Equilibria under our baseline conditions (no reputation) are characterized in their study.

¹⁵I decide to employ four participants per role because collusions are rare in markets with four or more sellers (Brandts and Potters, 2018).

type of problem they have¹⁶. The treatment can be either low-cost or high-cost. A high-cost treatment can fix major and minor problems, but a low-cost treatment can only fix minor ones. Note that physicians are liable to cure their patients in our setting¹⁷, so they cannot provide low-cost treatment recommendations to patients with major problems. Given that the primary focus is on the physicians' incentives for overtreatment due to monetary rewards, I further assume that, in the experiment, physicians can accurately diagnose patients' problems without incurring any cost. After a patient selects a physician, he observes the treatment recommendations offered by the physician. Consequently, the patient has two options: (1) accept the treatment recommendation and let the selected physician perform the corresponding treatment, or (2) pay a search cost K to select another physician from the three remaining physicians and undergo the treatment performed by the second selected physician¹⁸.

The payoff function π for each patient j is defined as the difference between a fixed value $V=130$ points, utility of recovering from the unknown problem, and price of the recommended treatment P_t ($P_H=115$ points; $P_L=75$ points). Clearly, patients prefer low-cost treatment recommendations to high-cost ones. Moreover, if patients choose to search for a second opinion, they pay an additional search cost of K .

$$\pi_j = \begin{cases} V - P_t & \text{if the patient } j \text{ accepts} \\ V - P_t - K & \text{if the patient } j \text{ searches for a second recommendation} \end{cases} \quad (1.1)$$

¹⁶I implement the strategy method for two reasons: (1) so that physicians do not know whether a patient is on her first or second search, a condition required by the model in Mimra, Rasch and Waibel (2016b); (2) it collects more decisions from physicians per period.

¹⁷This assumption is valid in healthcare markets because leaving patients uncured not only violates the Hippocratic Oath and the World Medical Association's Declaration of Geneva but also incurs a medical malpractice lawsuit.

¹⁸For the simplicity of the model and the experiment, a patient must undergo the treatment with the second selected physician if he searches for a second opinion. I assume that the search cost increases dramatically with any further searches so that it would never be optimal for patients to conduct further searches after receiving a second opinion.

Another important feature of the above patients' payoff function is that if a patient does not conduct a second search, he/she cannot detect whether the physician provides an appropriate treatment recommendation. However, if a patient chooses to search for a second recommendation, discovering a physician's dishonesty becomes possible, although not guaranteed. Detection only occurs when the first recommendation is high-cost, whereas the second recommendation is low-cost.

Physician i receives payoff e which is calculated as the difference between the price and cost of the accepted treatment by each patient j :

$$e_i = \sum_{j=1}^n (P_{ijt} - C_{ijt}) \quad (1.2)$$

where t indicates one of the particular types of treatment accepted, which is either high cost (H) or low cost (L). Specifically, a high-cost treatment is priced at $P_H=115$ points and costs a physician $C_H=80$ points. Low-cost treatment was priced at $P_L=75$ points and $C_L=60$ points. These parameters are chosen so that the physicians' profit for a high-cost treatment ($115-80=35$ points) is greater than that for a low-cost treatment ($75-60=15$ points). This is to ensure that physicians have incentives to overtreat patients, which refers to the case in which physicians provide a high-cost treatment recommendation to patients with only a minor problem in this experiment. Moreover, in order to mimic real-world markets where overtreatment always leads to an efficiency loss, the above payoff parameters ensure that the physicians' gain from overtreatment is smaller than the net loss to the patients. Note that a physician's payoff is summed over the number of patients who choose to accept their recommendations. If any patient does not accept a physician's recommendation, the payoff in this period is zero. For simplicity, I impose verifiability in this experiment so that physicians must provide the announced treatment. All of the above information is common knowledge, except for the patient's types.

At the end of each period, both physicians and patients observe their own payoffs in the current period. In addition to the payoffs, physicians also observe the number of patients visited and the number of high-cost and low-cost treatments they performed.

1.3.2 Treatment Design

This study aims to explore the impact of reputation on physicians' overtreatment and examine whether the impact of reputation varies with the level of patients' search costs. To this end, a 2×2 factorial design is applied. The treatments differ in two dimensions: the possibility of reputation building and the search cost. I parameterize the search cost K to be equal to either 7 points for the low search cost treatment or 14.5 points for the high search cost treatment.

Based on the theoretical model in Mimra, Rasch and Waibel (2016*b*), when reputation building is impossible, the low search cost condition ($K=7$) produces two types of equilibrium predictions: pure-strategy equilibrium and mixed-strategy equilibrium. In the pure-strategy equilibrium, physicians always overtreat patients with minor problems, and patients never search for a second opinion. In the mixed-strategy equilibrium, physicians' tendency to overtreat is balanced by the patient's tendency to search. Given the above parameters, one would observe that physicians overtreat patients with a minor problem in either 74.69% or 7.81%, and patients search for a second opinion in either 99.69% or 59.81%, respectively, when they receive a major treatment recommendation.

Under the high search cost condition ($K=14.5$), if reputation is absent, the unique equilibrium overlaps with the pure-strategy equilibrium mentioned above, with complete overtreatment and no search. The intuition is that never searching for a second opinion (independent of physicians' overtreatment level) is optimal for patients when the search cost is prohibitively high. Thus, physicians' best response is to always overtreat patients with a minor problem.

The key treatment variation in this study is achieved by switching the reputation mechanisms on and off. For the no-reputation (NR) condition, physicians’ ID numbers are shuffled after each market period. In these treatments, patients cannot distinguish between physicians; therefore, they can only choose randomly in the following market period. In the conditions with reputation (R), physicians’ IDs are fixed between periods¹⁹. Starting from the second period, patients can browse their own past interactions with physicians in their first and second searches when making decisions²⁰. *Table 1.1* summarizes the four treatments.

Table 1.1: Experimental Set-up: Conditions

	Low Search Cost (7)	High Search Cost (14.5)
No Reputation	NR ₇	NR _{14.5}
Reputation	R ₇	R _{14.5}

As mentioned in the introduction, all treatments in this study allowed patients to search for a second recommendation. This is because searching for a second recommendation is the only opportunity for buyers to assess a seller’s dishonesty in a credence good setting. Instead of removing the opportunity to search, my design implements the high search cost condition, which theoretically prevents all patients from conducting searches but still leaves some flexibility, so it is possible to observe choices that do not conform to the theory.

¹⁹Note that in all of the treatments, patients’ IDs are shuffled at the beginning of each period. This set-up prevents physicians from engaging in complex and strategic play with certain identified patients.

²⁰A rational buyer is always aware of his/her personal trade history. However, in case participants in the experiment might forget or misremember parts of their histories, I display this information as a reminder in the conditions with reputation, following the previous studies.

1.3.3 Experimental Protocol

I implemented a between-subject design such that each subject participated in only one of the four treatment conditions. There were four sessions for each treatment, with a total of 128 subjects. I recruited subjects from the University of Massachusetts Amherst subject pool using ORSEE (Greiner, 2015). I conducted the experiment using an online version of z-Tree (Fischbacher, 2007; Duch, Grossmann and Lauer, 2020).

At the beginning of each session, participants were invited to a Zoom meeting room with their audio and video turned on. They were asked not to leave the meeting room until the end of the session. The experimenter read the instructions aloud, followed by a set of control questions to ensure that everyone understood the instructions (see Appendix A.1 for the instructions and A.2 for the control questions). During the experiment, participants were assigned to an individual breakout room so that the experimenter could assist them privately. After the experiment, a short questionnaire was used to collect demographic information (see Appendix A.3).

The payment included a show-up bonus of \$5 and the cumulative payoff from the decisions made in all 20 periods. The participants received their payoff with an Amazon eGift card²¹ based on the exchange rate of 50 points = \$1. The average payoff per participant was \$16.47. The average session length was approximately one hour.

1.4 Hypotheses

In the analysis, four aspects are of prime interest: (i) whether the reputation mechanism reduces the level of physicians' overtreatment, (ii) whether the reputation mechanism reduces patients' search activities, (iii) whether introducing the reputa-

²¹The gift card was sent to each participant within three hours after the session ended.

tion mechanism boosts market efficiency, and (iv) whether the impact of reputation depends on the level of search costs.

Based on the findings in the literature, I present the following hypotheses regarding how reputation impacts rates of overtreatment, patient search, and efficiency. I also hypothesize how the level of search costs affects the role of reputation.

Hypothesis 1: The Reputation Condition (relative to No Reputation) will:

- a) decrease rates of overtreatment, b) decrease patient's need to search for a second opinion, and c) improve market efficiency.

The availability of patient search provides an efficient way for patients to monitor physician honesty and makes the reputation mechanism more informative. In repeated interactions, physicians' decisions to overtreat not only induce their patients to search for second opinions and undergo treatment with their competitors but also risk future business with their patients once their dishonesty is detected. The expectations of future businesses further incentivize physicians to provide treatment recommendations truthfully. Again, in this game, a rational patient who receives a low-cost treatment recommendation should accept it with certainty because liability is applied to the market. Under reputation conditions, patients are more likely to receive a low-cost treatment recommendation on their first visit, and as a result, their need for a second opinion decreases. Finally, less overtreatment and fewer patient searches in the market lead to higher efficiency.

Hypothesis 2: The reputation effects will decrease with the increase of search costs.

Under high-search-cost conditions, patients are less willing to conduct a second search when they receive a high-cost treatment recommendation during their first visit. In this case, the chance of not getting caught and not being punished decreases the expected penalty, which increases physicians' dishonesty. With increased search costs, the reputation mechanism becomes less informative and effective in correcting physicians' incentives for overtreatment. An extremely high search cost will eventually remove patient search as well as reputational effects from the market, turning the game back to the conditions with reputation, as in Dulleck, Kerschbamer and Sutter (2011) and Mimra, Rasch and Waibel (2016*a*), where a buyer who goes to a seller must be treated by that seller.

1.5 Results

The following section examines the impact of the reputation mechanism on subjects' decisions in the credence goods market. The study conducts its analysis based on three main outcomes of the market: (1) physician overtreatment, (2) patient search, and (3) market efficiency. I first report treatment comparisons for each outcome variable using two-tailed Mann-Whitney U tests. For these tests, I consider decision dependency by taking the average of the measures across all individuals and all market periods as one independent observation. I accompany the nonparametric results with panel data analysis using random-effects regressions clustered at the market level. I also control for decision time trends and basic subject demographics, including gender, age, GPA, and prior enrollment in economics courses, in the regression models.

1.5.1 Overtreatment

I first investigate whether the reputation mechanism reduces the level of overtreatment in the market. Again, overtreatment, in this context, refers to the provision of

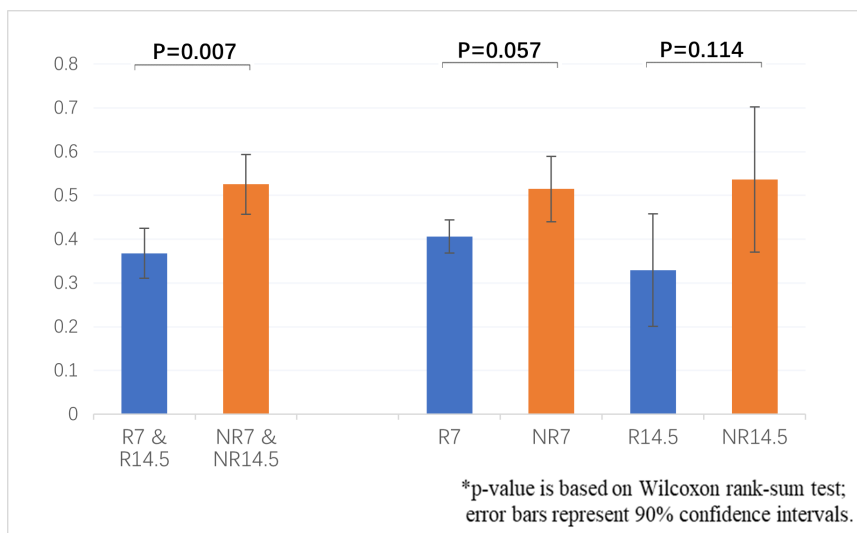
high-cost treatments for minor problems. I employ two distinct methods to measure the level of overtreatment. First, I analyze the “overtreatment strategy” based on all the treatment recommendations for minor problems submitted by a physician in her strategy profile.²² For each physician, the level of overtreatment in her strategy during a given period is measured by the frequency of her recommending high-cost treatments to patients with minor problems. This measurement produces a clear indicator of a physician’s honesty level. Second, I measure the “actual overtreatment” experienced by the patients. It factors in the patients’ choices of physicians. For each patient, the actual overtreatment value of 1 (0) corresponds to a situation in which she is (not) eventually overtreated during a given period when she has a minor problem, and it is null if she has a major problem during a given period. Under the above definitions, “actual overtreatment” occurs less frequently than “overtreatment in strategy,” the difference of which reflects the ability of patients to detect dishonest recommendations and avoid overtreatment by selecting a different physician.

1.5.1.1 Overtreatment in Strategy

Figure 1.1 shows the average level of overtreatment in strategy across all four treatments and between the two reputational conditions (pooled across two search costs). Physicians choose to overtreat patients 52.54% of the time under conditions without reputation, a highly comparable result to that found in Mimra, Rasch and Waibel (2016b, 52.7% overtreatment). When the reputation mechanism is introduced, physicians significantly reduce their tendency to recommend major recommendations for minor problems: only 36.76% of the recommendations are overtreated (R_7 & $R_{14.5}$ vs. NR_7 & $NR_{14.5}$, $p=0.007$). This provides strong support for Hypothesis 1.

²²Recall that in the experimental design, a physician needs to make recommendations for each patient, even if the patient does not end up selecting her.

Figure 1.1: Level of Overtreatment in Strategy



Next, I examine the reputation effect under different search costs. The main result is largely driven by the significant difference in the low-search-cost condition (R_7 vs. NR_7 , $p=0.057$). Under the high search cost condition, the effect of reputation on overtreatment is still negative but not statistically significant at the 10% level ($R_{14.5}$ vs. $NR_{14.5}$, $p=0.114$). This result is in line with Hypothesis 2, suggesting that the reputation effect is diminished with a high search cost compared with a low search cost. As mentioned earlier, the intuition behind this is that when the search cost is prohibitively high, such as in the case of $R_{14.5}$ and $NR_{14.5}$, patients would be less willing to conduct costly searches. This, in turn, reduces the chance of detecting physician dishonesty and slows the reputation-building process. In Section 1.5.2, I present the results regarding how search costs affect search frequency.

The random effects regression results in *Table 1.2* confirm that physicians are less likely to overtreat patients when a reputation mechanism is present. In particular, under full specification (Model 3), the percentage of overtreatment drops by 10.7% when the reputation mechanism is introduced. The effect is statistically significant at the 1% level, with or without additional controls on basic demographics and the

Table 1.2: Random effects panel OLS: level of overtreatment in strategy

	Model 1	Model 2	Model 3
Reputation	-0.15781*** (0.045)	-0.10833*** (0.032)	-0.10683*** (0.035)
High Search Cost	-0.02734 (0.045)	0.02214 (0.069)	0.02364 (0.063)
Rep x High		-0.09896 (0.086)	-0.09857 (0.083)
Period			-0.00105 (0.003)
Control			✓
Constant	✓	✓	✓
Observations	1280	1280	1280

Notes. Standard errors (reported in parentheses) are robust and clustered at the market level.

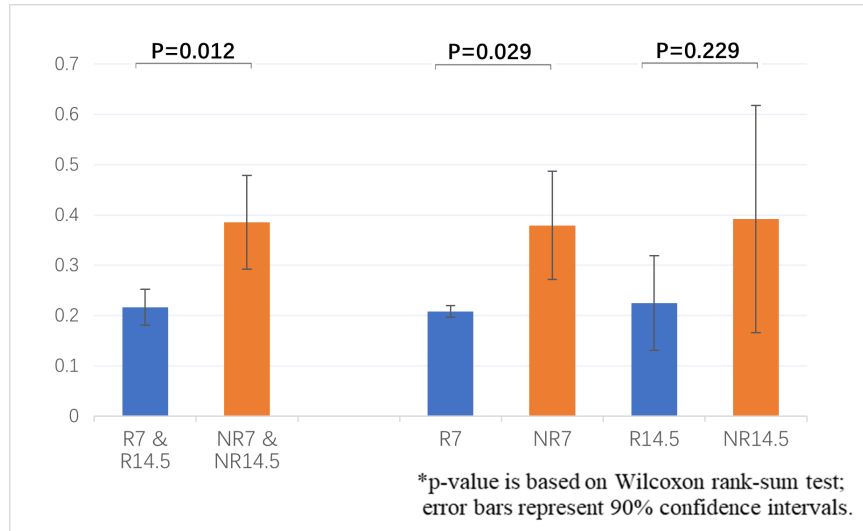
***:p<0.01, **:p<0.05, *:p<0.1

market period. These findings are consistent with Hypothesis 1. However, the interaction effect between the two treatment dimensions is not statistically significant at the 10% level, which contradicts Hypothesis 2.

1.5.1.2 Actual Overtreatment

Compared to the overtreatment decisions submitted by physicians in the strategy format, the level of actual overtreatment provides a more accurate representation of the prevalence of overtreatment in the market. Under conditions without a reputation, patients experience overtreatment from their physicians approximately 38.55% of the time. The results are also comparable to the findings of Mimra et al. (2016b, 36.46% overtreatment). As shown in *Figure 1.2*, patients are significantly less likely to be overtreated in conditions with a reputation than in those without. The average level of actual overtreatment drops to 21.67% after the reputation mechanism is introduced (R_7 & $R_{14.5}$ vs. NR_7 & $NR_{14.5}$, $p=0.012$). When examining the reputation effect under two different search costs separately, I conclude that the effect is primarily attributed

Figure 1.2: Level of Actual Overtreatment



to the reduction in the level of actual overtreatment in the conditions with a low search cost (R_7 vs. NR_7 , $p=0.029$). There is no significant difference in the actual overtreatment level between $R_{14.5}$ and $NR_{14.5}$ ($p=0.229$). This finding is consistent with Hypothesis 2.

The results from the random-effects probit regressions, as shown in *Table 1.3*, further confirm the significant and negative impact of reputation on the level of actual overtreatment. The marginal effect evaluated while holding all other variables at their mean suggests that patients are 21.6% less likely to receive overtreatment when reputation-building is possible. When comparing this result with the reputation effect on overtreatment in strategy, as shown in Section 1.5.1.1, I conclude that about half of the actual overtreatment is due to the change in recommendations selected by the physicians: They are 10.7% less dishonest on average. The other half of the effect results from patients being able to identify honest physicians and interact with them. Finally, I do not observe any interaction effect between the two treatment dimensions, which does not support Hypothesis 2.

Table 1.3: Random effects panel Probit: level of actual overtreatment

	Model 1 (dy/dx)	Model 2 (dy/dx)	Model 3 (dy/dx)
Reputation	-0.16673*** (0.051)	-0.16875*** (0.039)	-0.21546*** (0.047)
High Search Cost	0.0133 (0.050)	0.01136 (0.086)	-0.01292 (0.083)
Rep x High		0.00411 (0.096)	0.02604 (0.094)
Period			-0.00305 (0.004)
Control			✓
Constant	✓	✓	✓
Observations	960	960	960

Notes. Standard errors (reported in parentheses) are robust and calculated using the delta method.

The coefficients show the marginal effects of each variable, with all other variables held at their means.

***:p<0.01, **:p<0.05, *:p<0.1

1.5.2 Patient Search

I also use two methods to measure patients' search activities: conditional and unconditional search rates. Conditional search applies exclusively when a patient receives a high-cost recommendation on her first visit. The measure takes the value of 1 if a patient seeks a second opinion and 0 otherwise. The conditional search rate reflects the frequency of patients conducting a second search when faced with a high-cost recommendation, providing insight into their trust in their physicians in such cases.²³ Unlike conditional search, unconditional search takes into account patients' search decisions regardless of the type of recommendation they received on their first visit. The unconditional search rate is computed by dividing the total number of patient searches by the total number of transactions. This measure reflects the

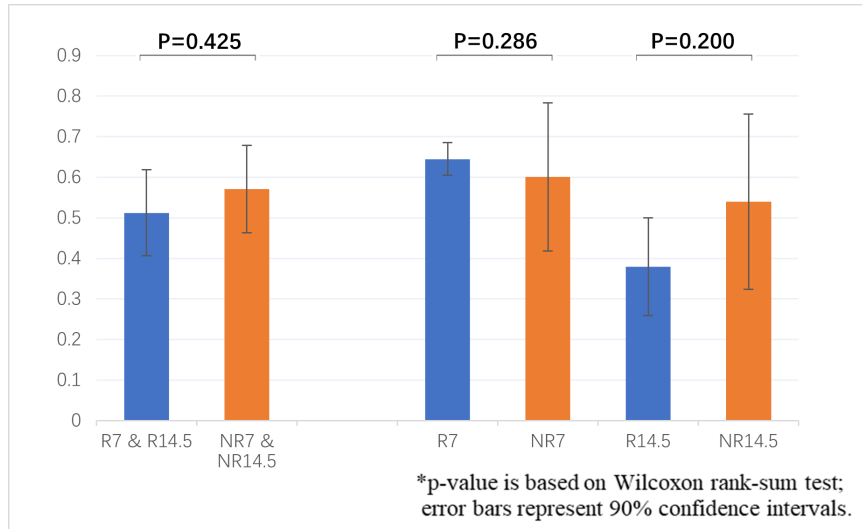
²³Recall, in our experimental setting, if a patient observes a low-cost recommendation, she can infer that the physician must be offering the proper treatment honestly.

amount of inefficient search cost incurred and, therefore, is directly linked to market efficiency.

1.5.2.1 Conditional Search

Figure 1.3 shows the average search rates conditional on the high-cost recommendations. Upon receiving a high-cost treatment recommendation on their first physician visit, patients search less often for a second opinion in conditions with a reputation than in those without a reputation (51% vs. 55.4%). However, this reduction is not statistically significant (R_7 & $R_{14.5}$ vs. NR_7 & $NR_{14.5}$, $p=0.425$). There are also no significant differences for R_7 vs. NR_7 ($p=0.286$) and $R_{14.5}$ vs. $NR_{14.5}$ ($p=0.2$). The results indicate that the presence of a reputation mechanism does not alter patients' trust in physicians. As the results in Section 1.5.1 show a significant improvement in physicians' honesty, it is surprising that patients do not seem to incorporate physician behavior change into their best response.

Figure 1.3: Conditional Search Rate



The panel probit regressions (see Table 1.4) report the results for the two dependent variables. In the first three models, the dependent variable is whether or not a patient searches for a second opinion when receiving a high-cost recommendation.

Table 1.4: Random effects panel Probit: patient search

	Conditional Patient Search			Unconditional Patient Search		
	Model 1 (dy/dx)	Model 2 (dy/dx)	Model 3 (dy/dx)	Model 4 (dy/dx)	Model 5 (dy/dx)	Model 6 (dy/dx)
Reputation	-0.03370 (0.063)	0.05981 (0.074)	0.06410 (0.082)	-0.10153*** (0.036)	-0.05870 (0.049)	-0.06830 (0.055)
High Search Cost	-0.14847** (0.063)	-0.05967 (0.103)	-0.02700 (0.114)	-0.09231*** (0.035)	-0.04952 (0.055)	-0.03829 (0.064)
Rep x High		-0.18363* (0.110)	-0.21553* (0.126)		-0.08451 (0.059)	-0.09659 (0.069)
Period			-0.00321 (0.003)			-0.00557** (0.003)
Control			✓			✓
Constant	✓	✓	✓	✓	✓	✓
Observations	724	724	724	1280	1280	1280

Notes. Standard errors (reported in parentheses) are robust and calculated using the delta method. The coefficients show the marginal effects of each variable, with all other variables held at their means. ***:p<0.01, **:p<0.05, *:p<0.1

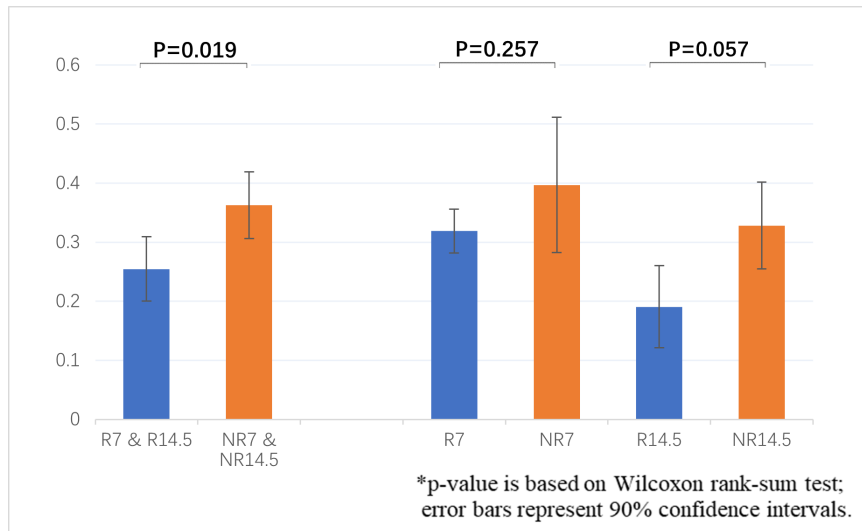
The regressions report a similar null result on the effect of reputation, which is consistent with the results from the nonparametric analysis. In Model 1, the results show that patients are less willing to search for a second opinion when the search cost is high, which is not surprising given the Law of Demand. In Models 2 and 3, adding the treatment interaction term and other controls, I find that the interaction effect is significant at the 10% level: patients are 21.55% less likely to search for a second opinion if the physicians' reputation is observable and the search cost is high. The results suggest that patients may perceive reputation and search as substitutes: patients rely more on the physician reputation system if the cost for additional search increases, and vice versa.

1.5.2.2 Unconditional Search

Figure 1.4 shows that reputation is vital in reducing costly patient searches. The unconditional search rate pooled over both search cost treatments drops from 36.25%

without reputation to 25.47% with reputation. The effect is highly significant (R_7 & $R_{14.5}$ vs. NR_7 & $NR_{14.5}$, $p=0.019$). Clearly, the reduction is mainly driven by the decisions made in the high search cost condition ($R_{14.5}$ vs. $NR_{14.5}$, $p=0.057$). Under the low search cost condition, patients' search decisions entail higher noise, which implies that the comparison of the unconditional search rate between reputational conditions is no longer statistically significant (R_7 vs. NR_7 , $p=0.257$). The above nonparametric results echo Hypothesis 2 and the result observed on overtreatment in Section 1.5.1: a high search cost leads to fewer searches, which makes reputational information difficult to accumulate.

Figure 1.4: Unconditional Search Rate



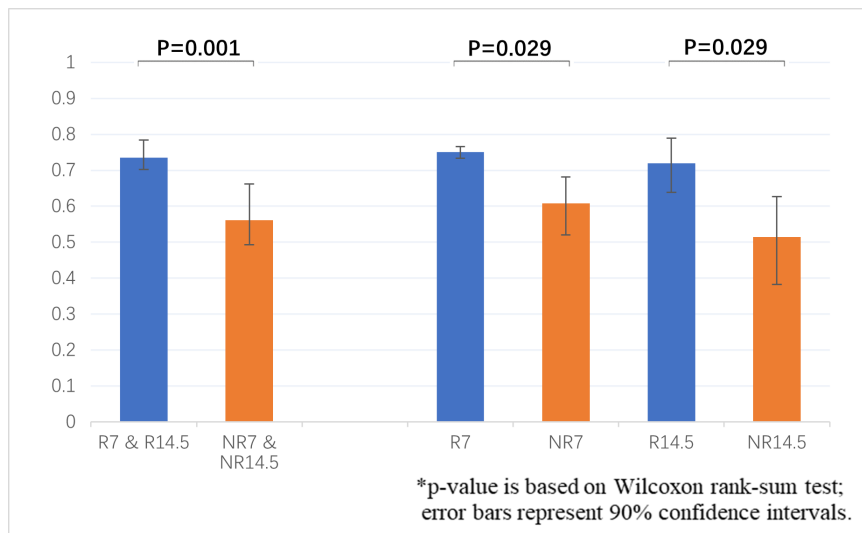
In *Table 1.4*, the dependent variable in the last three models is whether a patient searches for a second opinion. Accordingly, the results from Model 4 are consistent with those of the nonparametric test. Patients search for a second opinion 10% less often when they can observe their transaction histories. Moreover, they are less likely to search for a second opinion when the search cost is high. However, the significance of reputation and high search cost disappears when more factors are controlled for in

Models 5 and 6. There is no interaction effect between reputation and search costs on patients' search activities.

An additional comparison between the conditional and unconditional search rates helps shed light on the source of the reputation effect on costly searches. By comparing Models 1 and 4, I conclude that the effect of reputation on reducing costly searches is caused by the desirable change in physicians' recommendations to more honest ones. This is not due to a change in patients' trust in physicians.

1.5.3 Market Efficiency

Figure 1.5: Relative Market Efficiency



Based on my experimental setup, market efficiency is contingent upon actual overtreatment, search rates, and level of search costs. As reputation decreases overtreatment and reduces costly searches, I expect to see an improvement in market efficiency. To show this, I first compute the relative market efficiency by normalizing the sum of patients' and physicians' surpluses per market and per period in a $[0,1]$. More specifically, I achieve this by dividing the difference between the actual total surplus and the minimal possible surplus of the market by the difference between the maximum possible surplus and the minimum possible surplus of the market in a given

period. Hence, a relative efficiency of 0 represents the minimal possible surplus of the market and corresponds to a situation in which physicians always overtreat and patients always search for a second opinion; a relative efficiency of 1 represents the maximal possible surplus, which is achieved when physicians never overtreat their patients and patients never search for a second opinion.

As shown in *Figure 1.5*, market efficiency is significantly higher in conditions with a reputation than in those without. Relative market efficiency rises from 56.09% to 73.52% on average when the reputation mechanism is available (R_7 & $R_{14.5}$ vs. NR_7 & $NR_{14.5}$, $p=0.001$). The efficiency gain remains significantly positive after controlling for the level of search costs (R_7 vs. NR_7 , $p=0.029$; $R_{14.5}$ vs. $NR_{14.5}$, $p=0.029$). The highest efficiency is observed for R_7 (75.08%). This is not surprising, as the low search cost and reputation reduce wasteful searches and overtreatment, respectively.

Table 1.5: Random effects panel OLS: relative market efficiency

	Model 1	Model 2	Model 3
Reputation	0.17436*** (0.035)	0.14371*** (0.034)	0.17220*** (0.054)
High Search Cost	-0.06185* (0.035)	-0.09251 (0.060)	-0.03437 (0.059)
Rep x High		0.06132 (0.068)	0.01719 (0.061)
Period			0.00480 (0.004)
Control			✓
Constant	✓	✓	✓
Observations	320	320	320

Notes. Standard errors (reported in parentheses) are robust and clustered at the market level.

***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$

The results from the regression models in *Table 1.5* confirm the nonparametric findings and are robust to the inclusion of treatment interaction terms and demographic controls. The presence of reputation significantly improves relative market

efficiency by 17.22%. In addition, the results of Model 1 reveal a weakly significant negative relationship between search costs and market efficiency, but the effect disappears after I add more controls. Lastly, in both Model 2 and Model 3, I find no evidence indicating that the level of search costs influences the reputation effect on market efficiency.

1.6 Conclusion

Asymmetric information in healthcare services often leads to inefficient market outcomes, as physicians exploit their informational advantage by recommending unnecessary high-cost treatments. This study implements a lab experiment to investigate whether the opportunity for reputation building can effectively restrain overtreatment by physicians and improve market efficiency when patients can validate the necessity of prescribed treatments by seeking second opinions.

Previous experiments on credence goods studied the effect of providing patients with historical transaction data but required patients to accept any treatment recommendations and forbid patient search. As a result, patients have little chance of verifying whether a past recommendation is an overtreatment or a proper one. To tap into the full potential of the reputation mechanism, I relax this assumption and allow patients to seek a second opinion from a different physician. Complementing Mimra, Rasch and Waibel (2016*b*), the experimental results demonstrate that physicians' reputation concerns motivate them to provide treatment recommendations more truthfully in repeated interactions. This reduction in physician overtreatment also significantly reduces patients' need for costly searches. Taken together, market efficiency is dramatically improved after reputation is introduced. Although reputation in this study is induced through repeated interactions and personal histories, the findings provide valuable insights for designing a meaningful reputation mechanism in the healthcare market to tackle the issue of overtreatment.

Moreover, I explore how the effect of reputational concerns varies with search costs. The hypothesis is that reputational incentives to mitigate overtreatment are weaker when patients search less frequently due to the burden of high search costs. The findings exhibit some supporting evidence for the differential reputation effect under different search costs from the nonparametric analysis. When the search cost is high, physicians' actions are less likely to be affected by the presence of a reputation mechanism. This result has important policy implications: patients with higher copays and deductibles for their health insurance suffer a higher search cost. Those who cannot afford to search for a second opinion may receive a disproportionate number of improper diagnoses and expensive treatment recommendations. This may further exacerbate the current public concern regarding healthcare inequality in society.

This study provides empirical evidence that reputation can indeed play a crucial role in credence goods markets. However, gathering reputational information by patients themselves is inefficient and costly both in terms of time and money. An intriguing direction for future research would be to explore an alternative reputation mechanism that spreads more rapidly and broadly, all while ensuring the protection of patient privacy. Such a mechanism could potentially contribute to further improvements in physicians' integrity and market efficiency. Furthermore, the current design assumes that physicians are homogeneous in their incentives to provide treatment recommendations. Future studies can appropriately relax this assumption. Many previous studies have entertained the idea that healthcare service providers have different levels of competence and specialize in different treatments (e.g., Wolinsky, 1993). Mixing heterogeneous physicians in the same market may generate an interesting extension of the current experiment.

CHAPTER 2

THE EFFECT OF AGGREGATE TREATMENT INFORMATION ON PHYSICIAN OVERTREATMENT

2.1 Introduction

In healthcare services, physicians are tasked to diagnose patients' illnesses and recommend treatments for them. The physicians, however, usually have more information than their patients in terms of what the proper medical tests and treatments are. This information asymmetry leads some physicians to provide more expensive services than what patients actually need as physicians can earn higher profits by doing so. The markets with these properties are termed "credence goods markets" (Darby and Karni, 1973). The fraudulent behavior where physicians provide excessive and unnecessary services is referred to as "overtreatment".¹

Empirical evidence indicates that financial incentives are a significant driver of physicians' overtreatment (e.g., Domenighetti et al. 1993; Delattre and Dormont 2003; Iizuka 2007; Brownlee 2010; Currie, Lin and Meng 2014; Gottschalk, Mimra and Waibel 2020).² A salient example of overtreatment in childbirth can be found in Johnson and Rehavi (2016). According to their study, OB/GYNs tend to favor cesarean delivery, a procedure with a high reimbursement rate under fee-for-service,

¹This study focuses on physicians' incentives to overtreat their patients. There are two other common types of seller fraudulent behaviors in credence goods markets, undertreatment and overcharging.

²Overtreatment is sometimes driven by physicians' risk aversion toward malpractice liability. This study specifically examines how physicians' financial incentives contribute to overtreatment; therefore, it assumes that physicians can make accurate diagnoses without incurring any costs.

over natural childbirth. In contrast, physician mothers, who are typically more knowledgeable about appropriate levels of medical care, are less likely to get C-sections and tend to experience better health outcomes than nonphysician mothers in similar circumstances. Although overtreating patients increases physicians' payoffs, such misbehavior leaves more costly bills to patients and entails large amounts of wasted resources as more complex tests and treatments typically have higher costs. In the US healthcare market, "medically unnecessary" treatments are the primary cause of waste, amounting to approximately \$210 billion in excess spending each year (McGinnis et al., 2013). The rampancy of overtreatment has been widely seen as a major hindrance to market efficiency and growth. Therefore, it is crucial to find effective mechanisms to deter overtreatment fueled by physicians' financial incentives.

Implementing a traditional reputation mechanism whereby the quality of services provided by physicians is tracked would not be implementable (and would therefore fail to achieve the goal of reducing overtreatment) because patients are unable to tell whether a costly major treatment is necessary for their recovery even after the service is completed. To make a reputation mechanism functional, patients need to have additional channels to monitor physician honesty. There is evidence that overtreatment rates significantly decrease in repeated physician-patient interactions when patients can verify the necessity of expensive treatments through second opinions (Fong, Liu and Meng, 2022; Ge, 2023). The key mechanism is that dishonesty can be inferred when patients receive conflicting recommendations from different physicians (even though this is a credence-good setting). In a repeated interaction environment, this detection mechanism in turn provides an incentive for physicians to curb their overtreatment rate and thereby build a good reputation. However, there exists a serious drawback to the aforementioned mechanism: seeking a second opinion is costly in terms of money and time. According to Wagner and Wagner (1999), 19% of individuals who visited a doctor in 1994 sought a second opinion, with an associated

cost of approximately \$3.2 billion. Moreover, unlike other credence goods where the reputation can easily be disseminated as buyers of bad service willingly share their knowledge by posting bad online reviews (i.e., for car mechanics, tax accountants, etc.), buyers in the healthcare market (i.e., patients) are more likely to have concerns over their medical privacy. In the United States and many other countries, medical records are subject to strict privacy laws and regulations (i.e., the Health Insurance Portability and Accountability Act - HIPAA). As a result, patients are more reluctant to share their detailed experiences with others via online review platforms³, not only limiting patients' ability to use publicly available information to assess physicians' tendency toward overtreatment but also restricting physicians' capacity to establish indirect reputation among future patients.

In recent years, to address a high rate of overtreatment in healthcare markets, policymakers and third-party entities (e.g., Leapfrog Group and Consumer Reports) have developed a new type of publicly available information that keeps track of the volume (*quantities*) of various services provided by each hospital or individual physician. Such information is primarily collected from hospital billing records or annual hospital surveys and aggregated over all patients (see an example of the hospital report on C-section Rates provided by Consumer Reports in 2014⁴). The aggregate data, therefore, protects the privacy of patients' personal health information. In addition to this, policymakers may also mandate hospitals to publish such aggregate information on their websites, empowering patients to make more informed healthcare decisions. As an example, hospitals in New York are now obligated by the state to disclose maternity procedure data from the last five years (Lisa, 2022). Compared to reliance on

³It is worth noting that while online doctor review sites do exist, the majority of patient reviews focus on the qualities that patients can directly experience, such as a doctor's kindness and patience. These types of reviews do not provide informative insights into the quantity of services, making them less effective in influencing doctors' decisions regarding overtreatment.

⁴<https://thechildbirthprofession.com/wp-content/uploads/2014/05/Consumer-Reports-Cesarean-Hospital-Report.pdf>.

second opinions, revealing the physicians' treatment records at the aggregate level, in principle, offers patients a less costly method to monitor physicians' honesty. Patients can consider the frequency with which a physician employs expensive treatments as a potential indicator of the physician's inclination towards overtreatment. For instance, a significantly high cesarean delivery rate for a particular hospital or physician should raise a red flag among future patients.

Motivated by this recent type of policy, I investigate whether disclosing physicians' aggregate records could lead to fewer overtreatments and replace patients' reliance on their own costly search. In order to answer these questions, I design a lab experiment where multiple physicians and patients repeatedly interact in a credence goods market environment. In each period, uninformed patients who suffer from either a minor or major medical issue need to select a physician who can provide a recommendation for either a high-cost or low-cost treatment. A high-cost treatment can treat both major and minor medical issues, whereas a low-cost treatment can only handle minor issues. Before undergoing the procedure, patients have one chance to decline the treatment recommended by their selected physicians and pay a certain fee to seek an alternative physician to interact with. In the experiment, I vary the level of patient information. In the control group, patients can only see their own medical records (i.e., private information). In the two treated groups, not only do patients have knowledge of their own past transactions but they can also observe public information that reveals the treatment records aggregated at each physician level.⁵ The reason for setting up two treatment groups is that I control for whether physicians have (or not) the same public information as patients do. This design allows me to separately identify the mechanisms that may lead to the effectiveness of such a policy: if the efficiency gain is

⁵To make the public information more informative, I reveal the high-cost treatments recommended by each physician rather than those performed in the treated groups. A more detailed explanation will be provided in Section 3.

purely driven by increased patients' ability to detect overtreating physicians, the two treated groups would generate similar results. If the efficiency gain is largely driven by physicians' tendency to match their overtreatment rate with their competitors', the treated group that physicians have access to public information should outperform the treated group that physicians have no access to.

I argue that a controlled laboratory experiment has a clear advantage when investigating the overtreatment problem and its solution in healthcare services. Identifying physician overtreatment behavior is challenging in secondary data analysis because it is often ambiguous to label certain treatments (e.g., cesarean delivery) as either a medically necessary intervention or a case of overtreatment. A laboratory experiment provides a unique opportunity to assign and observe the actual needs of the patients, enabling the clear identification of any instances of physicians' excessive treatments. This serves as the fundamental prerequisite for detecting the causal impact of providing public aggregate information on physicians' strategic motives to overtreat. Moreover, the unique design of this study enables me to manually control physicians' access to public information, allowing for an investigation of the performance of public aggregate information under both conditions.

In my experimental analysis, I find that disclosing physicians' aggregate records provides a similar level of overtreatment compared to the market with private information. This may be because the likelihood of being overtreated is already low when patients can make decisions based on their personal records. Adding physicians' aggregate records does not lead to a further decrease in the level of overtreatment. Moreover, I do not observe any significant reduction in patients' costly searches in the two treated groups as compared to the control group. It seems that an additional channel to monitor physicians' honesty does not reduce patients' reliance on second opinions. When combining these results from those reported in Ge (2023), I conclude that compared to a one-shot transaction, enabling physicians to build direct repu-

tation with patients via repeated interaction is the key factor in improving market efficiency. In this context, revealing physicians' aggregate records does not further enhance market outcomes.

Another noteworthy finding of this study is that the presence of the physician's aggregate records leads physicians to align their overtreatment rates more closely with those provided by their competitors. The effect is more salient when physicians have direct access to such information. This is accompanied by a greater disparity in the average overtreatment levels between markets in the treated groups. This finding suggests that increased information transparency does not necessarily result in reduced overtreatment levels; in some markets, it could yield the opposite outcome. The experiment results in this study show that, on average, the efficiency gains from the markets with lower overtreatment rates happen to offset the efficiency losses from the markets with higher overtreatment rates in the treated groups. This presents another potential reason to explain why I find a lack of significant effect regarding physicians' aggregate records. The finding brings an important policy insight that might have been overlooked: policies intended to improve information transparency may act as a coordination device for physicians and could help create localized norms which may not benefit patients.

The rest of this paper is structured as follows. Section 2 provides a comprehensive review of the relevant literature and discusses the contribution of this study. Section 3 outlines the experimental design. Section 4 provides the hypotheses. Subsequently, Section 5 reports the experimental results. Finally, Section 6 concludes the paper by discussing the policy implications and suggesting future directions for research.

2.2 Literature Review

Following the pioneering work on credence goods markets by Darby and Karni (1973), several studies set out to analyze the impact of various mechanisms that

could mitigate inefficiencies in credence goods markets. These institutions include competition (Dulleck, Kerschbamer and Sutter, 2011; Brosig-Koch, Hehenkamp and Kokot, 2017), second opinions (Mimra, Rasch and Waibel, 2016*a*), price regulations (Mimra, Rasch and Waibel, 2016*a*), payment schemes (Hennig-Schmidt, Selten and Wiesen, 2011; Green, 2014; Brosig-Koch et al., 2016), digital platforms (Balafoutas et al., 2013; Liu, Brynjolfsson and Dowlatabadi, 2021), and monitoring (Angerer, Glätzle-Rützler and Waibel, 2021). Kerschbamer and Sutter (2017) and Balafoutas and Kerschbamer (2020) provide comprehensive reviews of the literature on when and how credence goods can be provided more efficiently. In what follows, I shortly introduce the studies investigating the impact of reputation in credence goods markets and discuss the contributions of this study to the existing literature in this field.

Most prior studies examine how reputational concerns related to service quality influence sellers' behaviors (supply and pricing decisions). Evidence from lab experiments shows that the opportunity for reputation-building through both direct interactions and public information may reduce the likelihood of sellers providing too little (undertreatment) and of overcharging (charging for services not rendered), especially when there are no institutional remedies against fraudulent behaviors of sellers (Dulleck, Kerschbamer and Sutter, 2011; Angerer, Glätzle-Rützler and Waibel, 2021). Luca and Vats (2013), Kolstad (2013), and Chartock (2021) study online rating platforms and find that reputational concerns provide hospitals and physicians with powerful incentives to improve their own performance to remain competitive.

When it comes to the impact of reputation on a seller's decision to overtreat, the results from both lab (Dulleck, Kerschbamer and Sutter, 2011) and field (Schneider, 2012) experiments indicate that the potential for future interactions does not lead to a decrease in overtreatment. This finding can be attributed to the lack of motivation for sellers to provide treatments truthfully when buyers are unable to distinguish between appropriate and excessive treatments based on the services received.

The work of Gerlach and Li (2022), Fong, Liu and Meng (2022), and Ge (2023) highlights the importance of monitoring sellers' honesty to facilitate trust-building and reduce overtreatment. The approach proposed by Gerlach and Li (2022) is to enable buyers to purchase any treatment, regardless of sellers' recommendations. This "freedom of choice" design allows buyers to occasionally validate the necessity of sellers' high-cost treatment recommendations by purchasing low-cost alternatives. The authors have demonstrated the effectiveness of the "freedom of choice" approach through their theoretical analysis and experimental results.

Fong, Liu and Meng (2022) and Ge (2023) investigate whether second opinions can be utilized as a means to discipline sellers in repeated games. In Fong, Liu and Meng (2022), the authors build a theoretical model where long-lived sellers interact with a sequence of short-lived buyers. According to their model, a seller will be forced to leave the market and be replaced by a new entrant if his recommendations for unnecessary treatments are detected by any buyers through second opinions. The authors show that when search costs are low and discount factors are high, buyer search can effectively mitigate overtreatment and improve market efficiency. The work of Ge (2023) is built upon the findings of Mimra, Rasch and Waibel (2016*b*). Mimra, Rasch and Waibel (2016*b*) show that enabling buyers to seek second opinions effectively reduces overtreatment in one-time interactions due to sellers' fear of losing their current business if they recommend unnecessary treatments. Ge (2023) further extends their argument by suggesting that the risk of losing future business can further discipline sellers' recommending behavior in repeated games, and he supports this claim through a laboratory experiment. Since Ge's (2023) focus is on the healthcare market, buyers in his experiment are restricted to make decisions based on their own histories.

As previously mentioned, while patient search is a common practice in the health-care market and can effectively curb physicians' overtreatment, it contributes to the

financial burden on patients. The newly emerged type of public information, physicians' aggregate records, appears to provide a more cost-effective means of disciplining physicians' behaviors without compromising patient privacy. This study contributes to the existing literature by experimentally exploring the effectiveness of physicians' aggregate records in reducing physicians' overtreatment and patients' search. Frankel and Schwarz (2014) provide a theoretical foundation for the effectiveness of aggregate treatment records. In their model, the authors demonstrate that sellers' incentives for overtreatment can be completely corrected if buyers can choose the sellers based on sellers' records of past actions. They argue that as the volume of interactions and discount rate increase, sellers will be indifferent between providing appropriate and excessive treatments because they get more immediate payoffs but less future business from overtreatment, considering that buyers tend to choose sellers who performed less profitable treatments in the past.

This study is also closely related to three articles comparing the impact of private and public information on market outcomes in markets with information asymmetries. Huck, Lünser and Tyran (2012) conduct a trust game to simulate an experience goods market. In their private information scenario, buyers are only informed about their own past interactions when making decisions. In their full information scenario, buyers have access to all past trade histories in the market, including their own. Following Huck, Lünser and Tyran (2012), Mimra, Rasch and Waibel (2016*a*) modify the game to study whether additional public information influences sellers' decisions regarding undertreatment and overcharging in credence goods markets. The experimental results from both Huck, Lünser and Tyran (2012) and Mimra, Rasch and Waibel (2016*a*) show that markets operate at high levels of efficiency when patients have access to their own trade experiences, so providing more information has no additional benefits. A recent study from Angerer et al. (2021) investigates the effect of a public feedback mechanism, in the form of a five-star rating system, on addressing the issue

of undertreatment and overcharging in healthcare markets. They design two distinct market environments: one with direct reputation and another without direct reputation. The market with direct reputation simulates a scenario where both physicians and patients are long-lived and they have prior experience with one another, while the market without direct reputation simulates a scenario that patients are short-lived and they rely only on feedback from previous patients. The results from their experiment suggest that feedback mechanisms are most effective in the market without direct reputation but have no effect in the market with direct reputation.

Their work differs from this study in two key aspects. Firstly, they focus on either experience goods markets or sellers' decisions regarding undertreatment and overcharging in credence goods markets. In contrast, this study investigates whether the availability of additional public information can effectively curb overtreatment. Secondly, unlike their experimental designs, which restrict the observation of public information to buyers in their treated groups, this study incorporates a design with two treated groups that encompasses both scenarios: whether sellers have access to public information or not.

2.3 Experimental Design

This study adopts a general credence goods framing in order to avoid framing effects that may confound the main results. For the purpose of this study, I refer to sellers as physicians and buyers as patients. Following the design of Mimra, Rasch and Waibel (2016*b*) and Ge (2023), I construct a basic market structure with exogenous prices.⁶

⁶Fixed pricing is a prevalent practice in US healthcare markets, where prices are determined through a negotiation process at a centralized level.

2.3.1 Market Procedure

Each market operates with four patients and four physicians per period. The subjects are randomly assigned to one market and to one of these two roles at the beginning of each session. The assignments remain unchanged throughout a total of 20 periods. Within a period, each subject is given an ID number. Physicians' IDs are fixed throughout the experiment in order to allow physicians to build a reputation throughout the experiment. Patients' IDs are re-assigned randomly at the beginning of each period. This design is to prevent physicians from implementing differentiation strategies toward patients.

In a given market period, the decision sequence for patients and physicians is as follows. First, each patient is randomly assigned a type of problem, either a major problem with a probability of $H=0.25$, or a minor problem with a probability of $(1-H)=0.75$; these probabilities are common knowledge. Even though patients recognize that they have a problem, they are unaware of the specific type of problem they have. Physicians are able to perfectly diagnose the problem at no cost.⁷ In the next step, each patient must choose one of the four physicians to receive a treatment recommendation. Simultaneously, each physician needs to provide a treatment recommendation for each patient after observing the type of problem they have. The available treatment options include low-cost and high-cost treatments. While high-cost treatment is capable of treating both major and minor problems, low-cost treatment can only address minor problems. Note that physicians are obligated to cure their patients, so they must provide a high-cost treatment recommendation to patients with major problems.⁸ After a patient selects a physician, he observes the treatment recommen-

⁷In this study, I preclude the possibility of misdiagnosing due to low diagnosis effort or general diagnosing mistakes. Additionally, I assume that physicians do not specialize in either high-cost or low-cost treatments.

⁸In healthcare markets, this assumption holds as leaving patients uncured not only violates the Hippocratic Oath and the World Medical Association's Declaration of Geneva but also leads to

dation offered by the physician. Then, the patient has two options: (1) to accept the recommendation and let the selected physician perform the treatment, or (2) to decline the current treatment recommendation, pay a search cost of K to select a different physician from the remaining three, and receive the treatment from the newly-selected physician. If the patient opts for the latter choice, he must proceed with the treatment under the newly selected physician. For the sake of experiment simplicity, no additional searches are allowed after the second one.

The payoff function π for each patient j is determined by the difference between a fixed value of 130 points, which represents the utility of recovering from an unknown problem, and the price of the recommended treatment, either $P_H = 115$ points or $P_L = 75$ points. It is clear that patients prefer a low-cost treatment recommendation to a high-cost one. Additionally, if a patient chooses to seek a second opinion, they incur an additional search cost of K , $K = 7$ points.

$$\pi_j = \begin{cases} V - P_t & \text{if the patient } j \text{ accepts} \\ V - P_t - K & \text{if the patient } j \text{ searches for a second recommendation} \end{cases} \quad (2.1)$$

Physician i receives payoff e , which is calculated as the difference between the price and cost of the accepted treatment by each patient j :

$$e_i = \sum_{j=1}^n (P_{ijt} - C_{ijt}) \quad (2.2)$$

where t indicates one of the particular types of treatment accepted, which is either high-cost (H) or low-cost (L). Specifically, a high-cost treatment is priced at $P_H = 115$ points and costs a physician $C_H = 80$ points. A Low-cost treatment is priced at $P_L = 75$ points and costs a physician $C_L = 60$ points. These parameters are chosen so that the

medical malpractice lawsuits. When liability is applied, a rational patient who receives a low-cost treatment recommendation can accept it without any doubt.

physicians' profit for a high-cost treatment ($115-80=35$ points) is greater than that for a low-cost treatment ($75-60=15$ points). This is to ensure that physicians have incentives to overtreat patients, which refers to the case in which physicians provide a high-cost treatment recommendation to patients with only a minor problem in this experiment. Note that a physician's payoff is calculated based on the number of patients who accept their recommendations. The physician will not receive any payoffs in this period if no patients accept his treatment recommendations. To keep the experiment simple, I impose verifiability so that physicians must provide the treatment that they recommend.

At the end of each period, both physicians and patients can see their own payoffs for that period. In addition to payoffs, physicians can observe the number of patients they treat, as well as the number of high-cost and low-cost treatments they administer.

In order to mimic real-world markets where overtreatment always leads to an efficiency loss, the above payoff parameters ensure that the physicians' gain from overtreatment is smaller than the net loss to the patients. All of the above information is common knowledge, except for the patient's types (which can be fully observed by physicians but not by patients).

2.3.2 Treatment Design

As mentioned in the introduction, there are one control group with private information and two treatment groups with public information. In the control group with Private Records (PRI), patients are able to browse their own past interactions with the physicians in their first and second searches when making decisions (see *Appendix B Figure B.5*). In this condition, patients can detect overtreatment only through a mismatch of treatment recommendations. This control group mimics the real-world healthcare market where physicians can only build direct reputation with patients.

In the first treated group with Aggregate Records revealed to Patients only (AGG_P), instead of showing patients’ own past transaction histories, I display the percentage of the high-cost treatments recommended by each physician in all previous periods and at the aggregated level (see *Appendix B Figure B.6*). Recall that in the experimental design, a physician needs to make recommendations for each patient, even if the patient does not end up selecting her. Note that the patients can still keep track of their own trade experience with each physician on their own, although such information is not directly shown on the patient’s decision interface.

In the second treated group with Aggregate Records revealed to both Patients and Physicians (AGG_P&PH), physicians can also observe the percentage of the major treatments recommended by each physician, just as patients can (see *Appendix B Figure B.7*). This treated group bears a close resemblance to the real-world healthcare market where transparency legislation is enforced. *Table 2.1* summarizes the three experimental groups.

Table 2.1: Experimental Set-up: Conditions

	Cond.	Description
Control Group	PRI	Private Records
Treated Groups	AGG_P	Aggregate Records visible to patients only
	AGG_P&PH	Aggregate Records visible to patients and physicians

2.3.3 Experimental Protocol

I implemented a between-subjects design such that each subject participated in only one of the three experimental conditions. There are four markets in each condition with a total of 96 subjects. Subjects were recruited from the subject pool at the University of Massachusetts Amherst using ORSEE (Greiner, 2015). The experiment was conducted on an AWS (Amazon Web Services) server using the online version of z-Tree (Fischbacher, 2007; Duch, Grossmann and Lauer, 2020).

At the beginning of each session, subjects were invited to join the Zoom meeting room while keeping their audio and video off to protect their privacy. They were asked to remain in the meeting room until the session ended. The instructions were read aloud by the experimenter, and subsequently, subjects were required to respond to a set of control questions to ensure their comprehension of the instructions (see Appendix B.1 and B.2 for the instructions and control questions, respectively). During the experiment, subjects were individually assigned to breakout rooms, allowing the experimenter to offer private assistance as needed. A brief questionnaire was administered after the experiment to gather demographic information (see Appendix B.3).

The payment included the show-up fee of \$5 and the cumulative payoff from the decisions made in all 20 periods. The subjects received their payoff with an Amazon eGift card⁹ based on the exchange rate of 50 points = \$1. The average total payoff per subject is \$16.94. The average session length, including instructions and a pre-experiment quiz, was approximately 65 min.

2.4 Hypotheses

In the analysis, three aspects are of prime interest: (i) the overtreatment level, (ii) the search rate, and (iii) the market efficiency. This section elaborates on how to measure these aspects, formulates corresponding hypotheses, and provides logical explanations for each hypothesis.

I have no priori directions for comparing market outcomes between the two treated groups. The main reason is that providing physicians with access to their aggregate records could potentially lead to either heightened competitiveness among them or the facilitation of coordination. These two logical arguments result in completely

⁹The gift card was sent to each subject within three hours after the session ended.

opposing outcomes. Therefore, I attribute equal significance to both treated groups for the following hypotheses.

2.4.1 Overtreatment Level

Overtreatment is defined as a situation where patients with a minor problem are provided with high-cost treatment recommendations. Following Mimra, Rasch and Waibel (2016*b*) and Ge (2023), I apply two distinct methods to measure overtreatment. First, I analyze the “overtreatment in strategy” based on all the recommendations submitted by a physician in her strategy profile. The strategy method allows me to evaluate, for each physician, how many patients with a minor problem in the market the physician would have overtreated. Next, I measure the level of “actual overtreatment” patients experience. Actual overtreatment is a subset of overtreatment in strategy: it only considers physicians’ overtreatment that has been selected by patients and actually performed. In summary, overtreatment in strategy indicates a physician’s intended honesty level, whereas actual overtreatment reflects patients’ choices and is thus directly related to observed levels of market efficiency.

Based on the above two definitions for measuring the overtreatment level, I formulate the following hypotheses:

Hypothesis 1: For the overtreatment level in strategy, $PRI > AGG_P = AGG_P\&PH$

Hypothesis 2: For the actual overtreatment level, $PRI > AGG_P = AGG_P\&PH$

In the experiment, physicians establish a good reputation by recommending cost-effective treatments as much as possible, given that patients tend to select physicians who propose fewer high-cost treatments in the preceding periods and shun those who suggest more high-cost treatments. Hence, when the information on physi-

cians' past recommending actions becomes more transparent in the treated groups of AGG_P and AGG_P&PH, physicians are expected to be more motivated to minimize overtreatment among their patients.

Moreover, I anticipate observing a more pronounced decline in the actual overtreatment level for the following reasons: firstly, the increased honesty in treatment recommendations from physicians in Agg_P and Agg_P&PH is expected to decrease the likelihood of patients experiencing overtreatment; secondly, disclosing the physicians' aggregate records empowers patients to make better-informed choices over physicians.

2.4.2 Search Rate

I also utilize two measures to gauge patients' search behavior: conditional and unconditional search rates. The conditional search is applicable only when a patient is recommended a high-cost treatment in her first search. The conditional rate provides a straightforward way to measure patients' trust in their physicians.¹⁰ The unconditional search considers all patients' decisions about seeking a second opinion, regardless of the type of treatment received on their first visits. To sum up, the conditional search rate provides a more precise measure of patient trust, whereas the unconditional search rate is more closely associated with the calculation of market efficiency.

Based on the above two definitions of patient search, I formulate the following hypotheses:

Hypothesis 3: For the conditional search rate, $PRI > AGG_P = AGG_P\&PH$

¹⁰Note that physicians are obligated to cure patients. Patients can accept a low-cost treatment recommendation without hesitation.

Hypothesis 4: For the unconditional search rate, $PRI > AGG_P = AGG_P\&PH$

In the treated groups of AGG_P and AGG_P&PH, patients are provided with an additional means of monitoring physicians by comparing physicians' past recommendation patterns. Given that each physician is tasked with providing treatment recommendations to the same patient group, it is easy for patients to identify dishonest physicians through the aggregate records. Thus, I argue that patients' trust in the recommendation received increases in the treated groups, leading to a decrease in the conditional search rate.

Given that the level of overtreatment in strategy decreases in the treated groups, it is expected that patients will have a higher probability of receiving low-cost treatment recommendations during their first visit. Consequently, there will be a decreased need for patients to seek a second opinion in the treated groups. Thus, I expect to observe a more salient reduction in the unconditional search rate.

2.4.3 Market Efficiency

As I mentioned above, market efficiency depends on actual overtreatment level and unconditional search rate. To show this, I define a relative efficiency measure by normalizing the sum of patients' and physicians' surpluses per market and period. This normalization ensures that the efficiency measure falls within the range of 0 to 1. Based on the above definition of relative market efficiency, I formulate the following hypothesis:

Hypothesis 5: For the relative market efficiency, $AGG_P = AGG_P\&PH > PRI$

As the actual overtreatment rates and the unconditional search rates are expected to decrease in the treated groups, I anticipate an improvement in the market efficiency with the presence of physicians' aggregated treatment records.

2.5 Results

In this section, I test the above hypotheses and discuss the impact of physicians' aggregate records. I first report treatment comparisons for each outcome variable using two-tailed Mann-Whitney U tests. Considering decision dependency, I take the average of a particular measure across all individuals and all market periods as one independent observation.¹¹ For more detailed analysis and robustness checks, I complement the non-parametric test results with random effects regressions clustered at the market level. Specifically, I use Probit regressions for binary dependent variables and OLS regressions for continuous dependent variables. For each outcome variable, I present the results from three distinct specifications: Model (1) concentrates solely on the condition variables; Model (2) includes controls for basic subject demographics, including gender, age, GPA, and prior enrollment in economics courses; Model (3) goes a step further by incorporating the time period variable along with its interactions with the condition variables to investigate the potential presence of a time trend. Additionally, to compare the difference between the two treated groups, a Wald test is employed, utilizing the coefficients of these two treated groups obtained from Model (2).

¹¹When I conduct the Mann-Whitney U tests, I also try two alternative assumptions on decision independency: (1) correlated periods with independent individuals, and (2) independent periods with correlated individuals. Both analyses yield similar conclusions compared to those obtained from the main assumption used in this study.

2.5.1 Overtreatment

Following the aforementioned Hypothesis 1 and 2, I start my analysis by examining whether disclosing the aggregate records can effectively decrease the level of overtreatment.

2.5.1.1 Overtreatment in Strategy

Firstly, I analyze the overtreatment in strategy. For each physician, I calculate the average level of her decisions to overtreat patients with a minor problem in each period.

Figure 2.1: Level of Overtreatment in Strategy

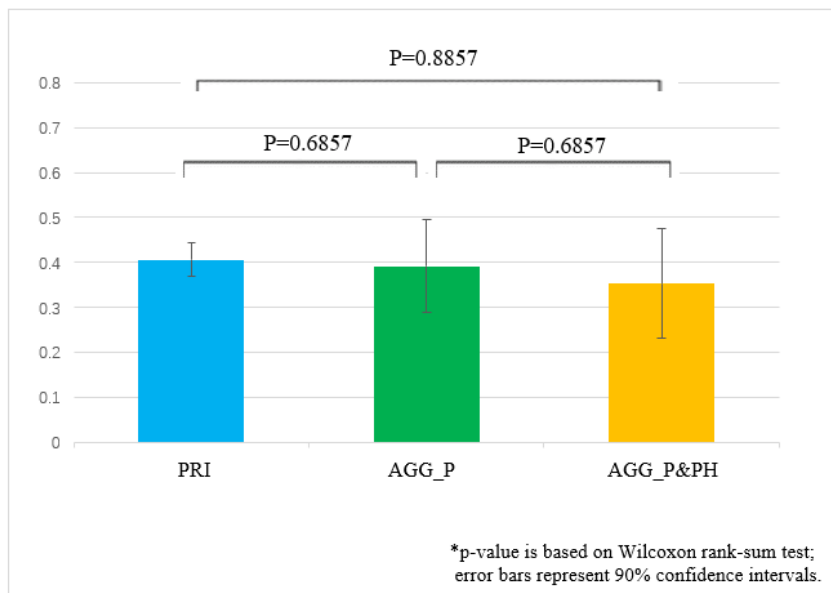


Figure 2.1 shows the average level of overtreatment in strategy across all three experimental groups. Physicians choose to overtreat patients 40.60% of the time when patients can only obtain their private records. When the aggregate records are disclosed, physicians on average reduce their tendency to recommend high-cost treatments but only by a small degree: the overtreatment level is 39.19% in AGG_P and 35.29% in AGG_P&PH. According to the non-parametric test results, the re-

Table 2.2: Random effects panel OLS: level of overtreatment in strategy

	Model 1	Model 2	Model 3
Condition (vs. PRI)			
AGG_P	-0.01406 (0.043)	-0.01032 (0.038)	-0.09708** (0.049)
AGG_P&PH	-0.05312 (0.049)	-0.06679 (0.048)	-0.01457 (0.072)
Period			0.00018 (0.003)
Period x AGG_P			-0.00826** (0.004)
Period x AGG_P&PH			-0.00497 (0.005)
Control		✓	✓
Constant	✓	✓	✓
Observations	960	960	960

Notes. Standard errors (reported in parentheses) are robust and clustered at the market level.

***:p<0.01, **:p<0.05, *:p<0.1

duction does not reach statistical significance (PRI vs. AGG_P, p=0.6857; PRI vs. AGG_P&PH, p=0.8857;). Furthermore, no significant differences are found between the two treated groups (AGG_P vs. AGG_P&PH, p = 0.6857).

Next, let's look at the results from the regression models. *Table 2.2* reports the estimated marginal effects of how the aggregate records impact physicians' decisions to overtreat. The results from Model (1) and Model (2) in *Table 2.2* are consistent with the findings of the non-parametric tests, indicating that as more information is provided to patients, the average changes in physicians' responses are minimal and the effects are insignificant, which runs contrary to Hypothesis 1.

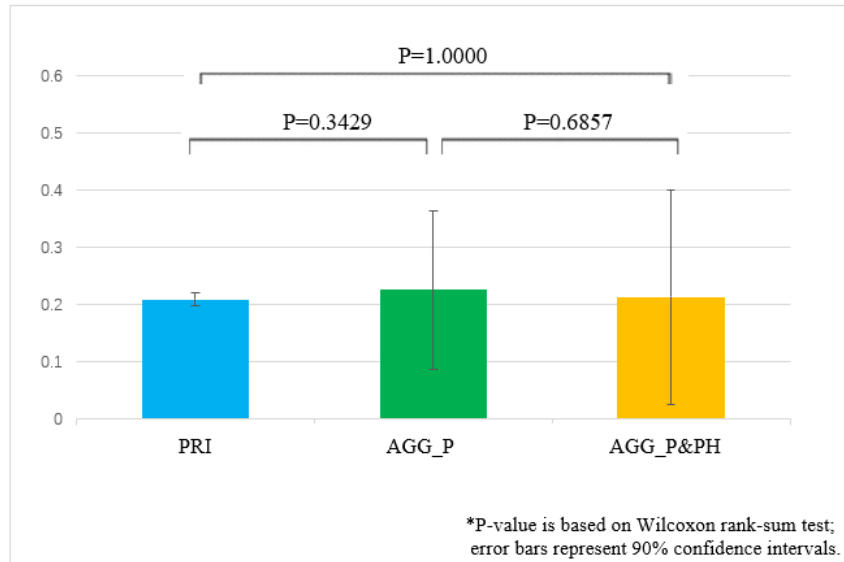
After controlling for the experiment period and its interaction with experimental conditions, I observe a dynamic pattern in AGG_P in Model (3). The level of overtreatment in strategy in AGG_P is 9.7% higher than that in PRI in the first period but it gradually decreases over time. The findings suggest that physicians in

AGG_P may initially underestimate the effect of aggregate information, resulting in a statistically significant and economically large increase in overtreatment at the start of the experiment (nearly 10% higher than in later periods). This overtreatment, however, gradually disappears as indicated by the continuous reduction of overtreatment over each additional period. On average, each additional period in AGG_P condition leads to 0.826% less overtreatment in strategy. This period effect in the AGG_P condition suggests that when the aggregate records are observable by patients only, the market will first experience a significantly higher undesirable overtreatment rate. However, physicians' overtreatment behavior evolves over time and eventually dissipates. Since the overall average overtreatment is similar in PRI and AGG_P, I argue that the PRI condition may be more desirable than the AGG_P condition as it produces more stable results from the very start.

Consistent with the findings from the Rank-sum test, the Wald test confirms that there is no statistically significant difference in the overtreatment level in strategy between the two treated groups ($p=0.139$). It seems that having access to the competitors' records does not change physicians' decisions regarding overtreatment.

2.5.2 Actual Overtreatment

Figure 2.2: Level of Actual Overtreatment



Next, I investigate whether disclosing aggregate records reduces the likelihood of patients being overtreated by analyzing the level of actual overtreatment. For each individual patient, the actual overtreatment of 1 (0) corresponds to a situation in which she is (not) overtreated during this period when she has a minor problem, and it is a null value if a patient has a major problem during this period.

As shown in *Figure 2.2*, patients are more likely to be overtreated in AGG_P and AGG_P&PH than in PRI. The average level of actual overtreatment rises from 20.83% in PRI to 22.50% in AGG_P and to 21.25% in AGG_P&PH. However, I do not observe any significant difference in the actual overtreatment level among the three experimental groups (PRI vs. AGG_P, $p=0.3429$; PRI vs. AGG_P&PH, $p=1.000$; AGG_P vs. AGG_P&PH, $p = 0.6857$).

Table 2.3: Random effects panel Probit: level of actual overtreatment

	Model 1 (dy/dx)	Model 2 (dy/dx)	Model 3 (dy/dx)
Condition (vs. PRI)			
AGG_P	0.01368 (0.052)	-0.02566 (0.068)	-0.02400 (0.068)
AGG_P&PH	-0.00325 (0.073)	-0.00259 (0.067)	-0.00251 (0.066)
Period			-0.00649 (0.005)
Period x AGG_P			-0.00649 (0.005)
Period x AGG_P&PH			0.00025 (0.007)
Control		✓	✓
Constant	✓	✓	✓
Observations	720	720	720

Notes. Standard errors (reported in parentheses) are robust and calculated using the delta method.

The coefficients show the marginal effects of each variable, with all other variables held at their means.

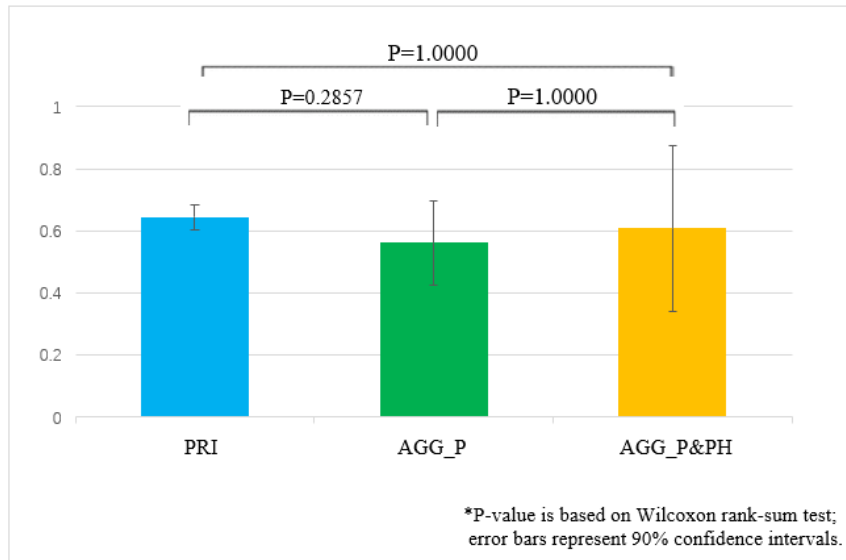
***:p<0.01, **:p<0.05, *:p<0.1

The panel Probit regressions in *Table 2.3* confirm the nonparametric results that there is no statistically significant effect of the aggregate records on the level of actual overtreatment, a finding that is contrary to Hypothesis 2. The results indicate that disclosing physicians' aggregate records neither reduces the likelihood of patients receiving overtreatment nor helps patients make more informed decisions when selecting physicians. Moreover, the Wald test does not detect any significant differences in the coefficients between the two treated groups (p=0.852).

2.5.3 Patient Search

Subsequently, I examine whether revealing aggregate records reduces conditional and unconditional search rates.

Figure 2.3: Conditional Search Rate



2.5.3.1 Conditional Search Rate

For each individual patient, the conditional search of 1 (0) corresponds to a situation in which she does (does not) search for a second opinion when she received a high-cost recommendation on her first visit in the current period. It is a null value if a patient received a low-cost recommendation on her first visit in the current period. Hence, the conditional search rate in each market and each period is computed as the frequency of patients conducting a second search when faced with a high-cost recommendation. I expect to use this measurement to investigate whether additional public information can increase patient trust in the high-cost treatment recommendations provided by their physicians.

Figure 2.3 shows the average search rate conditional on the high-cost recommendations across the three experimental groups. Upon receiving a high-cost treatment recommendation on their first physician visit, patients search less often for a second opinion when patients can obtain the aggregate records. The conditional search rate drops from 64.56% in PRI to 54.66% in AGG_P and to 58.07% in AGG_P&PH.

Table 2.4: Random effects panel Probit: conditional patient search

	Model 1 (dy/dx)	Model 2 (dy/dx)	Model 3 (dy/dx)
Condition (vs. PRI)			
AGG_P	-0.10492* (0.061)	-0.06045 (0.103)	-0.06338 (0.103)
AGG_P&PH	-0.04797 (0.101)	-0.03527 (0.094)	-0.03543 (0.096)
Period			-0.00330 (0.006)
Period x AGG_P			-0.00293 (0.005)
Period x AGG_P&PH			0.01269 (0.008)
Control		✓	✓
Constant	✓	✓	✓
Observations	474	474	474

Notes. Standard errors (reported in parentheses) are robust and calculated using the delta method.

The coefficients show the marginal effects of each variable, with all other variables held at their means.

***:p<0.01, **:p<0.05, *:p<0.1

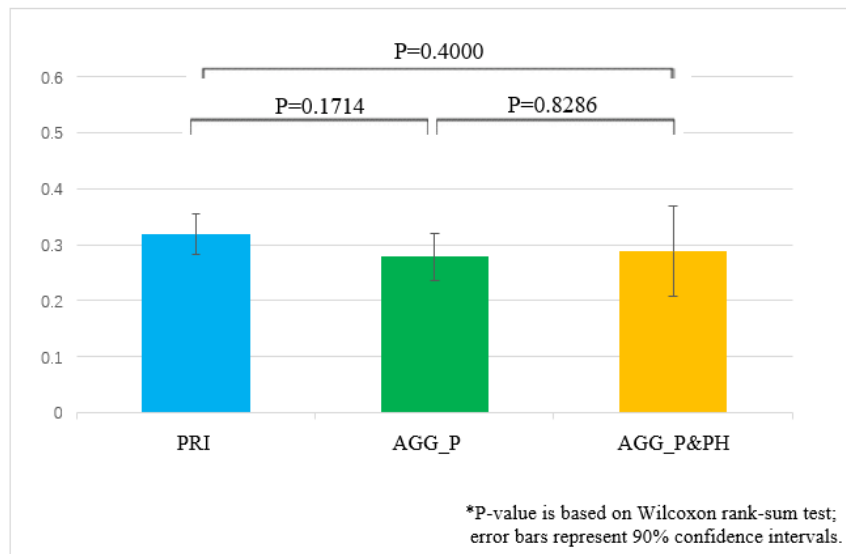
However, the reductions are not statistically significant (PRI vs. AGG_P, p=0.2857; PRI vs. AGG_P&PH, p=1.0000). Moreover, there is no significant difference in conditional search rates between the treated groups (AGG_P vs. AGG_P&PH, p=1.0000). These results suggest that the availability of the physicians' aggregate records fails to alter patients' trust in physicians and reduce their reliance on second opinions.

The corresponding panel Probit regressions are reported in *Table 2.4*. Inconsistent with the results from the nonparametric analysis, the results in Model (1) show that patients are 10% less willing to search for a second opinion when only patients have access to the physicians' aggregate records. However, the effect is only significant at the 10% level and this significance disappears as more control variables are added

in Model (2) and Model (3). No effect is detected for the AGG_P&PH condition in either of the three models. Overall, these results show, at best, weak support that patients' access to patients' aggregate records can increase patients' trust in treatment recommendations. Additionally, utilizing the coefficients obtained from Model (2), the Wald test shows no significant difference between the two treated groups ($p=0.852$).

2.5.3.2 Unconditional Search Rate

Figure 2.4: Unconditional Search Rate



The only difference between the measurements of conditional and unconditional search rates lies in the denominator. The denominator of the unconditional search rate considers the total number of transactions, irrespective of the type of recommendations received during patients' first visit. Compared to the unconditional search rate, this measurement also takes into account whether patients receive more truthful treatment recommendations during their first visit.

Figure 2.4 shows that the unconditional search rate is also lower in AGG_P and AGG_P&PH than PRI on average. It drops from 31.88% in PRI to 27.81% in

Table 2.5: Random effects panel Probit: unconditional patient search

	Model 1	Model 2	Model 3
Condition (vs. PRI)			
AGG_P	-0.04322*	-0.03374	-0.03459
	(0.023)	(0.048)	(0.049)
AGG_P&PH	-0.03023	-0.02763	-0.02772
	(0.034)	(0.039)	(0.040)
Period			-0.00611
			(0.006)
Period x AGG_P			-0.00906***
			(0.003)
Period x AGG_P&PH			0.00458
			(0.004)
Control		✓	✓
Constant	✓	✓	✓
Observations	960	960	960

Notes. Standard errors (reported in parentheses) are robust and calculated using the delta method.

The coefficients show the marginal effects of each variable, with all other variables held at their means.

***:p<0.01, **:p<0.05, *:p<0.1

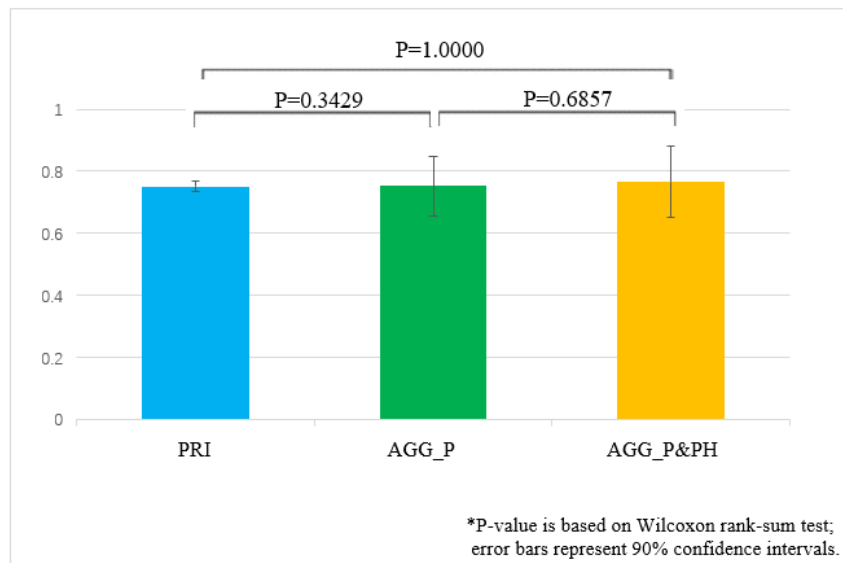
AGG_P and to 28.75% in AGG_P&PH. However, no statistically significant difference is detected among the three groups (PRI vs. AGG_P, p=0.1714; PRI vs. AGG_P&PH, p=0.8286; AGG_P vs. AGG_P&PH, p = 0.4000). Again, the results provide no evidence that disclosing the physicians' aggregate records changes the frequency of patients seeking second opinions.

The results of the regression models are presented in *Table 2.5*. In Model (1), I observe a weakly significant decrease of 4% in the frequency of patients seeking a second opinion in the AGG_P condition compared to the PRI condition. However, the significance disappears after I control for more variables in Model (2) and Model (3), similar to what I find in the analysis of the conditional search. In Model (3), by incorporating period and its interactions with experimental conditions as control variables, I find that patients' tendency to conduct a second search in AGG_P signif-

icantly decreases over time. The findings echo those reported in *Table 2.2* Model (3). As physicians tend to provide more truthful treatment recommendations over time in AGG_P, patients are less likely to receive a high-cost treatment recommendation in their initial search. According to the Wald test, the frequency of patient search is not affected by physicians' access to the aggregate treatment information ($p= 0.889$).

2.5.4 Market Efficiency

Figure 2.5: Relative Market Efficiency



To measure the relative market efficiency achieved in every market and period, I divide the difference between the actual total surplus and the minimal surplus of the market by the difference between the maximal possible surplus and the minimal surplus of the market in a given period. A relative efficiency of 0 represents the minimal possible surplus of the market and corresponds to a situation in which physicians always overtreat their patients with a minor problem and patients always search for a second recommendation regardless of the initial recommendation they receive; a relative efficiency of 1 represents the maximal possible surplus, which is achieved when physicians never overtreat and patients never search for a second opinion.

Table 2.6: Random effects panel OLS: relative market efficiency

	Model 1	Model 2	Model 3
Condition (vs. PRI)			
AGG_P	0.00152 (0.038)	0.01923 (0.045)	0.00613 (0.079)
AGG_P&PH	0.01517 (0.045)	0.05906 (0.040)	0.12995 (0.093)
Period			0.00572 (0.005)
Period x AGG_P			0.00125 (0.006)
Period x AGG_P&PH			-0.00675 (0.007)
Control		✓	✓
Constant	✓	✓	✓
Observations	240	240	240

Notes. Standard errors (reported in parentheses) are robust and clustered at the market level.

***:p<0.01, **:p<0.05, *:p<0.1

Upon providing patients with aggregate records, the relative market efficiency slightly increases from 75.08% to 75.24% in AGG_P and to 76.60% in AGG_P&PH on average. Nevertheless, the efficiency gain is not statistically significant according to the non-parametric test results presented in *Figure 2.5* (PRI vs. AGG_P, p=0.3429; PRI vs. AGG_P&PH, p=1.000; AGG_P vs. AGG_P&PH, p = 0.6857).

The results from the regression models in *Table 2.6* confirm the nonparametric findings and are robust to the inclusion of period, interaction terms and demographic controls. Despite the slight improvement in market efficiency with the disclosure of physicians' aggregate records, the improvements are not statistically significant. Although the findings contradict Hypothesis 5, they are not surprising considering the lack of significant changes in the actual overtreatment rate and the slight decrease in the unconditional search rate. The Wald test also shows that market efficiency does

not change significantly regardless of whether physicians have access to the aggregate treatment information or not ($p=0.476$).

2.5.5 Variance in Physicians' Decisions within Markets

Table 2.7: Random effects panel OLS: market-level SD of physicians' overtreatment in strategy

	Model 1 (dy/dx)	Model 2 (dy/dx)	Model 3 (dy/dx)
Condition (vs. PRI)			
AGG_P	-0.04937* (0.025)	-0.04374* (0.024)	-0.00518 (0.035)
AGG_P&PH	-0.08771*** (0.028)	-0.06677*** (0.024)	-0.06254** (0.027)
Period		0.00014 (0.001)	0.00234 (0.002)
Period x AGG_P			-0.00421 (0.003)
Period x AGG_P&PH			-0.00240 (0.003)
Control		✓	
Constant	✓	✓	✓
Observations	240	240	240

Notes. Standard errors (reported in parentheses) are robust and clustered at the market level.

***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$

While comparing each outcome variable across the three experimental conditions, I observe a greater disparity between the markets in the AGG_P and AGG_P&PH conditions (as shown in Figure 1-5). In this section, I further investigate whether high variances also exist in physicians' decisions within markets in the two treated groups. To this end, I compute the standard deviation of physicians' overtreatment levels in strategy in each market and period as a measure of the variability in physicians' decisions within markets. I then utilize this measure as the dependent variable and perform Random-effect Panel regressions. The specifications I use differ slightly from

those reported above. In Model (1), I focus exclusively on the condition variables. In Model (2), I control for the time period variable and basic subject demographics aggregated at the market level, while in Model (3), I control for the time period and its interactions with condition variables. *Table 2.7* reports the results from these regression models.

Surprisingly, both Model (1) and Model (2) show that physicians' decisions are significantly less spread out in a given market in the treated groups, and the effect is more prominent in the AGG_P&PH condition. In Model (3), the impact of AGG_P&PH condition remains statistically significant at the 5% level. The results suggest that as transparency of reputational information increases, physicians adjust their overtreatment rate to match that of their peers. Note that even though physicians in the AGG_P condition do not have direct access to information on their competitors' past recommendations, they can easily make inferences based on their past earnings.

By combining this finding with the previously observed large volatility in the average overtreatment rates between markets in the treated group, it becomes evident that physicians' alignment of strategies with their competitive peers, facilitated by the increased information transparency, does not necessarily lead to a reduction in overtreatment levels. This entails a risk where, in certain markets, physicians might employ this information to coordinate with each other, potentially leading to an escalation in the market's overtreatment level. Such an outcome clearly contradicts the intended policy goal. According to my experimental data, in the treated groups, markets with lower overtreatment levels and markets with higher overtreatment levels happened to offset each other.

Furthermore, these findings seem to provide a plausible explanation for why patients find it challenging to adopt the physicians' aggregate records as an alternative monitoring method, as initially expected. This is because physicians within a given

market tend to recommend high-cost treatments at a comparable frequency. As a result, patients are left with no choice but to rely on second opinions to validate the necessity of these high-cost treatment recommendations.

2.6 Conclusion

The healthcare market is characterized by an information asymmetry where physicians possess more knowledge about the appropriate treatments than patients. While combining the reputation mechanism with patient search has been demonstrated to effectively curb overtreatment, monitoring physicians' honesty through second opinions is a costly and time-consuming process. In this study, motivated by the emergence of public information tracking the quantities of services offered by healthcare providers, I experimentally investigate its potential in addressing the efficient loss arising from physician overtreatment and patient search.

The empirical investigation conducted in this study reveals that the disclosure of physicians' aggregate records has minimal impact on physicians' inclination to overtreat, patients' decisions to search for second opinions, and overall market efficiency when patients can rely on their own experiences. The findings of this study align with prior research (Huck, Lünser and Tyran, 2012; Mimra, Rasch and Waibel, 2016*a*; Angerer, Glätzle-Rützler and Waibel, 2021), which demonstrates that public information in both experienced goods markets and credence goods markets do not carry many additional benefits in terms of market outcomes when market participants draw on personal relationships. One plausible explanation proposed in these studies is that when buyers can make decisions based on their own experiences, sellers' misbehaviors are already substantially mitigated. As a result, additional information becomes less likely to yield further improvements. Therefore, the findings suggest that the policy of disclosing physicians' aggregate records may not achieve the anticipated improvement of market outcomes in the scenarios that physicians, such as General

Practitioners, have more repeated interactions with patients and can establish direct reputation.

This study also presents another notable finding. The data from the experimental sample show that there is reduced variation in the level of overtreatment in strategy within markets but increased volatility between markets when physicians' aggregate records are disclosed. It indicates that the increased information transparency through public information allows physicians to align their recommending strategies with those of their competitors rather than engaging in competition by recommending fewer unnecessary treatments. Certainly, this complicates the assessment of physicians' aggregate records on market outcomes, which serves as another potential explanation for the fewer significant results observed in my study. The finding provides valuable insights for policymakers contemplating the disclosure of physicians' aggregate records. It seems that complementary policies are necessary to prevent overtreatment from escalating to higher levels when the localized market norms are unfavorable to patients.

To the best of my knowledge, this is the first study using a lab experiment to investigate the effect of physicians' aggregate records on market outcomes. It provides many avenues for future research. In this study, I solely focus on the scenario of repeated interactions between physicians and patients. However, it is worth noting that healthcare markets encompass both first-time and repeated interactions between these parties. Therefore, it is crucial for future research to explore the effectiveness of the physicians' aggregate records in scenarios where physicians cannot easily establish direct reputation, simulating one-time interactions. I consider it to be a fruitful direction for future research, as previous work has shown that public information, such as patients' rating systems, is beneficial in terms of reducing undertreatment and overcharging in the one-time physician-patient interactions in healthcare markets (Angerer, Glätzle-Rützler and Waibel, 2021).

Another avenue of research can focus on the variation in search costs among patients. In my experimental set-up, search costs are uniformly set at 7 points for all patients, which can be seen as a moderate amount compared to the potential benefits of obtaining a low-cost treatment from a different physician. However, in reality, the costs associated with seeking second opinions vary from patient to patient. It is crucial to study the distributional impact of disclosing physicians' aggregate records for different patient groups in terms of their varying search costs.

Lastly, the findings of this study indicate that physicians adjust their own recommending strategies when they have more knowledge about their competitors' past actions. In my experiment, all physicians are compensated through the fee-for-service system, which provides them with financial incentives to overtreat their patients. In real-world healthcare markets, there are many hospitals implementing alternative payment schemes. For example, under the HMO (Health Maintenance Organization) payment scheme, physicians receive a fixed, predetermined payment per patient, regardless of the actual services provided or the costs incurred. Generally, the physicians under this payment scheme have fewer incentives for overtreatment. Hence, it is intriguing to explore how the availability of physicians' aggregate records to the public would influence market outcomes, given the presence of different payment schemes in the market.

This study can be regarded as a starting point for evaluating the effectiveness of physicians' aggregate records, particularly investigating its performance in the context where physicians can build direct reputation to their patients through repeated interactions. However, real-world healthcare markets have various conditions that differ from the setups in this study. A more comprehensive understanding of this newly-emerging type of public information is necessary before we can accurately assess its effectiveness.

CHAPTER 3

MARKET POWER IN THE NEW ENGLAND ELECTRICITY MARKET: EVIDENCE FROM NUCLEAR REFUELING OUTAGES¹

3.1 Introduction

Electricity market restructuring, often referred to as deregulation, has been a significant global trend over the past few decades. In the United States, this process began in the 1990s as an effort to dismantle traditional utility monopolies and introduce competition into wholesale electricity markets. California took the lead in implementing these initiatives, followed by the New England states, and subsequently expanded to the mid-Atlantic region. Nowadays, approximately two-thirds of the US electricity demand is serviced by the deregulated electricity markets.² Under the restructured model, electricity generation is separated from transmission and distribution. Electricity generation has become competitive, with multiple energy suppliers selling electricity in wholesale markets, while transmission and distribution have remained regulated, as these aspects of the electricity system are natural monopolies. The primary objective of restructuring is to let competition in wholesale markets drive down electricity prices and offer consumers a choice of suppliers.

¹This chapter is coauthored with Xiaolin Zhou. We are indebted to Matt Woerman for his invaluable guidance and advice. We also thank Juan-Camilo Cárdenas, Yongjoon Park, Christian Rojas, Rong Rong, and seminar participants at the AERE 2023 Summer Conference and UMass ResEcon IO Reading Group for their insightful comments and suggestions.

²United States Environmental Protection Agency. (2023, February 5). Power Market Structure. <https://www.epa.gov/green-power-markets/power-market-structure>

Although economic theory suggests that increased competition in a market leads to lower prices and reduced market power, the reality in electricity markets has proven to be more complicated. Despite the introduction of competition, market power has remained prevalent in deregulated electricity markets, resulting in consumers continuing to bear high electricity prices and unreliable power supply. This is primarily due to the unique features of electricity, such as capacity constraints, expensive storage, homogeneity of the product, and perfectly inelastic demand. These features enable energy suppliers with small market shares to exercise significant market power (Borenstein and Bushnell, 2000; Borenstein, Bushnell and Wolak, 2002). In some cases, suppliers exercising market power has even led to catastrophic consequences, as exemplified by the 2000–01 California electricity crisis, which exposes the vulnerabilities of deregulated electricity markets and highlights the need for effective oversight and safeguards to prevent abuse of market power.

In light of these lessons, regulators have persistently improved the functioning of deregulated electricity markets by bolstering their monitoring of market participants³ and implementing a range of complementary market mechanisms, such as the forward reserve market, ancillary services, and capacity market. Considering the dynamic nature of market conditions, it is essential to periodically conduct investigations to assess whether the exercise of market power by suppliers has been effectively curbed. Moreover, many electricity markets in the United States, as well as around the world, are currently transitioning to a low-carbon market, with renewable energy sources gradually replacing traditional fossil fuels. It is also important to consider the potential challenges that market power may pose to this transition.

³While the entire U.S. electricity market is overseen by the Federal Energy Regulatory Commission (FERC), each regional market operator, either Independent System Operator (ISO) or Regional Transmission Organization (RTO), has been granted substantial autonomy in implementing a variety of market power mitigation regulations. See Graf et al. (2021) for more detailed information.

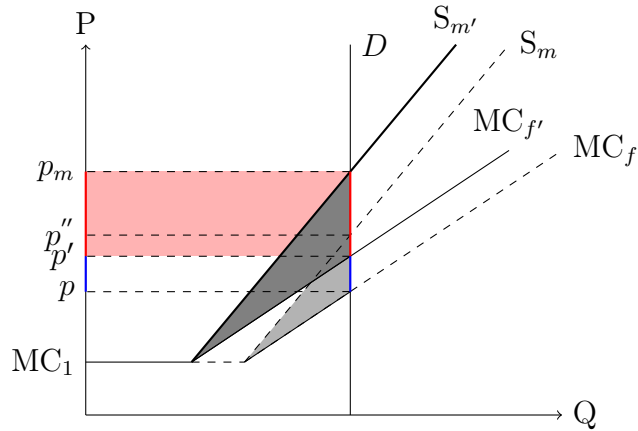


Figure 3.1: Welfare impact of market power

In this paper, we use nuclear refueling outages as a test for market power in deregulated wholesale electricity markets. Nuclear power has the lowest marginal cost among all non-renewable energy and, therefore, it always sits at the bottom of the supply curve when the reactors operate. However, each nuclear reactor must undergo periodic refueling approximately every 18 months and refueling typically leads to month-long outages.⁴ The temporary baseload supply shortage in the market provides other energy suppliers with more incentives to exercise market power to increase their markup. *Figure 3.1* illustrates how and why suppliers want to exercise market power during nuclear outages. During nuclear refueling outages, the marginal cost curve shifts leftward from MC_f to $MC_{f'}$ as higher-cost generators are dispatched to compensate for the gap left by reduced nuclear generation. This shift pushes the competitive price from p to p' . At this moment, the remaining suppliers can capture higher markup if they increase the price they ask for by the same magnitude, as evidenced by the comparison between the dark-shaded area and the lightly-shaded area in the figure. The red-shaded area represents the welfare transfer from consumers

⁴Nuclear reactor refueling is a time-consuming process that involves shutdown, cooling, removal of spent fuel rods, and installation of new fuel rods.

to suppliers, which is simply the area of price increases by demand. The welfare impact of manipulating bids is much larger when the marginal cost curve is closer to the vertical axis.

We select the New England electricity market as the subject of study because nuclear power is one of the main energy resources in this market, which allows us to observe the bidding behaviors of energy suppliers at different levels of nuclear utilization. Using day-ahead hourly market data from 2016 to 2018 in New England, we observe that, on average, the wholesale market-cleared price increased by \$4.9/MWh when at least one of the four nuclear reactors in the region is offline after accounting for other influencing factors. Next, through a series of analyses on the bidding behavior of energy suppliers, we find that approximately 36% of the price effect is explained by increased market power. The results raise the concern that the exercise of market power during nuclear refueling outages can result in a sizable welfare transfer from consumers to suppliers. We estimate this transfer to be around \$43 to \$45 million on average per year during our study period.

The findings of this paper should alert regulators and market designers to the exercise of market power by energy suppliers during negative baseload supply shocks. More importantly, they should recognize that this issue may escalate in severity during the transition to a low-carbon market. This is because most renewable resources are intermittent in nature, dependent on factors such as sunlight for solar power and wind for wind power. During periods when the sun doesn't shine and the wind doesn't blow, the remaining fossil-fuel power plants will have dramatically increased market power, which will empower them with greater control over determining market-clearing prices.

This paper makes three main contributions to the economics literature. First, we propose a novel approach for detecting the exercise of market power. The most common method for achieving this is by utilizing the Lerner Index, which compares

the market price to the competitive benchmark price (see Borenstein, Bushnell and Wolak, 1999). However, to compute the counterfactual competitive benchmark price, researchers need to have information regarding suppliers' marginal costs. Due to the absence of essential information required for calculating marginal costs in the New England electricity market, we are unable to use this method. Instead, we use an alternative approach by conducting a direct comparison of suppliers' bidding behaviors during two distinct periods: one with nuclear refueling outages in the region and one without. By analyzing the differences in bidding behaviors between these two periods, we can explore any potential evidence of suppliers exercising market power. In a perfectly competitive market, the suppliers' bidding behaviors should remain consistent across both periods as long as there are no significant changes in their production costs. It is important to emphasize that our approach is designed to detect the exercise of market power during the temporary baseload supply shortage rather than accurately measuring the full scope of market power. The latter requires more knowledge regarding the competitiveness of the market.

Second, this paper contributes to the literature on market power and strategic firm behavior in deregulated wholesale electricity markets. Extensive studies document market power's presence in deregulated wholesale electricity markets across various regions worldwide. Some well-known examples include Borenstein, Bushnell and Wolak (2002), Wolak (2003), Puller (2007), and Bushnell, Mansur and Saravia (2008) from the California market, Hortaçsu and Puller (2008) and Woerman (2019) from the Texas market, Mansur (2001 and 2008) from the PJM market, Green and Newbery (1992), Wolfram (1999), and Wolak and Patrick (2001) from the British market, Reguant (2014) and Ito and Reguant (2016) from the Spain market, Bask, Lundgren and Rudholm (2011) in Nordic market, and McRae and Wolak (2009) from the New Zealand market. The New England market has been studied by Bushnell and Saravia (2002), Hadsell (2011) and Kim (2019, 2022). Bushnell and Saravia (2002)

and Hadsell (2011) use various approaches to measure market efficiency for the years 1999-2001 and 2003-2007, respectively. Kim (2019, 2022) focus on examining the effect of the volatility of natural gas prices on market power during the winter of 2013. This paper adds to their work by extending the analysis of market power beyond their time frame.

Finally, this paper contributes to the expanding body of literature investigating the sources of market power in the electricity market. Prior research has identified forward contracting (Bushnell, Mansur and Saravia, 2008), capacity limits (Borenstein, Bushnell and Wolak, 1999), transmission constraints (Borenstein, Bushnell and Stoft, 2000; Woerman, 2019), heterogeneity of cost shocks (Kim, 2022), and dynamic costs (Reguant, 2014) as important determinants of the incentives to exercise market power. This paper demonstrates that the negative temporary baseload supply shock caused by nuclear periodic refueling also prompts energy suppliers to strategically adjust their bidding strategies to increase their markup.

3.2 New England Wholesale Electricity Market

The New England wholesale electricity market caters to approximately 6.5 million households and businesses and covers the states of Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont. In the years 2016-2018, nearly \$7.2 billion was transacted annually in this market (ISO-NE *Key Grid and Market Stats*). The wholesale electricity market has been overseen by the Independent System Operator of New England (ISO-NE) since the implementation of restructuring in 1997. The responsibility of ISO-NE is to manage the reliable operation of the regional power grid, conduct market operations, and facilitate the transaction of electricity in the wholesale market.

This wholesale market operates under the uniform-price multi-unit auction, a common practice in deregulated electricity markets. This auction determines the

clearing price and quantity of electricity by establishing a single price that satisfies the market's demand, known as the Energy Component Price (ECP). All successful transactions are settled at the ECP for both electricity suppliers and consumers. However, variations in final prices exist across different location nodes due to transmission constraints and congestion within the system. These factors lead to slight deviations from the ECP, resulting in what is known as the Locational Marginal Price (LMP). The LMP adjusts prices at specific location nodes to incorporate the size of congestion costs and marginal losses.

3.2.1 Day-Ahead, Real-Time, and Ancillary Services Markets

ISO-NE operates two primary markets: the day-ahead energy market and the real-time energy market. The day-ahead energy market enables the supply and demand sides to make commitments to sell or purchase electricity one day prior to the actual generation day. This market is cleared on an hourly basis. The real-time energy market allows market participants to buy and sell electricity to meet immediate demand. This market balances the supply and demand of electricity in near real-time, typically every five minutes, ensuring that there is enough power available to meet the current needs of the region. In addition to these two markets, ISO-NE also manages ancillary services markets such as reserve capacity and forward capacity, ensuring reliability and procuring resources for unexpected situations and future demand.

This paper focuses on the firm behavior and market outcomes in the day-ahead energy market for the following reasons. First, over 95% of the electricity supplied during the next day is scheduled in the day-ahead auction in New England (Kim, 2022). Therefore, understanding the dynamics of this auction is crucial for comprehending the overall wholesale market. Second, compared to the real-time auction, the day-ahead auction offers a more advantageous framework for studying firms' strategic decisions. This is because the day-ahead auction allows firms to plan their operations

and make strategic decisions based on anticipated market conditions with a longer time horizon. In contrast, the real-time auction deals with immediate adjustments to unexpected changes, limiting the scope for strategic decision-making.

3.2.2 Resource Mix

The New England electricity market consists of 123 firms operating over 300 power plants to supply electricity. These power plants utilize a wide range of energy sources for electricity generation, reflecting the diverse resource mix in New England. The energy sources include traditional fuels, such as natural gas, coal, oil, and nuclear power, and renewable energy sources, such as wind power and solar energy. *Table 3.1* presents an overview of power generation categorized by fuel type from 2016 to 2018. The distribution of generation remained consistent during this period. The primary sources for electricity generation were natural gas and nuclear power.

Table 3.1: Fuel Mix (% of Native New England Generation)

	2016	2017	2018
Natural Gas	49%	48%	49%
Nuclear	31%	31%	30%
Hydro	7%	8%	8%
Other*	7%	7%	7%
Wind	2%	3%	3%
Coal	2%	2%	1%
Oil	0%	1%	1%

Note: The “Other” fuel category includes landfill gas, methane, refuse, solar, and steam.

Source: ISO New England (2019)

Nuclear Power Between 2016 and 2018, nuclear power generation accounted for 4,000 MW (13%) of the total capacity fuel mix and approximately 30% of annual energy production in New England. There were three nuclear power plants in New England, Pilgrim Nuclear Power Station in Massachusetts, Seabrook Station in New Hampshire, and Millstone Power Station in Connecticut, with Millstone having two

reactors. These nuclear power plants use enriched uranium as their primary fuel for electricity generation. While these nuclear power plants have high fixed costs associated with construction, maintenance, and operation, their marginal costs are very low.

While in operation, the nuclear reactors generate electricity at full throttle⁵, providing a steady and continuous supply of electricity.⁶ However, each reactor has to swap out used fuel rods every 18 months, which results in a month-long outage. Prior to refueling, nuclear power plant operators are obligated to submit the requests to the ISO-NE for approval, with a minimum notice period of 15 days and a maximum of 730 days. Upon receiving approval from the ISO, the confirmed schedules for refueling outages are announced to all market participants. In New England, the gap left by nuclear refueling outages is mainly filled by natural gas generation (Davis and Hausman, 2016; Kim, 2019). Since nuclear power has a substantial share in electricity generation in New England, the refueling outage of any reactor in the region creates a significant supply shock in the market. This presents us with excellent exogenous shocks and a well-balanced set of observations to study other energy suppliers' strategic behaviors.

Natural gas Natural gas played a dominant role in New England's electricity generation, accounting for 49% (approx. 5800 MW per hour) of annual energy generation between 2016 and 2018. Compared to power plants utilizing other energy sources, natural gas power plants have relatively higher marginal costs. As a result, they are often located near the clearing price in electricity markets. In the context of

⁵There are two exceptions to this continuous full generation. Firstly, there is a gradual decrease and subsequent increase in the generation of nuclear reactors a few days before and after a refueling outage. Secondly, the Pilgrim Nuclear Power Station's reactor, due to its age, requires frequent maintenance and repairs, which occasionally result in the reactor not operating at its maximum capacity.

⁶It is worth noting that, although nuclear power plants have significant market shares and low marginal production costs, they are unable to exercise market power because, unlike natural gas power plants, nuclear power plants have limited flexibility in adjusting their generation capacity.

a uniform price auction, these natural gas power plants have a higher ability to exercise market power by strategically adjusting their bids to influence the final clearing price. Therefore, natural gas power plants are the primary focus of this paper.

Table 3.2: Major Five Firms Gas Generation Capacity in 2016

	Firm	Capacity(MW)	# of Plants
1	Exelon Generation Company, LLC	2639.419	4
2	Calpine Energy Services, LP	1930.353	5
3	Dynegy Marketing and Trade, LL	1674.981	7
4	GDF Suez Energy Marketing NA	1002.45	4
5	Repsol Energy North America Co	644.016	3
	Top 5	7889	23
	Total	13752	75

Table 3.2 shows summary statistics by firm for the five largest gas generation firms in the market in 2016. The three largest firms held shares of 19%, 14%, and 12%, respectively, in terms of natural gas generation capacity, and their market shares were even less. According to traditional measures of competitiveness, such as the Herfindahl–Hirschman Index, the market did not exhibit a high degree of concentration that would raise concerns about market power. However, as mentioned in the introduction, these measurements are poor indicators for the existence of market power in electricity markets due to the unique nature of electricity (Wolfram, 1999; Wolak, Borenstein and Bushnell, 2002). Therefore, to reach a reliable conclusion on the presence of market power in an electricity market, conducting an empirical investigation is indispensable.

3.3 Data

This paper aims to examine the effect of nuclear refueling outages on market-clearing prices and market power in the New England wholesale electricity market.

One notable advantage of the deregulated market is its high level of transparency, as the rich market transaction data is publicly accessible through ISO-NE.

We study the period from 2016 to 2018, during which the nuclear power capacity in this region remains unchanged.⁷ We further restrict our sample period to the spring and fall months (March 1 to May 31 and September 1 to November 30). Excluding the summer and winter seasons provides many additional advantages in our analysis. Firstly, it allows for a more accurate estimation of nuclear refueling outage impact since such outages have never occurred during the high-demand summer and winter months. Secondly, extremely cold winters and hot days can introduce additional complexities into the energy market, such as electricity transmission congestion and natural gas pipeline constraints. Lastly, coal and oil-fired power plants were rarely in operation in the spring and fall months during our study period. Hence, we do not need to control the cost of coal and gasoline in our analysis.

We construct a panel dataset, incorporating five important features: hourly day-ahead cleared prices, hourly day-ahead net demand, hourly day-ahead offer curves submitted by each power plant, daily fuel cost, and daily nuclear utilization. In the subsequent part of this section, we provide a detailed discussion of each data source and explain how we construct the main variables of interest.

Day-ahead cleared prices The data of hourly day-ahead market-cleared prices is available from *Day-Ahead Energy Market Hourly LMP Report* published by ISO-NE. As mentioned earlier, there are two types of prices in the wholesale market, the ECP and the LMP. The ECP represents the uniform cleared price for the entire market, while the LMP reflects the location-specific price that takes congestion and marginal loss into account. In our analysis, we focus on the ECP rather than the LMP due to the complexity of the LMP system and the lack of detailed information

⁷The Yankee Nuclear Power Station in Yankee was decommissioned in December 2014, and the Pilgrim Nuclear Power Station in Massachusetts ceased operations in 2019.

regarding the ISO’s market clearing algorithm. In practice, the differences in LMPs across nodes are not significant during spring and autumn when severe transmission constraints are less likely to occur.

Day-ahead net demand In the short term, end consumers have limited ability to adjust their electricity consumption patterns, and they usually face retail prices that are set in advance and do not vary with wholesale prices. As a result, the electricity demand in the wholesale market is not highly responsive to changes in wholesale prices. Given that demand directly impacts market-clearing prices and the bidding strategy of suppliers, our analysis incorporates demand as a control variable. The data of the hourly day-ahead cleared demand was obtained from ISO-NE *Day-Ahead Energy Market Hourly Demand Report*. Additionally, a portion of the electricity demand in New England is met by imports from Canada and New York. We deducted the imported electricity from the overall market demand to calculate the net demand that must be fulfilled by the internal market supply.

Bidding behavior When a power plant wants to participate in the day-ahead energy market, it is required to submit a step-function offer curve to ISO-NE for each of its units one day before the scheduled electricity generation. The offer curve specifies how much electricity a unit will produce at every price, with each unit being allowed to bid up to 10 steps. *Energy Offer* data published by ISO-NE contains all the submitted offer curves, which enables the observation of hourly bidding behaviors of suppliers in this market. However, the specific identity of each power plant and its affiliation with a particular firm remain unknown because the plant IDs in the dataset are represented by unstructured codes.

In our analysis of strategic decisions made by energy suppliers, we focus on “economic” power plants that have the ability to adjust their output as needed, such as natural gas-fired power plants, and we exclude “must-run” power plants that operate

continuously at relatively consistent output levels and lack the capability to control their output, such as nuclear and large-scale hydropower.⁸

In *Figure 3.2*, we constructed two example market-wide aggregate offer curves, which sum the quantity of all “economic” unit-level offer curves at two specific time points. These two time points exhibit similar demand and gas prices but significantly different nuclear utilization rates. On September 27, 2018, nuclear power was nearly fully utilized at 93%, but on October 11, 2018, it dropped to 30% due to three out of the four reactors being in outage. The figure clearly illustrates a notable rightward shift in the supply curve during nuclear outages, indicating that suppliers adjust their bidding behaviors in response to these events.

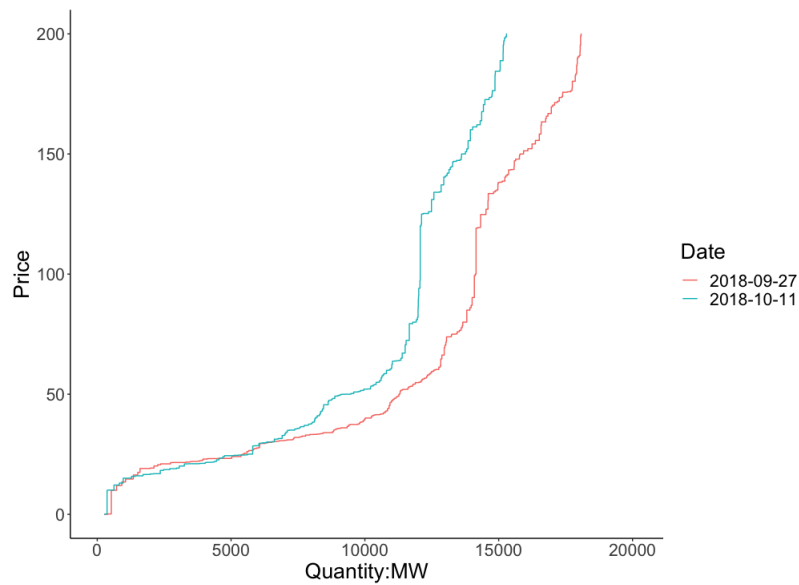


Figure 3.2: Market-wide aggregated offer curves.
2018/09/27: nuclear utilization at 93%;
2018/10/11: nuclear utilization at 30%.

Fuel cost The primary component of variable costs for a fossil-fueled power plant is the fuel cost. Since natural gas is the predominant fuel for electricity generation

⁸Although we lack specific details on ISO-NE’s classification process, it is reasonable to infer that the majority of power plants categorized as “economic resources” are fueled by natural gas.

in New England, natural gas prices greatly influence wholesale electricity prices. We incorporate the natural gas price in our analysis to approximate the marginal cost function of natural gas power plants. We obtained daily Henry Hub natural gas prices from EIA.⁹ Henry Hub serves as a major pricing center for natural gas in the United States and is widely accepted as a benchmark for gas trading. When there are no pipeline constraints, Henry Hub gas prices are representative of the purchasing costs for natural gas power plants in New England.

Nuclear power reactor status In our analysis, nuclear utilization serves as the primary variable of interest. To gather information on nuclear power plant operations, we collected data from the Nuclear Regulatory Committee. Based on this information, we created two variables to describe nuclear outages: one is binary and the other is continuous. Firstly, we created a dummy variable to indicate the specific timeframe when at least one nuclear reactor in the region was offline. The cutoff threshold for this was set at 84% of nuclear utilization. Additionally, we computed a continuous variable of the daily nuclear utilization rate by dividing the total daily nuclear generation by the overall nuclear capacity.

See *Table 3.3* for descriptive statistics of the variables used in the empirical analysis, grouped into two categories based on the presence or absence of any offline nuclear reactors.

3.4 Empirical Analysis

The goal of this analysis is to examine how negative supply shocks resulting from nuclear periodic refueling influence (i) market-clearing prices and (ii) bidding strategies employed by “economic” energy suppliers.

⁹Algonquin Gas Transmission Pipeline (AGT) and Tennessee Gas Pipeline (TGP) deliver most of the natural gas from Henry Hub to New England.

Table 3.3: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
251 days with nuclear outages							
Price (\$/MWh)	6,024	36.356	18.741	4.360	24.200	42.447	179.200
Demand (MWh)	6,024	12,238	1,907.716	8,145	10,748	13,533.5	21,864
Gas (\$/MMBtu)	251	2.930	0.659	1.490	2.710	3.200	4.700
298 days without nuclear outages							
Price (\$/MWh)	7,152	26.694	12.459	1.000	18.700	31.242	120.750
Demand (MWh)	7,152	12,641	2,179.510	7,346	10,955	14,061	22,426
Gas (\$/MMBtu)	298	2.708	0.440	1.570	2.670	3.000	3.320

Note: This table displays the information on hourly market cleared prices, hourly market net demand, and daily natural gas prices categorized into two groups: days with nuclear outages and days without nuclear outages.

First, we estimate two following equations to check whether the clearing price goes up in the event of nuclear outages and quantify the magnitude of this impact.

$$P_t = \alpha_{outage} \mathcal{I}\{\% \text{ nuclear} < 0.84\} + \beta_1 X_t^D + \beta_2 X_t^S + \omega_t + \epsilon_t \quad (3.1)$$

$$P_t = \alpha_{util} \{\% \text{ nuclear}\} + \beta_1 \mathbf{X}_t^D + \beta_2 \mathbf{X}_t^S + \omega_t + \epsilon_t \quad (3.2)$$

where P_t on the left-hand side of both equations represents the market-cleared price at hour t . \mathcal{I} is the dummy variable indicating whether at least one reactor in the region is offline. α_{outage} and α_{util} are the parameters of interest in *Equation (3.1)* and *Equation (3.2)*, respectively. While α_{outage} represents the discrete average effect of nuclear refueling outages on the market-cleared price, α_{util} illustrates the continuous relationship between nuclear utilization and cleared prices. X_t^D and X_t^S are demand side and supply side controls, namely day-ahead cleared demand and natural gas price. ω_t is the time-fixed effect that absorbs the effect of emission costs among others. Both equations are estimated fixed-effect regression models.

Identifying the occurrence of higher prices during nuclear refueling outages is insufficient to prove that “economic” energy suppliers are exercising market power. The price effect could be solely attributed to a shift in the marginal generator to

one with higher marginal costs. Therefore, our next step is to further analyze the suppliers’ bidding behaviors. We estimate two following equations to examine whether “economic” energy suppliers modify their bidding behaviors in response to nuclear outages.

$$BiddingPrice_{it} = \lambda_{outage} \mathcal{I}\{\% \text{ nuclear} < 0.84\} + \beta_1 X_t^D + \beta_2 X_t^S + \omega_t + \gamma_i + \epsilon_{it} \quad (3.3)$$

$$BiddingPrice_{it} = \lambda_{util} \{\% \text{ nuclear}\} + \beta_1 X_t^D + \beta_2 X_t^S + \omega_t + \gamma_i + \epsilon_{it} \quad (3.4)$$

for generating unit i on time t . On the left-hand side of both equations is each unit’s bidding price at a certain percentage of its capacity.¹⁰ Given that the suppliers’ offer curves are step functions, we have to choose specific points along these curves. Following Woerman (2019), we look at the bidding prices submitted at three quantities of production: 65%, 75%, and 85%. The rationale for focusing on high quantities is that it increases the likelihood of observing the exercise of market power. For natural gas power plants, they opt for bidding at lower prices at low quantities to avoid the fixed cost associated with shutdown and restart processes and tend to bid above their marginal cost at high quantities. λ_{outage} and λ_{util} are the parameters of interest in *Equation (3.3)* and *Equation (3.4)*, respectively. λ_{outage} in *Equation (3.3)* enables us to check whether the “economic” energy suppliers adjust their bidding strategies during the nuclear refueling outage. λ_{util} in *Equation (3.4)* further captures the continuous relationship between nuclear utilization and suppliers’ bidding behaviors. The variables we control on the right-hand side of both equations are almost the same as those in *Equations (3.1)* and *(3.2)*. The only difference is that we control γ_i , the unit fixed effect, in *Equations (3.3)* and *(3.4)*. This fixed effect absorbs unobserved het-

¹⁰Computing the suppliers’ markup and using them as dependent variables is a more common approach. However, the energy offer data provided by ISO-NE masks all participant IDs, making it extremely difficult to merge with CEMS data from the US Environmental Protection Agency (EPA). The CEMS data is crucial for calculating the heat rates and emission costs of units.

erogeneity at the unit level, such as heat rates. We use fixed-effect regression models to estimate both equations.

We finally estimate the welfare effects during nuclear refueling outages. As previously mentioned, due to the inelastic nature of short-term electricity demand, the increase in market prices does not change the quantity of electricity generated and consumed. As a result, the exercise of market power during nuclear refueling outages does not lead to welfare loss in the short term but rather causes transfers from consumers to suppliers. We estimate welfare transfer by multiplying the coefficients of the main variable of interest, λ_{outage} , in *Equation (3.3)* with the total electricity generation during the days with nuclear outages.

3.5 Results

We start with estimating the effect of nuclear refueling outages on market prices as described in *Equations (3.1)* and *Equation (3.2)*. The results from the fixed-effect regressions are shown in *Table 3.4*. We estimated four different specifications for each equation. Because these specifications yield consistent and similar results, we focus on Columns (7) and (8), which include week-year FEs, hour FEs, and control variables. Column (7) indicates a statistically significant increase of \$4.86 in the market-cleared price when at least one reactor is offline. Column (8) gives a point estimate of the marginal effect of lost nuclear power generation on market prices. For each percentage decrease in nuclear utilization, the market-cleared price increases by \$0.23. With the average nuclear utilization rate at 67% during the days with nuclear outages, our estimation suggests the average effect size to be \$7.59.

After confirming that nuclear refueling outages have a significant impact on market prices, we proceed to estimate the discrete average impact of these outages on suppliers' bidding behaviors, as described in *Equation (3.3)*. The results from the fixed-effect regressions are displayed in *Table 3.5*. This table contains the results of

Table 3.4: Nuclear refueling outages & utilization rates – Effect on market prices

Dependent Variable: Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Variables</i>								
$\mathcal{I}\{\% \text{ nuclear} < 0.84\}$	4.693** (2.279)	-22.89** (9.909)	4.818** (1.537)	-22.40** (7.961)	4.680** (0.2928)	-22.92** (1.232)	4.863** (0.2739)	-22.52** (1.131)
% nuclear								-56.83** (2.859)
Gas			-53.40** (18.49)	-53.61** (18.43)			-56.61** (2.858)	10.85** (0.4666)
Gas ²			10.74** (3.419)	10.73** (3.398)			10.87** (0.4666)	0.0055** (0.0002)
Demand			0.0042** (0.0002)	0.0042** (0.0002)			0.0055** (0.0002)	
<i>Fixed-effects</i>								
Week FEs	✓	✓	✓	✓	✓	✓	✓	✓
Year FEs	✓	✓	✓	✓	✓	✓	✓	✓
Week × Year FEs	✓	✓	✓	✓	✓	✓	✓	✓
Hour FEs					✓	✓	✓	✓
Observations	13,176	13,176	13,176	13,176	13,176	13,176	13,176	13,176
R ²	0.54900	0.55055	0.78993	0.79114	0.66843	0.67000	0.81043	0.81164

Note: This table presents the results of regression estimations for Equations 3.1 and 3.2, utilizing various combinations of control variables. The dependent variable is the market-cleared price. The independent variables of interest are the indicator of nuclear outages in Columns (1), (3), (5), and (7), and the nuclear utilization rate in Columns (2), (4), (6), and (8). Clustered standard errors are shown in parentheses. Significance: ** $p < 0.01$, * $p < 0.05$, $p < 0.1$.

Table 3.5: Nuclear refueling outages – Effect on suppliers’ bidding price

Dependent Variable: Model:	Offer price	
	(1)	(2)
Panel A: Offer price at 65% $\mathcal{I}\{\% \text{ nuclear} < 0.84\}$	1.782** (0.7842)	0.7247 (1.308)
Panel B: Offer price at 75% $\mathcal{I}\{\% \text{ nuclear} < 0.84\}$	1.834** (0.7826)	1.848** (0.7834)
Panel C: Offer price at 85% $\mathcal{I}\{\% \text{ nuclear} < 0.84\}$	1.727** (0.7661)	1.743** (0.7677)
Control	✓	✓
Unit FEs	✓	✓
Hour FEs		✓
Week FEs	✓	✓
Year FEs	✓	✓
Week×Year FEs	✓	✓
Observations	775,122	775,122

All regressions include controls for Gas, Gas², and net demand.

Clustered standard errors are shown in parentheses.

*Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.*

regression estimations for suppliers’ bidding prices at three production levels: 65%, 75%, and 85% in Panels A to C, respectively. For each quantity level, the table displays the outcomes from two different specifications: Column (1) excludes the hour fixed effect, while Column (2) includes it. Let’s first look at the regression results for the bidding prices at 65% capacity in Panel A. The results from the two regression models differ greatly. Column (1) shows a significantly positive effect of nuclear refueling outages. It indicates that, on average, suppliers bid \$1.78 higher during nuclear refueling outages. However, the effect size diminishes to \$0.72, and the significance disappears after we account for the hourly fixed effect in Column (2). The

inconsistent results may reflect the thermal power plants' preference for not bidding higher at lower quantities to ensure uninterrupted operation. In Panel B, we observe the consistent and significant effects in both regressions. Suppliers tend to bid \$1.85 higher at their 75% capacity during nuclear refueling outages. Moving on to Panel C, the effects of nuclear outages on bidding prices at 85% capacity are similar to those observed at 75% capacity, although the effect size drops to \$1.74.

Table 3.6: Nuclear utilization rate – Effect on suppliers' bidding price

Dependent Variable: Model:	Offer price	
	(1)	(2)
Panel A: Offer price at 65% % Nuclear	-9.508*** (2.494)	-9.554*** (2.492)
Panel B: Offer price at 75% % Nuclear	-9.829*** (2.465)	-9.878*** (2.463)
Panel C: Offer price at 85% % Nuclear	-9.782*** (2.656)	-9.836*** (2.654)
Control	✓	✓
Unit FEs	✓	✓
Hour FEs		✓
Week FEs	✓	✓
Year FEs	✓	✓
Week×Year FEs	✓	✓
Observations	775,122	775,122

All regressions include controls for Gas, Gas², and net demand.

Clustered standard errors are shown in parentheses.

*Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.*

Next, we delve deeper into examining the continuous relationship between nuclear utilization and suppliers' bidding behaviors, as outlined in *Equation 3.4*. Similarly, *Table 3.6* reports bidding prices for all three quantities and presents two specifica-

tions for each quantity level. The results reveal consistently significant positive effects across all six regressions. We find that for each one less percentage of nuclear utilization in the region, suppliers tend to bid an additional \$0.096, \$0.099, and \$0.098 for their 65%, 75%, and 85% capacity, respectively. Given that the average nuclear utilization rate during the days with nuclear outages is 65%, the average effect size is \$3.168.

It is noteworthy that there were 28 days when three nuclear reactors in the region were offline during our study period. The corresponding nuclear utilization rate was around 30%. According to our models, suppliers would increase their bidding prices by \$6.86 during these days. Such a substantial change in suppliers' bidding behaviors should draw our attention and concern.

Lastly, we proceed to estimate welfare transfer during nuclear power outages. From *Table 3.5*, we find that suppliers did strategically bid for higher prices when at least one of the four reactors was offline. Assuming the day-ahead market cleared at 75% or 85% capacity of a unit, this results in approximately \$1.74 – \$1.85 price increase attributable to the exercise of market power. Multiplying this number by the total generation during the days with nuclear outages, we estimate a welfare transfer of around \$128 to \$136 million from consumers to suppliers during our study period, with an average of \$43 to \$45 million per year. This accounts for roughly 0.625% of the overall market transaction volume.

3.6 Conclusion

We utilize the phenomenon of periodic nuclear refueling outages as a test for market power in the electricity market. Using hourly wholesale market data from New England, we find that the market-cleared price increased by \$4.9/MWh, of which market power contributed to approximately 36% of the effect. A back-of-the-envelope calculation shows there is a \$43 million dollar welfare transfer from consumers to

suppliers due to the exercise of market power. Although the focus of this study is the New England wholesale electricity market, its findings provide valuable insights for other deregulated markets as well.

The findings of this paper call the regulators into action. While prior research has established that the exercise of market power is usually observed during high-demand summer and winter months, this paper provides empirical evidence that it also happens during the spring and fall seasons. Therefore, market operators should remain vigilant in overseeing suppliers' bidding behaviors in the spring and fall months, particularly in cases of temporary baseload supply shocks.

In addition, it is crucial for regulators to be fully prepared during the transition to low-carbon markets, as our findings further suggest. The reason behind this necessity lies in the fact that most renewable energy is derived from intermittent sources. During periods when these sources are not actively generating power, there will be a considerable likelihood of market power abuses occurring due to baseload supply shortages. This trend has already been observed in the California electricity market. According to Butters, Dorsey and Gowrisankaran (2021), as solar panel generation had been increasing in California from 2015 to 2019, daytime electricity prices dramatically decreased, while nighttime prices continued to rise. Therefore, regulators must consider strategies to mitigate the potential risks of market power abuses caused by the increase in intermittent power generation.

The electricity market is an intricate market with various factors that need to be taken into account. Despite our best efforts to control it, there are still certain limitations that necessitate further investigation. Firstly, our analysis only considers the scheduled maintenance of nuclear reactors and does not account for the maintenance activities conducted by other energy suppliers. Prior studies have highlighted that, apart from bidding at higher prices, firms owning multiple power plants can strategically schedule maintenance to manipulate market-clearing prices. Future research can

delve into this aspect and explore whether suppliers employ such tactics to exercise market power during temporary baseload supply shocks.

Another limitation of this paper is our estimations regarding market welfare. In order to achieve more precise estimations, it is necessary to compute the marginal cost of each unit. Unfortunately, due to the limitations presented by the available data, we are unable to carry out this calculation as part of this research endeavor. However, there have been studies (such as Ryan, 2021 and Kim, 2022) suggesting an alternative method to estimate units' marginal costs based on their bidding strategies. Their approach may open up a promising avenue of in-depth investigation into how temporary baseload supply shocks affect market welfare.

APPENDIX A

APPENDIX OF CHAPTER 1

A.1 Instructions for NR_7 and R_7

To save space, I report the instructions for NR_7 , underline the parts that differ from those in R_7 , and show the variation in curly brackets. The instructions for $NR_{14.5}$ and $R_{14.5}$ are skipped because the only difference between NR_7 and $NR_{14.5}$, as well as R_7 and $R_{14.5}$, is that buyers need to pay 7.5 points more in $R_{14.5}$ and $NR_{14.5}$ when they search for a second opinion.

Instructions

Thank you for participating in this experiment. In case you have a question, you can pop up your question through the chat. I will respond to your question after I finish reading the instructions.

You are paid \$5 show-up bonus to be here on time. You can earn additional money depending on your decisions during the experiment. Below, the instructions specify how participants make decisions in today's experiment. To be able to earn more money, you will need to understand the instructions. After we finish reading the instructions together, you will be asked to take a short quiz to test your understanding. You WILL NOT be able to proceed to the experiment until you answer all quiz questions correctly. So please pay attention when I read the instructions.

The experiment is about the decisions of buyers and sellers in a market. Buyers need to choose a seller to purchase a service; sellers need to choose the type of service to offer to a buyer. In today's experiment, both sellers and buyers can earn "points"

from their market transactions. At the end of the experiment, we will calculate your dollar earnings according to the following exchange rate:

$$50 \text{ points} = \$1$$

Market Roles

At the beginning of the experiment, you will be randomly assigned to be either a buyer or a seller. You will keep playing in the SAME role for the entire experiment. On the first screen of the experiment, you will see which role you are assigned to.

There will be 4 sellers and 4 buyers in a market. These eight people remain in the same market throughout the experiment. Once you are assigned to a market and given a role, you start interacting with other players in the market repeatedly. We call each of these repeated interactions a “period”. There are 20 periods in total.

Within a period, you are given an ID number. Seller IDs can be either S1, S2, S3, or S4. Buyer IDs can be either B1, B2, B3, or B4.

Attention: Both Sellers’ IDs and Buyers’ IDs will be re-assigned randomly at the beginning of each period. That is to say, decision-makers behind each ID are different for each new period. However, the role each player plays is fixed throughout the experiment.

[In R₇: Attention: Sellers’ IDs are fixed throughout the experiment. The same ID number always represents the same seller. However, buyers’ IDs will be re-assigned randomly at the beginning of each period. That is to say, buyers behind each ID are different in each new period.]

Decision Sequence

Once buyer and seller IDs are assigned, they make decisions in sequence to complete a market transaction. The instructions below detail the sequence. At the beginning of a period, each buyer is randomly assigned a service problem: either Problem X or Problem Y. The type of problem is determined randomly and is unaffected by other buyers' problems. Problem X happens with a 25% chance. Problem Y happens with a 75% chance. Buyers do not know which type of service problem they have. To fix the unknown problem, the buyers have to interact with the sellers.

Unlike buyers, sellers can identify the type of problem each buyer has. Sellers can solve a buyer's problem by choosing one of the two possible actions, Action 1 or Action 2. Knowing that a buyer has Problem X, a seller MUST choose Action 1 to solve it. But if a buyer has Problem Y, a seller could choose to fix the problem using either Action 1 or Action 2.

Moreover, the seller makes a total of four decisions, one towards each buyer. The following screenshot shows the decision interface for a particular seller. Note, since B2 has Problem X, this seller has no choice but choosing Action 1. For the other three buyers who have Problem Y, this seller could choose either Action 1 or Action 2.

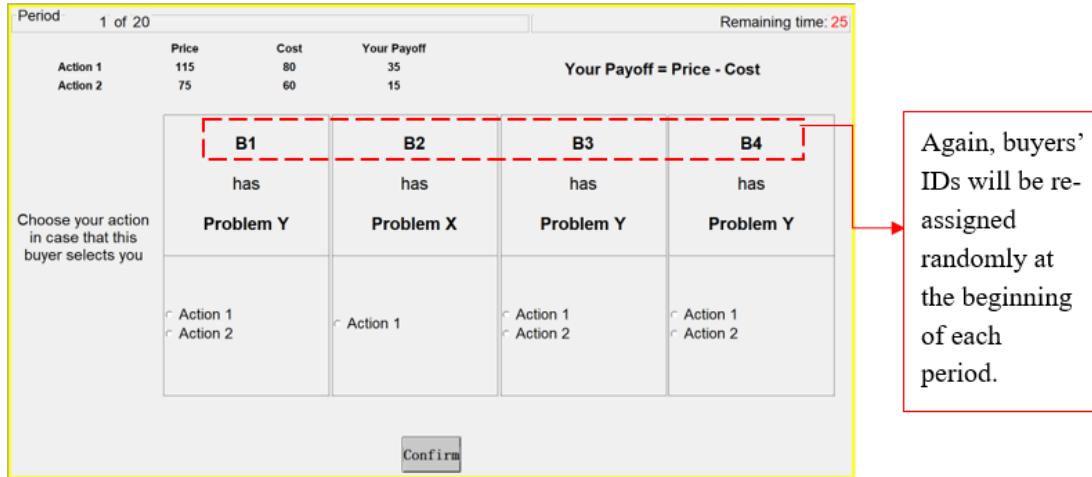


Figure A.1: Example seller's decision screen

Note that although a seller is choosing an action for each buyer, it does not necessarily mean that the seller will sell the service to all buyers. You can consider these seller's choices as "recommendations" to the buyers. Only in the case that a buyer chooses to interact with a particular seller and accept the proposed action, that seller's choice would be used to calculate his actual payoffs. For example, if a seller chooses Action 1 for all four buyers, but only B1 and B3 accept the actions provided by this particular seller, the seller's choice for B2 and B4 will not be used for payoff calculation.

In this experiment, while sellers make their “recommendations”, buyers need to select one of the four sellers in the market to interact with. Buyers make this selection without knowing what action each seller had chosen for him/her. The following screenshot shows the decision interface for a particular buyer.

Period 1 of 10 Remaining time: 0

Action	Price	Cost	Your Payoff
Action 1	115	80	15
Action 2	75	60	55

Your Payoff = 130 - Price

This is your first search in the current period

Please choose the seller you wish to interact with.

- S1
- S2
- S3
- S4

Confirm

Again, sellers' IDs are fixed throughout the experiment. However, the display order will be changed for each period.

Figure A.2: Example buyer's first decision screen

After the buyer selects a seller, s/he will observe the action offered by the selected seller. Now, the buyer has two options: (1) accept the action or (2) pay a cost of 7 points to search for another seller from the remaining three sellers. If the buyer chooses to accept the action, s/he pays the seller the price for the action and the period ends. If the buyer chooses to search for another seller, s/he pays a cost of 7 points and receives the action offered by the new seller. The action offered by this second seller is final. The buyer must accept it. The buyer does NOT have the option to select again, nor can s/he go back to accept the offer from the first seller.

The following screenshot shows an example of the decision interface for a particular buyer when s/he has selected a seller (in this case S1) and is asked to either accept the action (in this case, Action 1) or to search for another seller.

Period	1 of 20			Remaining time: 9
	Price	Cost	Your Payoff	
Action 1	115	80	15	Your Payoff = 130 - Price
Action 2	75	60	55	

You selected **S1** and he/she chose to perform **Action 1**.

Please decide whether to search for another seller.
 Notice: searching for another seller entails an additional cost of **7 points**.

Accept the current action.
 Search for another seller.

Figure A.3: Example buyer's second decision screen

If the buyer selects “Search for another seller” in the screen shown above, s/he will see the interface below to select another seller from the remaining three sellers. Then the buyer receives the action offered by the new seller and pays the corresponding price. Notice that S1 is not on the option list. That is because this buyer chose to not accept the action offered by S1 in the previous decision screen.

Period	1 of 10			Remaining time: 28
	Price	Cost	Your Payoff	
Action 1	115	80	8	Your Payoff = 130 - 7 - Price
Action 2	75	60	48	
This is your second search in the current period				
Please choose the seller you wish to interact with.				
<input type="radio"/> S2 <input type="radio"/> S3 <input type="radio"/> S4				
<input type="button" value="Confirm"/>				

Figure A.4: Example buyer’s third decision screen

[Exclusively available in R₇: Starting from the second period, each buyer can see a history table on the right half of the screen when s/he makes decisions. The table displays interactions from all previous periods, including information on the selected seller's ID and the action offered by that seller in both the first and the second searches. If a buyer did not go through a second search, the related field will show a pound sign “#”.

The following screenshot shows an example of the interface for a particular buyer during Period 4. In this example, the first row of the table shows you that this particular buyer knows that in Period 1, s/he selected S1 during the first search and S1 chose Action 2. This particular buyer accepted S1's action; therefore, there is no information provided for the second search.]

Period 4 of 20		Remaining time: 15																											
Action 1	Price 115	Cost 80	Your Payoff 15	Your Payoff = 130 - Price																									
Action 2	75	60	55																										
<p>This is your first search in the current period</p> <p>Please choose one of sellers that you would like to interact with.</p> <ul style="list-style-type: none"> <input type="radio"/> S1 <input type="radio"/> S2 <input type="radio"/> S3 <input type="radio"/> S4 			<table border="1"> <thead> <tr> <th colspan="5">Trade History</th> </tr> <tr> <th>Period</th> <th>First Search</th> <th>First Seller Action</th> <th>Second Search</th> <th>Second Seller Action</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>S1</td> <td>2</td> <td>#</td> <td>#</td> </tr> <tr> <td>2</td> <td>S4</td> <td>1</td> <td>S3</td> <td>1</td> </tr> <tr> <td>3</td> <td>S1</td> <td>1</td> <td>#</td> <td>#</td> </tr> </tbody> </table>		Trade History					Period	First Search	First Seller Action	Second Search	Second Seller Action	1	S1	2	#	#	2	S4	1	S3	1	3	S1	1	#	#
Trade History																													
Period	First Search	First Seller Action	Second Search	Second Seller Action																									
1	S1	2	#	#																									
2	S4	1	S3	1																									
3	S1	1	#	#																									
<input type="button" value="Confirm"/>																													

Figure A.5: Example buyer's first decision screen in Period 4 in R₇

In summary, in each period, sellers know the type of problem each buyer has and recommend an action to solve each problem. Simultaneously, each buyer chooses a seller without knowing what action that seller has chosen. The buyers then observe the chosen seller's action and decide on whether to accept the action or to search for a second seller. Note: in case the buyer did not accept the first seller, that buyer must accept the action from the seller chosen in the second search. There is no more search after that. Also, the buyer cannot go back to the first selected seller once the second search begins.

Payoffs

You may wonder how Action 1 and Action 2 impact your payoffs. Here are the details.

For Action 1, a seller charges a buyer the price of 115 points and pays the cost of 80 points.

For Action 2, a seller charges a buyer the price of 75 points and pays the cost of 60 points.

Seller's earnings for a particular period are the sum of payoffs from all buyers who choose to accept that seller's action. The acceptance includes offers accepted during both buyers' first and second searches. For each accepted offer,

$$Payoff_{\text{seller}} = Price - Cost$$

If Action 1 is offered and accepted, the seller's payoff is 35 points (=115-80).

If Action 2 is offered and accepted, the seller's payoff is 15 points (=75-60).

If none of the seller's offers is accepted, the seller receives 0 points for the period.

A seller's total earnings are the sum of earnings across all 20 periods.

On the other hand, buyer payoff is calculated as follows:

$$Payoff_{\text{buyer}} = \begin{cases} 130 - Price_{\text{first}} & \text{if the buyer accepts the first seller's action} \\ 130 - Price_{\text{second}} - 7 & \text{if the buyer searches for a second seller} \end{cases}$$

If Action 1 is offered and accepted in the first search, the buyer payoff is 15 points (= 130 - 115).

If Action 2 is offered and accepted in the first search, the buyer payoff is 55 points (= 130 - 75).

If either of these actions is accepted in the second search, the buyer payoff will be 7 points lower. That is, 8 points for Action 1 and 48 points for Action 2.

A buyer's total earnings are the sum of payoffs across all 20 periods.

Notice that a buyer cannot find out what type of problem s/he has even after the payoff is determined and revealed (the buyer observes their own payoff, not the seller's payoff).

The information about the prices and costs of two different actions and your payoff will be listed on the top of the interface. These parameters are fixed throughout the experiment.

A.2 Control Questions for NR_7 and R_7

The control questions for NR_7 and R_7 are exactly the same. Again, the only difference between NR_7 and $NR_{14.5}$, as well as R_7 and $R_{14.5}$, is that buyers need to pay 7.5 points more in $R_{14.5}$ and $NR_{14.5}$ when they search for a second opinion. As a result, the answer to Q4 is 0.5 rather than 8 in $R_{14.5}$ and $NR_{14.5}$.

Please answer the following questions:

Q1: How many periods in total are there in this experiment?	Answer: 20
<p>Consider the following hypothetical scenario:</p> <p>Suppose Buyer B1 selected Seller S2 as his first-choice seller and observed that S2 chose to perform Action 1. Then, B1 decided to switch to S3 in the second search and observed that S3 chose to perform Action 1 as well.</p>	
Q2: How many points did S2 earn in this scenario?	Answer: 0
Q3: How many points did S3 earn in this scenario? Hint: The price of Action1 – The cost of Action 1.	Answer: 35
Q4: How many points did B1 earn in this scenario? Hint: 130 - The price of Action 1 - 7 (search cost).	Answer: 8
Q5: Will B1 be able to go back to S2 and accept the offer from S2 if s/he chooses to search for a second seller? Yes/No	Answer: No

A.3 Questionnaires

1. Please enter your UMass e-mail address where we can send you the Amazon eGift card.

2. Which year were you born?

3. What gender are you identifying with?

- Male
- Female
- Other

4. Which academic cohort do you belong to as of this Fall semester?

- Freshman
- Sophomore
- Junior
- Senior
- Graduate Student
- Non-degree Seeker

5. Which school are you majored in?

- College of Education
- College of Engineering

- College of Humanities and Fine Arts
- College of Information and Computer Sciences
- College of Natural Sciences
- College of Nursing
- College of Social and Behavioral Sciences
- Isenberg School of Management
- School of Public Health and Health Sciences
- Stockbridge School of Agriculture
- Commonwealth Honors College
- Undetermined

6. What is your current GPA?

- 3.5 – 4.0
- 3.0 – 3.49
- 2.0 – 2.99
- Below 2.0

7. Have you ever taken any classes in economics?

- Yes
- No

APPENDIX B

APPENDIX OF CHAPTER 2

B.1 Instructions for PRI, AGG_P, and AGG_P&PH

To save space, I report the instructions for PRI, underlining the parts that differ from those in AGG_P and AGG_P&PH. I also show the variation for AGG_P in parentheses and the variation for AGG_P&PH in curly brackets.

Instructions

Thank you for participating in this experiment. In case you have a question, you can pop up your question through the chat. I will respond to your question after I finish reading the instructions.

You are paid \$5 show-up bonus to be here on time. You can earn additional money depending on your decisions during the experiment. Below, the instructions specify how participants make decisions in today's experiment. To be able to earn more money, you will need to understand the instructions. After we finish reading the instructions together, you will be asked to take a short quiz to test your understanding. You WILL NOT be able to proceed to the experiment until you answer all quiz questions correctly. So please pay attention when I read the instructions.

The experiment is about the decisions of buyers and sellers in a market. Buyers need to choose a seller to purchase a service; sellers need to choose the type of service to offer to a buyer. In today's experiment, both sellers and buyers can earn "points" from their market transactions. At the end of the experiment, we will calculate your dollar earnings according to the following exchange rate:

50 points = \$1

Market Roles

At the beginning of the experiment, you will be randomly assigned to be either a buyer or a seller. You will keep playing in the SAME role for the entire experiment. On the first screen of the experiment, you will see which role you are assigned to.

There will be 4 sellers and 4 buyers in a market. These eight people remain in the same market throughout the experiment. Once you are assigned to a market and given a role, you start interacting with other players in the market repeatedly. We call each of these repeated interactions a “period”. There are 20 periods in total.

Within a period, you are given an ID number. Seller IDs can be either S1, S2, S3, or S4. Buyer IDs can be either B1, B2, B3, or B4.

Attention: Sellers’ IDs are fixed throughout the experiment. The same ID number always represents the same seller. However, buyers’ IDs will be re-assigned randomly at the beginning of each period. That is to say, buyers behind each ID are different in each new period.

Decision Sequence

Once buyer and seller IDs are assigned, they make decisions in sequence to complete a market transaction. The instructions below detail the sequence. At the beginning of a period, each buyer is randomly assigned a service problem: either Problem X or Problem Y. The type of problem is determined randomly and is unaffected by other buyers' problems. Problem X happens with a 25% chance. Problem Y happens with a 75% chance. Buyers do not know which type of service problem they have. To fix the unknown problem, the buyers have to interact with the sellers.

Unlike buyers, sellers can identify the type of problem each buyer has. Sellers can solve a buyer's problem by choosing one of the two possible actions, Action 1 or Action 2. Knowing that a buyer has Problem X, a seller MUST choose Action 1 to solve it. But if a buyer has Problem Y, a seller could choose to fix the problem using either Action 1 or Action 2.

Moreover, the seller makes a total of four decisions, one towards each buyer. The following screenshot shows the decision interface for a particular seller. Note, since B2 has Problem X, this seller has no choice but choosing Action 1. For the other three buyers who have Problem Y, this seller could choose either Action 1 or Action 2.

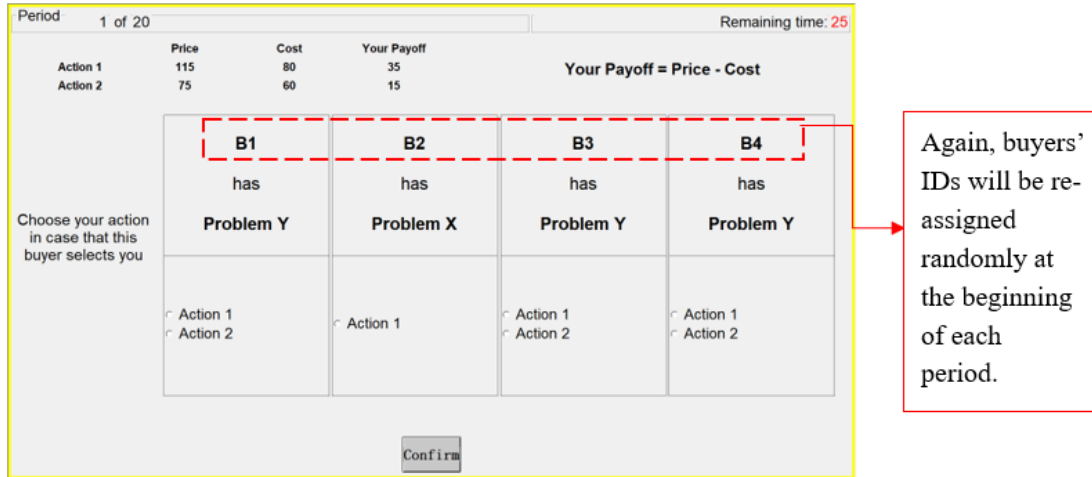


Figure B.1: Example seller's decision screen

Note that although a seller is choosing an action for each buyer, it does not necessarily mean that the seller will sell the service to all buyers. You can consider these seller's choices as "recommendations" to the buyers. Only in the case that a buyer chooses to interact with a particular seller and accept the proposed action, that seller's choice would be used to calculate his actual payoffs. For example, if a seller chooses Action 1 for all four buyers, but only B1 and B3 accept the actions provided by this particular seller, the seller's choice for B2 and B4 will not be used for payoff calculation.

In this experiment, while sellers make their “recommendations”, buyers need to select one of the four sellers in the market to interact with. Buyers make this selection without knowing what action each seller had chosen for him/her. The following screenshot shows the decision interface for a particular buyer.

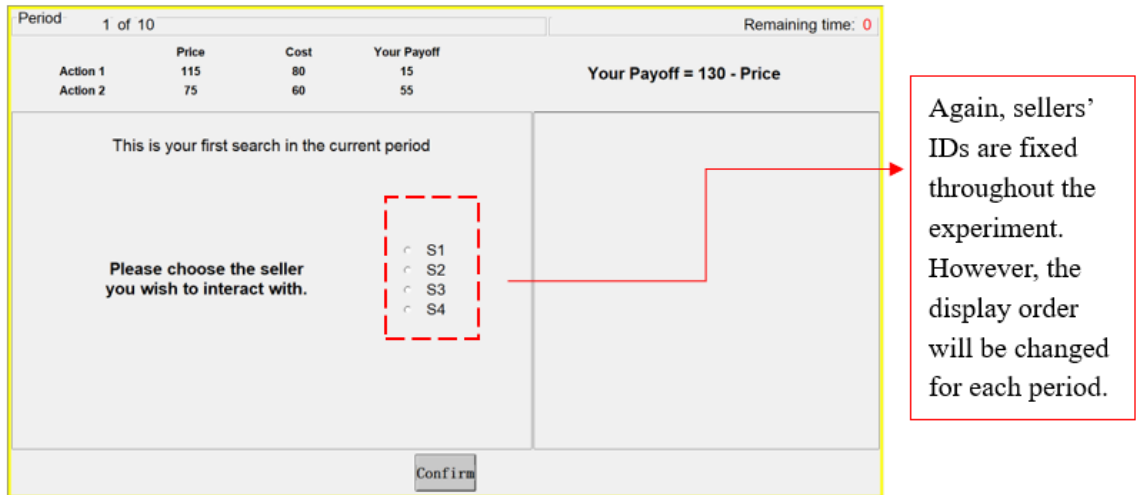


Figure B.2: Example buyer's first decision screen

After the buyer selects a seller, s/he will observe the action offered by the selected seller. Now, the buyer has two options: (1) accept the action or (2) pay a cost of 7 points to search for another seller from the remaining three sellers. If the buyer chooses to accept the action, s/he pays the seller the price for the action and the period ends. If the buyer chooses to search for another seller, s/he pays a cost of 7 points and receives the action offered by the new seller. The action offered by this second seller is final. The buyer must accept it. The buyer does NOT have the option to select again, nor can s/he go back to accept the offer from the first seller.

The following screenshot shows an example of the decision interface for a particular buyer when s/he has selected a seller (in this case S1) and is asked to either accept the action (in this case, Action 1) or to search for another seller.

Period	1 of 20			Remaining time: 9
	Price	Cost	Your Payoff	
Action 1	115	80	15	Your Payoff = 130 - Price
Action 2	75	60	55	

You selected **S1** and he/she chose to perform **Action 1**.

Please decide whether to search for another seller.
 Notice: searching for another seller entails an additional cost of **7 points**.

Accept the current action.
 Search for another seller.

Figure B.3: Example buyer's second decision screen

If the buyer selects “Search for another seller” in the screen shown above, s/he will see the interface below to select another seller from the remaining three sellers. Then the buyer receives the action offered by the new seller and pays the corresponding price. Notice that S1 is not on the option list. That is because this buyer chose not to accept the action offered by S1 in the previous decision screen.

Period	1 of 10			Remaining time: 28
	Price	Cost	Your Payoff	
Action 1	115	80	8	Your Payoff = 130 - 7 - Price
Action 2	75	60	48	
This is your second search in the current period				
Please choose the seller you wish to interact with.				
<input type="radio"/> S2 <input type="radio"/> S3 <input type="radio"/> S4				
<input type="button" value="Confirm"/>				

Figure B.4: Example buyer’s third decision screen

Starting from the second period, each buyer can see a history table on the right half of the screen when s/he makes decisions. The table displays interactions from all previous periods, including information on the selected seller's ID and the action offered by that seller in both the first and the second searches. If a buyer did not go through a second search, the related field will show a pound sign “#”.

The following screenshot shows an example of the interface for a particular buyer during Period 4. In this example, the first row of the table shows you that this particular buyer knows that in Period 1, s/he selected S1 during the first search and S1 chose Action 2. This particular buyer accepted S1's action; therefore, there is no information provided for the second search.

Period 4 of 20 Remaining time: 15

	Price	Cost	Your Payoff
Action 1	115	80	15
Action 2	75	60	55

Your Payoff = 130 - Price

This is your first search in the current period

Please choose one of sellers that you would like to interact with.

- S1
- S2
- S3
- S4

Trade History				
Period	First Search	First Seller Action	Second Search	Second Seller Action
1	S1	2	#	#
2	S4	1	S3	1
3	S1	1	#	#

Confirm

Figure B.5: Example buyer's first decision screen in Period 4 in PRI

[(Both AGG_P and AGG_P&PH: Starting from the second period, buyers can see a history table on the right half of the screen when they make decisions. The table displays the percentage of Action 1 recommended by each seller in all previous periods.]

The following screenshot shows an example of the interface for a buyer during Period 4. In this example, the first row of the table shows you that all the buyers know that in Period 1, the percentages of Action 1 recommended by S1, S2, S3, and S4 were 100%, 100%, 50%, and 25%, respectively. These percentages mean that in Period 1, S1 and S2 recommended Action 1 to four out of four buyers (100% of the time), S3 recommended Action 1 to two out of four buyers (50% of the time), and S4 recommended Action 1 to one out of four buyers (25% of the time). Additionally, the bottom row of the table displays the average percentage of Action 1 recommended in all past periods for each seller.)]

Period 4 of 20				Remaining time: 0				
Action 1	Price	Cost	Your Payoff	Your Payoff = 130 - Price				
Action 2	115	80	15					
	75	60	55					
<p>This is your first search in the current period</p> <p>Please choose the seller you wish to interact with.</p> <ul style="list-style-type: none"> <input type="radio"/> S3 <input type="radio"/> S4 <input type="radio"/> S1 <input type="radio"/> S2 				Percentage of Action 1 Recommended				
				Period	S1(%)	S2(%)	S3(%)	S4(%)
				1	100	100	50	25
				2	25	75	50	75
3	75	75	50	100				
Average		66.7	83.3	50.0	66.7			
Confirm								

Figure B.6: Example buyer's first decision screen in Period 4 in AGG_P and AGG_P&PH

[Exclusively available in AGG_P&PH: Starting from the second period, sellers can view the same information as buyers on the right half of the screen. Each seller's own ID number is highlighted in red with a red arrow at the bottom of the history table. The following screenshot provides an example of the decision interface for S1. The second column of the history table records the percentages of Action 1 recommended by S1 himself/herself in all previous periods, indicated by the red arrow.]

Period 4 of 20 Remaining time: 8

	Price	Cost	Your Payoff
Action 1	115	80	35
Action 2	75	60	15

Choose your action in case that this buyer selects you

B1	B2	B3	B4
has	has	has	has
Problem Y	Problem Y	Problem X	Problem Y
<input type="radio"/> Action 1 <input type="radio"/> Action 2	<input type="radio"/> Action 1 <input type="radio"/> Action 2	<input type="radio"/> Action 1	<input type="radio"/> Action 1 <input type="radio"/> Action 2

Confirm

Your Payoff = Price - Cost

Percentage of Action 1 Recommended				
Period	S1(%)	S2(%)	S3(%)	S4(%)
1	100	100	50	25
2	25	75	50	75
3	75	75	50	100
Average	66.7	83.3	50.0	66.7

Figure B.7: Example seller's decision screen in Period 4 in AGG_P&PH

In summary, in each period, sellers know the type of problem each buyer has and recommend an action to solve each problem. Simultaneously, each buyer chooses a seller without knowing what action that seller has chosen. The buyers then observe the chosen seller's action and decide on whether to accept the action or to search for a second seller. Note: in case the buyer did not accept the first seller, that buyer must accept the action from the seller chosen in the second search. There is no more search after that. Also, the buyer cannot go back to the first selected seller once the second search begins.

Payoffs

You may wonder how Action 1 and Action 2 impact your payoffs. Here are the details.

For Action 1, a seller charges a buyer the price of 115 points and pays the cost of 80 points.

For Action 2, a seller charges a buyer the price of 75 points and pays the cost of 60 points.

Seller's earnings for a particular period are the sum of payoffs from all buyers who choose to accept that seller's action. The acceptance includes offers accepted during both buyers' first and second searches. For each accepted offer,

$$Payoff_{\text{seller}} = Price - Cost$$

If Action 1 is offered and accepted, the seller's payoff is 35 points (=115-80).

If Action 2 is offered and accepted, the seller's payoff is 15 points (=75-60).

If none of the seller's offers is accepted, the seller receives 0 points for the period.

A seller's total earnings are the sum of earnings across all 20 periods.

On the other hand, buyer payoff is calculated as follows:

$$Payoff_{\text{buyer}} = \begin{cases} 130 - Price_{\text{first}} & \text{if the buyer accepts the first seller's action} \\ 130 - Price_{\text{second}} - 7 & \text{if the buyer searches for a second seller} \end{cases}$$

If Action 1 is offered and accepted in the first search, the buyer payoff is 15 points (= 130 - 115).

If Action 2 is offered and accepted in the first search, the buyer payoff is 55 points (= 130 - 75).

If either of these actions is accepted in the second search, the buyer payoff will be 7 points lower. That is, 8 points for Action 1 and 48 points for Action 2.

A buyer's total earnings are the sum of payoffs across all 20 periods.

Notice that a buyer cannot find out what type of problem s/he has even after the payoff is determined and revealed (the buyer observes their own payoff, not the seller's payoff).

The information about the prices and costs of two different actions and your payoff will be listed on the top of the interface. These parameters are fixed throughout the experiment.

B.2 Control questions (was the same in all three experimental conditions)

Please answer the following questions:

Q1: How many periods in total are there in this experiment?	Answer: 20
<p>Consider the following hypothetical scenario:</p> <p>Suppose Buyer B1 selected Seller S2 as his first-choice seller and observed that S2 chose to perform Action 1. Then, B1 decided to switch to S3 in the second search and observed that S3 chose to perform Action 1 as well.</p>	
Q2: How many points did S2 earn in this scenario?	Answer: 0
Q3: How many points did S3 earn in this scenario? Hint: The price of Action1 – The cost of Action 1.	Answer: 35
Q4: How many points did B1 earn in this scenario? Hint: 130 - The price of Action 1 - 7 (search cost).	Answer: 8
Q5: Will B1 be able to go back to S2 and accept the offer from S2 if s/he chooses to search for a second seller? Yes/No	Answer: No

B.3 Questionnaires

1. Please enter your UMass e-mail address where we can send you the Amazon eGift card.

2. Which year were you born?

3. What gender are you identifying with?

Male

Female

Other

4. Which academic cohort do you belong to as of this Fall semester?

Freshman

Sophomore

Junior

Senior

Graduate Student

Non-degree Seeker

5. Which school are you majored in?

College of Education

College of Engineering

- College of Humanities and Fine Arts
- College of Information and Computer Sciences
- College of Natural Sciences
- College of Nursing
- College of Social and Behavioral Sciences
- Isenberg School of Management
- School of Public Health and Health Sciences
- Stockbridge School of Agriculture
- Commonwealth Honors College
- Undetermined

6. What is your current GPA?

- 3.5 – 4.0
- 3.0 – 3.49
- 2.0 – 2.99
- Below 2.0

7. Have you ever taken any classes in economics?

- Yes
- No

BIBLIOGRAPHY

- Agarwal, Ritu, Che-Wei Liu, and Kislaya Prasad.** 2019. “Personal research, second opinions, and the diagnostic effort of experts.” *Journal of Economic Behavior & Organization*, 158: 44–61.
- Angerer, Silvia, Daniela Glätzle-Rützler, and Christian Waibel.** 2021. “Monitoring institutions in healthcare markets: Experimental evidence.” *Health Economics*, 30(5): 951–971.
- Angerer, Silvia, Daniela Glätzle-Rützler, Thomas Rittmannsberger, and Christian Waibel.** 2021. “The value of rating systems in healthcare credence goods markets.” *Available at SSRN 3965318*.
- Balafoutas, Loukas, Adrian Beck, Rudolf Kerschbamer, and Matthias Sutter.** 2013. “What drives taxi drivers? A field experiment on fraud in a market for credence goods.” *Review of Economic Studies*, 80(3): 876–891.
- Balafoutas, Loukas, and Rudolf Kerschbamer.** 2020. “Credence goods in the literature: What the past fifteen years have taught us about fraud, incentives, and the role of institutions.” *Journal of Behavioral and Experimental Finance*, 26: 100285.
- Bar-Isaac, Heski, Steven Tadelis, et al.** 2008. “Seller reputation.” *Foundations and Trends® in Microeconomics*, 4(4): 273–351.
- Bask, Mikael, Jens Lundgren, and Niklas Rudholm.** 2011. “Market power in the expanding Nordic power market.” *Applied Economics*, 43(9): 1035–1043.
- Borenstein, Severin, and James Bushnell.** 2000. “Electricity restructuring: deregulation or reregulation.” *Regulation*, 23: 46.
- Borenstein, Severin, James B Bushnell, and Frank A Wolak.** 2002. “Measuring market inefficiencies in California’s restructured wholesale electricity market.” *American Economic Review*, 92(5): 1376–1405.
- Borenstein, Severin, James Bushnell, and Frank Wolak.** 1999. “Diagnosing market power in California’s deregulated wholesale electricity market.”
- Borenstein, Severin, James Bushnell, and Steven Stoft.** 2000. “The Competitive Effects of Transmission Capacity in a Deregulated Electricity Industry.” *The RAND Journal of Economics*, 294–325.

- Brandts, Jordi, and Jan Potters.** 2018. “Experimental industrial organization.” In *Handbook of Game Theory and Industrial Organization, Volume II*. 453–474. Edward Elgar Publishing.
- Brosig-Koch, Jeannette, Burkhard Hehenkamp, and Johanna Kokot.** 2017. “The effects of competition on medical service provision.” *Health economics*, 26: 6–20.
- Brosig-Koch, Jeannette, Heike Hennig-Schmidt, Nadja Kairies-Schwarz, and Daniel Wiesen.** 2016. “Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision.” *Journal of Economic Behavior & Organization*, 131: 17–23.
- Brownlee, Shannon.** 2010. *Overtreated: Why too much medicine is making us sicker and poorer*. Bloomsbury Publishing USA.
- Bushnell, James B, Erin T Mansur, and Celeste Saravia.** 2008. “Vertical arrangements, market structure, and competition: An analysis of restructured US electricity markets.” *American Economic Review*, 98(1): 237–66.
- Bushnell, Jim, and Celeste Saravia.** 2002. “An empirical assessment of the competitiveness of the New England electricity market.”
- Butters, R Andrew, Jackson Dorsey, and Gautam Gowrisankaran.** 2021. “Soaking up the sun: Battery investment, renewable energy, and market equilibrium.” National Bureau of Economic Research.
- Chartock, Benjamin L.** 2021. “Quality Disclosure, Demand, and Congestion: Evidence from Physician Ratings.”
- Craig, David.** 2005. *Rip off*. Original Book Company.
- Currie, Janet, Wanchuan Lin, and Juanjuan Meng.** 2014. “Addressing antibiotic abuse in China: an experimental audit study.” *Journal of development economics*, 110: 39–51.
- Darby, Michael R, and Edi Karni.** 1973. “Free competition and the optimal amount of fraud.” *The Journal of law and economics*, 16(1): 67–88.
- Davis, Lucas, and Catherine Hausman.** 2016. “Market impacts of a nuclear power plant closure.” *American Economic Journal: Applied Economics*, 8(2): 92–122.
- Delattre, Eric, and Brigitte Dormont.** 2003. “Fixed fees and physician-induced demand: A panel data study on French physicians.” *Health economics*, 12(9): 741–754.
- Dellarocas, Chrysanthos.** 2006. “Reputation mechanisms.” *Handbook on economics and information systems*, 629–660.

- Domenighetti, Gianfranco, Antoine Casabianca, Felix Gutzwiller, and Sebastiano Martinoli.** 1993. "Revisiting the most informed consumer of surgical services: the physician-patient." *International journal of technology assessment in health care*, 9(4): 505–513.
- Duch, Matthias L, Max RP Grossmann, and Thomas Lauer.** 2020. "z-Tree unleashed: A novel client-integrating architecture for conducting z-Tree experiments over the Internet." *Journal of Behavioral and Experimental Finance*, 28: 100400.
- Dulleck, Uwe, and Rudolf Kerschbamer.** 2006. "On doctors, mechanics, and computer specialists: The economics of credence goods." *Journal of Economic literature*, 44(1): 5–42.
- Dulleck, Uwe, Rudolf Kerschbamer, and Matthias Sutter.** 2011. "The economics of credence goods: An experiment on the role of liability, verifiability, reputation, and competition." *American Economic Review*, 101(2): 526–555.
- Fischbacher, Urs.** 2007. "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental economics*, 10(2): 171–178.
- Fong, Yuk-fai, Ting Liu, and Xiaoxuan Meng.** 2022. "Trust building in credence goods markets." *American Economic Journal: Microeconomics*, 14(1): 490–528.
- Frankel, Alexander, and Michael Schwarz.** 2014. "Experts and their records." *Economic Inquiry*, 52(1): 56–71.
- Ge, Ming.** 2023. "It Takes Two Hands to Clap: Effects of Reputation and Search in Healthcare Markets."
- Gerlach, Heiko, and Junqian Li.** 2022. "Experts, trust and competition." *Journal of Economic Behavior & Organization*, 198: 552–578.
- Gottschalk, Felix, Wanda Mimra, and Christian Waibel.** 2020. "Health services as credence goods: A field experiment." *The Economic Journal*, 130(629): 1346–1383.
- Graf, Christoph, Emilio La Pera, Federico Quaglia, and Frank A Wolak.** 2021. "Market Power Mitigation Mechanisms for Wholesale Electricity Markets: Status Quo and Challenges." *Work. Pap. Stanf. Univ.*
- Green, Ellen P.** 2014. "Payment systems in the healthcare industry: An experimental study of physician incentives." *Journal of economic behavior & organization*, 106: 367–378.
- Green, Richard J, and David M Newbery.** 1992. "Competition in the British electricity spot market." *Journal of political economy*, 100(5): 929–953.
- Greiner, Ben.** 2015. "Subject pool recruitment procedures: organizing experiments with ORSEE." *Journal of the Economic Science Association*, 1(1): 114–125.

- Hadsell, Lester.** 2011. “Inefficiency in deregulated wholesale electricity markets: the case of the New England ISO.” *Applied Economics*, 43(5): 515–525.
- Hennig-Schmidt, Heike, Reinhard Selten, and Daniel Wiesen.** 2011. “How payment systems affect physicians’ provision behaviour—an experimental investigation.” *Journal of health economics*, 30(4): 637–646.
- Hortaçsu, Ali, and Steven L Puller.** 2008. “Understanding strategic bidding in multi-unit auctions: a case study of the Texas electricity spot market.” *The RAND Journal of Economics*, 39(1): 86–114.
- Huck, Steffen, Gabriele K Lünser, and Jean-Robert Tyran.** 2012. “Competition fosters trust.” *Games and Economic Behavior*, 76(1): 195–209.
- Huck, Steffen, Gabriele Lünser, Florian Spitzer, and Jean-Robert Tyran.** 2016. “Medical insurance and free choice of physician shape patient overtreatment: A laboratory experiment.” *Journal of Economic Behavior & Organization*, 131: 78–105.
- Hu, Wei.** 2017. “Design Health Insurance for Overtreatment.” *Available at SSRN 2970922*.
- Iizuka, Toshiaki.** 2007. “Experts’ agency problems: evidence from the prescription drug market in Japan.” *The Rand journal of economics*, 38(3): 844–862.
- ISO New England.** 2019. *2018 Annual Markets Report*. Holyoke, MA.
- Ito, Koichiro, and Mar Reguant.** 2016. “Sequential markets, market power, and arbitrage.” *American Economic Review*, 106(7): 1921–57.
- Johnson, Erin M, and M Marit Rehani.** 2016. “Physicians treating physicians: Information and incentives in childbirth.” *American Economic Journal: Economic Policy*, 8(1): 115–141.
- Kerschbamer, Rudolf, and Matthias Sutter.** 2017. “The economics of credence goods—a survey of recent lab and field experiments.” *CESifo Economic Studies*, 63(1): 1–23.
- Kim, Harim.** 2019. “Cleaner but volatile energy? The effect of coal plant retirement on market competition in the wholesale electricity market.”
- Kim, Harim.** 2022. “Heterogeneous impacts of cost shocks, strategic bidding, and pass-through: evidence from the New England electricity market.” *American Economic Journal: Microeconomics*, 14(2): 370–407.
- Kolstad, Jonathan T.** 2013. “Information and quality when motivation is intrinsic: Evidence from surgeon report cards.” *American Economic Review*, 103(7): 2875–2910.

- Lisa, Kate.** 2022. “Legislation seeks to address New York’s high cesarean birth rate.” <https://spectrumlocalnews.com/nys/central-ny/politics/2022/08/23/legislation-pushes-back-on-new-york-s-high-cesarean-birth-rate>.
- Liu, Meng, Erik Brynjolfsson, and Jason Dowlatabadi.** 2021. “Do digital platforms reduce moral hazard? The case of Uber and taxis.” *Management Science*, 67(8): 4665–4685.
- Luca, Michael, and Sonal Vats.** 2013. “Digitizing doctor demand: The impact of online reviews on doctor choice.” *Cambridge, MA: Harvard Business School*.
- Lyu, Heather, Tim Xu, Daniel Brotman, Brandan Mayer-Blackwell, Michol Cooper, Michael Daniel, Elizabeth C Wick, Vikas Saini, Shannon Brownlee, and Martin A Makary.** 2017. “Overtreatment in the united states.” *PloS one*, 12(9): e0181970.
- Mansur, Erin.** 2001. *Pricing behavior in the initial summer of the restructured PJM wholesale electricity market*. University of California Energy Institute Berkeley, CA.
- Mansur, Erin T.** 2008. “Measuring welfare in restructured electricity markets.” *The Review of Economics and Statistics*, 90(2): 369–386.
- McGinnis, J Michael, Leigh Stuckhardt, Robert Saunders, Mark Smith, et al.** 2013. “Best care at lower cost: the path to continuously learning health care in America.”
- McRae, Shaun D, and Frank A Wolak.** 2009. “How do firms exercise unilateral market power? Evidence from a bid-based wholesale electricity market.”
- Mimra, Wanda, Alexander Rasch, and Christian Waibel.** 2016a. “Price competition and reputation in credence goods markets: Experimental evidence.” *Games and Economic Behavior*, 100: 337–352.
- Mimra, Wanda, Alexander Rasch, and Christian Waibel.** 2016b. “Second opinions in markets for expert services: Experimental evidence.” *Journal of Economic Behavior & Organization*, 131: 106–125.
- Pieper, Dawid, Simone Hess, and Tim Mathes.** 2017. “Compilation of second opinion programs in the German statutory health insurance.” *Gesundheitswesen (Bundesverband der Ärzte des Öffentlichen Gesundheitsdienstes (Germany))*, 80(10): 859–863.
- Puller, Steven L.** 2007. “Pricing and firm conduct in California’s deregulated electricity market.” *The Review of Economics and Statistics*, 89(1): 75–87.
- Reguant, Mar.** 2014. “Complementary bidding mechanisms and startup costs in electricity markets.” *The Review of Economic Studies*, 81(4): 1708–1742.

- Ryan, Nicholas.** 2021. “The competitive effects of transmission infrastructure in the indian electricity market.” *American Economic Journal: Microeconomics*, 13(2): 202–42.
- Schneider, Henry S.** 2012. “Agency problems and reputation in expert services: Evidence from auto repair.” *The Journal of Industrial Economics*, 60(3): 406–433.
- Sülzle, Kai, and Achim Wambach.** 2005. “Insurance in a market for credence goods.” *Journal of Risk and Insurance*, 72(1): 159–176.
- Tadelis, Steven.** 2016. “Reputation and feedback systems in online platform markets.” *Annual Review of Economics*, 8: 321–340.
- Wagner, Todd H, and Lisa Smith Wagner.** 1999. “Who Gets Second Opinions? Once a technique for controlling inappropriate care, second medical opinions have become a controversial practice under managed care.” *Health Affairs*, 18(5): 137–145.
- Woerman, Matt.** 2019. “Market size and market power: Evidence from the Texas electricity market.” *Energy Institute Working Paper*, 298.
- Wolak, Frank A.** 2003. “Measuring unilateral market power in wholesale electricity markets: the California market, 1998-2000.” *American Economic Review*, 93(2): 425–430.
- Wolak, Frank A, and Robert H Patrick.** 2001. “The impact of market rules and market structure on the price determination process in the England and Wales electricity market.”
- Wolak, Frank, Severin Borenstein, and James B Bushnell.** 2002. “Measuring Market Inefficiencies in California’s Restructured Wholesale Electricity Market.” *American Economic Review*, 92(5): 1376–1405.
- Wolfram, Catherine D.** 1999. “Measuring duopoly power in the British electricity spot market.” *American Economic Review*, 89(4): 805–826.
- Wolinsky, Asher.** 1993. “Competition in a market for informed experts’ services.” *The RAND Journal of Economics*, 380–398.