October 2017

Three Essays on Incentive Problems of Parties with Potential Conflict of Interest

Salar Ghamat
The University of Western Ontario

Supervisor
Dr. Gregory S. Zaric
The University of Western Ontario

Joint Supervisor
Dr. Hubert Pun
The University of Western Ontario

Graduate Program in Business

A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

© Salar Ghamat 2017

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Management Sciences and Quantitative Methods Commons, and the Operations and Supply Chain Management Commons

Recommended Citation
https://ir.lib.uwo.ca/etd/4909

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact tadam@uwo.ca, wlswdmin@uwo.ca.
I study the impact of different incentives on strategic decisions of parties that have the option of cooperating with each other. Incentive problems are well studied in various contexts, such as supply chain management and healthcare operations management. However, in the fast-changing business environment, there is a need to study and understand the new and emerging strategic behaviors of firms to adopt better incentive mechanisms and reach desired outcomes. This dissertation consists of three essays that examine the strategic behavior of parties under different incentive schemes.

In the first essay, I study the supply chain partnership of two potential competitors and evaluate the impact of limited capacity on their strategic behavior. An increasing number of original brand manufacturers (OBMs) do not have in-house production capability, and thus rely on competitive contract manufacturers (CCMs) on the supply side. This increasing demand puts CCMs in a capacity allocation dilemma between their own product and the OBM’s product. I derive the conditions that incentivize the two potential competitors to cooperate and compete (coopetition), compete, or only cooperate (supply chain partnership). I show that the OBM might multi-source its component demand only when competition in the final-product market is intense. Moreover, the CCM can be worse off from having more capacity, even when that CCM’s capacity is available for free.

The second and third essays are inspired by changes in healthcare funding models that reward quality care. In the second essay, I examine performance-based payment contracts to promote the optimal use of an optional diagnostic test that will help determine if a cancer patient would benefit from an advanced treatment option. This essay is inspired by three ongoing trends: tremendous increases in the cost of new advanced cancer drugs, development of new diagnostic tests to allow physicians to tailor treatments to patients, and changes in healthcare funding models that reward quality care. I model the interaction between two parties—a healthcare payer and an oncologist—where the oncologist has private information about the patient’s characteristics (adverse selection) and the payer does not know whether the optimal course of action is used by the oncologist (moral hazard). I demonstrate that, in the presence of information asymmetry, an oncologist should never test all patients, even when the diagnostic test is available for free. I also show that it is not
always socially optimal to make a diagnostic test compulsory, even if such a policy can be implemented for free.

In the third essay, I study gain-sharing agreements between a hospital and a healthcare provider that can only treat a patient and achieve the desired quality of care with collective effort. The Centers for Medicare and Medicaid Services (CMS) introduced a bundled payment model for lower extremity joint replacement (LEJR) that offers hospitals a fixed bundled payment for a patient’s treatment expenses during acute and post-acute care. This bundled payment model aims to incentivize hospitals to enter into agreements with providers to ensure that the total treatment cost and care quality meets the bundled payment requirements. However, I show that the bundled payment does not always incentivize a hospital to offer gain-sharing agreements to the provider. Furthermore, I show that the provider prefers a low bundled payment, such that the hospital needs the provider to reduce the total cost of treatment.

Keywords

Co-Authorship Statement

I hereby declare that this thesis incorporates some material that is a result of joint research. Essay 1 was co-authored with Dr. Hubert Pun and Dr. Xinghao (Shaun) Yan. Essay 2 was co-authored with Dr. Greg Zaric and Dr. Hubert Pun. As the first author, I was in charge of all aspects of the projects including formulating research questions, literature review, research design, model formulation and analysis, and preparing the first and the following complete of the manuscripts. With the above exceptions, I certify that this dissertation and the research to which it refers, is fully a product of my own work. Overall, this dissertation includes 3 original papers, with the first two essays are being currently under review in academic journals.

Ghamat, S., Pun, H., Yan, X. H. (July, 2017). Optimal Outsourcing Strategies when Capacity is Limited. Revise and Resubmit for 4\textsuperscript{th} round review (minor comments), Decision Sciences.

# Table of Contents

Abstract ............................................................................................................................... ii  
Co-Authorship Statement................................................................................................... iv  
Table of Contents ................................................................................................................ v  
List of Tables ................................................................................................................... viii  
List of Figures .................................................................................................................... ix  
List of Appendices ............................................................................................................. xi  
Chapter 1 ............................................................................................................................. 1  
  1 Introduction .................................................................................................................... 1  
    1.1 Overview of Thesis and Specific Essays ................................................................. 2  
    1.2 References ............................................................................................................. 5  
Chapter 2 ............................................................................................................................. 7  
  2 Optimal Outsourcing Strategies when Capacity is Limited ........................................... 7  
    2.1 Introduction ............................................................................................................. 7  
    2.2 Literature Review .................................................................................................. 10  
    2.3 The Model ............................................................................................................. 12  
    2.4 Equilibrium Solution ............................................................................................. 16  
    2.5 Value of Competition in Component Market ....................................................... 25  
    2.6 Buy-and-hold Option ............................................................................................ 28  
    2.7 Conclusion ............................................................................................................ 29  
    2.8 References ............................................................................................................. 31  
Chapter 3 ........................................................................................................................... 34  
  3 Contracts to Promote Optimal Use of Optional Diagnostic Tests in Cancer Treatment  ................................................................................................................................. 34  
    3.1 Introduction ............................................................................................................. 34  
    3.2 Literature Review .................................................................................................. 39
# Table of Contents

3.3 The Model ............................................................................................................. 42  
  3.3.1 Provider’s problem .................................................................................... 45  
  3.3.2 Payer’s problem ........................................................................................ 47  
3.4 System Optimum .................................................................................................. 50  
  3.4.1 Full information ........................................................................................ 57  
  3.4.2 Alternate payment models ........................................................................ 58  
3.5 Comparative Statistics .......................................................................................... 59  
3.6 Compulsory Testing ............................................................................................. 61  
3.7 Conclusions ........................................................................................................... 64  
3.8 References ............................................................................................................. 65  

Chapter 4 ........................................................................................................................... 72  
4 Care-Coordination: Gain-sharing Agreements in Bundled Payment Models .......... 72  
  4.1 Introduction ....................................................................................................... 72  
  4.2 Literature Review .............................................................................................. 75  
  4.3 Fee-for-Service (Strategy S) ............................................................................. 78  
  4.4 Bundled Payment: FFS with Target Price ....................................................... 79  
    4.4.1 Centralized system (Strategy C) ............................................................... 81  
    4.4.2 Decentralized system (Strategy D) ........................................................... 85  
  4.5 FFS with Target Price and Gain-sharing Agreements (Strategy G) ............... 87  
  4.6 Implications for the Payer (Numerical Analysis) .............................................. 92  
  4.7 Extensions ........................................................................................................... 95  
    4.7.1 Separate cost and quality .......................................................................... 95  
    4.7.2 Interdependency in cost functions ............................................................ 97  
  4.8 Conclusion ......................................................................................................... 98  
  4.9 References ......................................................................................................... 99  

Chapter 5 ......................................................................................................................... 102
List of Tables

Table 3.1: Parameter estimations for numerical examples ..................................................... 61

Table A.1: Equilibrium profits, prices, and demands of the firms when $w_T \leq \hat{w}_T$ .............. 118

Table A.2: Equilibrium profits, prices, and demands of the firms when there is no competition in the component market........................................................................................................ 119

Table A.3: Optimization problem for 9 subgames................................................................ 128
List of Figures

Figure 2.1: Relationship between customer’s utilities and demands ........................................ 14
Figure 2.2: Optimal outsourcing strategies when $M > M$ ...................................................... 18
Figure 2.3: Firms’ demand and price as a function of $k$ .......................................................... 21
Figure 2.4: Firms’ profit as a function of $k$ ............................................................................ 22
Figure 2.5: Optimal outsourcing strategies when $M \leq M$ ...................................................... 24
Figure 3.1: Sequence of events ............................................................................................... 44
Figure 3.2: Provider’s treatment choices ................................................................................ 46
Figure 3.3: Optimal treatment decisions for the system with information asymmetry .......... 52
Figure 3.4: Optimal treatment decisions for the system with full information....................... 58
Figure 3.5: Social welfare as a function of $\alpha$ multiplier in Case 4 ....................................... 60
Figure 3.6: Compulsory test versus Optional test when $CV = 0$ ............................................. 63
Figure 4.1: Payment scheme depending on performance ....................................................... 80
Figure 4.2: Optimal solution in the centralized system when $c_H = c_p = 2; b_p = 2; \alpha = 5\%$ .............................................................................................................................. 83
Figure 4.3: Profit and savings of the centralized system when $c_H = c_p = 2; b_p = 2; \alpha = 5\%$ .............................................................................................................................. 84
Figure 4.4: Optimal solution in the decentralized system when $c_H = c_p = 2; \alpha = 5\%$ .... 86
Figure 4.5: Profits and savings of the decentralized system when $c_H = c_p = 2; \alpha = 5\%$ .... 87
Figure 4.6: Sequence of events ............................................................................................... 88
Figure 4.7: Optimal solution in the decentralized system with gain-sharing agreement when $c_H = c_P = 2; b_P = 2; \alpha = 5\%$ ................................................................. 89

Figure 4.8: Profits and savings of the decentralized system with gain-sharing agreement when $c_H = c_P = 2; b_P = 2; \alpha = 5\%$ ................................................................. 91

Figure 4.9: Total spending when $c_H = c_P = 2; b_P = 2; \alpha = 5\%$ ................................................................. 92

Figure 4.10: Ratio of $\frac{\alpha^G}{\alpha^S}$ for different values of $c_t, b_p$, and $\alpha$ ................................................................. 94

Figure 4.11: Ratio of $\frac{\alpha^G}{\alpha^S \beta_H(e_H^*) \beta_P(e_P^*)}$ for different values of $c_t, b_p$, and $\alpha$ ................................................................. 95

Figure 4.12: Total spending when $c_H = c_P = 2; b_P = 2; \alpha = 5\%$ ................................................................. 96

Figure 4.13: Optimal solution in decentralized system with gain-sharing agreement when $c_H = c_P = 2; b_P = 2; \alpha = 5\%; \theta = 0.5$ ................................................................. 97

Figure 4.14: Ratio of $\frac{\alpha^G}{\alpha^S}$ for different values of $\theta$ ................................................................. 98

Figure A.1: Willingness-to-pay when $f_L(r) = 2 - 2r$ and $f_H(r) = 2r$ ................................................................. 124

Figure A.2: Willingness-to-pay when $f_L(r) = \begin{cases} 5r & \text{if } 0 \leq r < 0.4 \\ 2(1 - r)/0.6 & \text{if } 0.4 \leq r \leq 1 \end{cases}$ and $f_H(r) = \begin{cases} 2r/0.6 & \text{if } 0 \leq r < 0.6 \\ 2(1 - r)/0.4 & \text{if } 0.6 \leq r \leq 1 \end{cases}$ ................................................................. 124

Figure A.3: Willingness-to-pay when $f_L(r) = \begin{cases} 2r/0.49 & \text{if } 0 \leq r < 0.49 \\ 2(1 - r)/0.51 & \text{if } 0.49 \leq r \leq 1 \end{cases}$ and $f_H(r) = \begin{cases} 2r/0.51 & \text{if } 0 \leq r < 0.51 \\ 2(1 - r)/0.49 & \text{if } 0.51 \leq r \leq 1 \end{cases}$ ................................................................. 125
List of Appendices

Appendix-A ........................................................................................................................... 105
Appendix-B ........................................................................................................................... 120
Appendix-C ........................................................................................................................... 129
Chapter 1

1 Introduction

Recent technological advances have led to a substantial increase in the sophistication of products and services. Companies now have more customized and targeted products and services than ever before. As a result, it has become impossible for organizations to contain all procedures and production in-house. Instead, firms are becoming more specialized and focused, which has led to decentralized and fragmented supply chains. For example, outsourcing the production of components of a product is a very common practice among original brand manufacturers (OBMs) because doing so (instead of investing in production facilities) allows firms to invest in product development instead, which may provide greater competitive advantage. New Venture Research (NVR) estimates that the contract manufacturing industry will grow from $430 billion in 2015 to $580 billion in 2020 (NVR, 2016).

Similarly, in the healthcare industry, the growing number of treatment options and availability of personalized medicine have dramatically increased the cost of treatment and the number of care providers. For instance, the 21-gene assay for breast cancer (Oncotype Dx), introduced by Genomic Health Inc. in 2004, is a gene-expression profiling test that can help guide cancer treatment decisions. Likewise, there are many independent care providers that specialize in a specific procedure required in treating a patient’s condition. However, in a fragmented healthcare system enforcing standard protocols to ensure that physicians follow through with appropriate patient care is a barrier to adopting these new technologies (Davis et al., 2010). Therefore, as Kaplan and Porter (2012) describe, one of the challenges in terms of healthcare cost crises is the complex path of care due to the highly fragmented way in which healthcare is delivered.

It is well known that, decentralization creates inefficiency in supply chains. This inefficiency is caused by a lack of coordination among the different stages of a supply chain, and the lack of coordination among parties stems from two causes: (1) conflict between the respective interests and objectives of the parties, and (2) the distortion of
information as it moves between the stages of the supply chain. The purpose of this thesis is to examine different obstacles to coordination, and to understand firms’ strategic decisions under different incentive mechanisms. The results generated from this thesis add to the current knowledge on the effects of decentralization on supply chains. Moreover, I provide managerial insights for firms in both the manufacturing and healthcare industries on how to coordinate their supply chains using some incentive mechanisms. Supply chain partnership and coordination provide firms with higher revenues and opportunities to access more financial resources for future business investments. I illustrate how the new advancements in healthcare services require alternative payment mechanisms to achieve a higher quality of care and lower treatment costs. Finally, this thesis highlights the importance of customizing coordinating mechanisms to the specific system settings by showing the sensitivity of the coordination and strategic decisions of firms to these incentive mechanisms.

1.1 Overview of Thesis and Specific Essays

In Chapter 2, I study the supply chain partnership of two potential competitors—an OBM and a competitive contract manufacturer (CCM)—and evaluate the impact of limited capacity on their strategic behavior. It is common for firms in different stages of a decentralized supply chain to compete with each other in the final product market. There are many studies that examine the scenario of competition between supply chain partners (e.g., Venkatesh et al., 2006; Xu et al., 2010; Wang et al., 2013; Yang et al., 2014). These studies either analyze the partnership of competitors for the supply of a proprietary component, or do not incorporate the limited capacity on the supply side of the shared component.

In Chapter 2, I extend the literature to a more realistic setting, where there is competition in the component market and the CCM faces a resource allocation problem due to limited capacity when choosing a partnership strategy with its competitor. I derive the conditions that incentivize the two potential competitors to cooperate and compete (coopetition), compete, or only cooperate (supply chain partnership). In particular, I show how firms can optimally use supply chain partnership and resource allocation to overcome the operational and pricing obstacles in a decentralized supply chain. I show
that the OBM might multi-source its component demand only when competition in the final-product market is intense. I also find that when capacity increases, firms’ demand might decrease while retail price might increase. Moreover, even when capacity is available for free, the CCM can be worse off from having more capacity. Further, the results illustrate that the demand for the CCM’s product may increase when the competing products become more substitutable.

In Chapter 3, I study the incentive problem of a healthcare purchaser (payer) and a healthcare provider (provider) for using a diagnostic test; to do so, I investigate performance-based payment contracts to align the incentives of two parties to achieve optimal use of optional diagnostic tests. There are studies in healthcare operations management that analyze a provider’s choice regarding use of diagnostic tests in various healthcare settings (e.g., Yaesoubi and Roberts, 2011; Dai et al., 2015; Dai et al., 2016). I extend this literature by studying payment mechanisms that coordinates the incentives of a healthcare payer and a provider, in the presence of information asymmetry, while examining the impact of an optional diagnostic test on the provider’s decision-making process. I show that the advent of new diagnostic tests necessitates new reimbursement models because current payment models do not account for treatment choices of the provider, and thus cannot coordinate a decentralized healthcare setting. I also highlight the importance of information flow in the healthcare system by showing that when there is imperfect information flow between the payer and the provider, the payer should motivate the provider to only test certain types of patients, even if the diagnostic test can be available for free. Furthermore, I find that an increase in reputational concerns and altruistic behavior of a provider may increase social welfare. Finally, I show that it is not always socially optimal to make a diagnostic test compulsory, even if there is no cost to implement such a policy.

In Chapter 4, I study gain-sharing agreements between a hospital and a provider that can only treat a patient and achieve the desired quality of care collectively. There are some studies that evaluate gain-sharing agreements in healthcare operations management, especially in the pharmaceutical industry (Barros, 2011; Zhang et al., 2011; Mahjoub et al., 2013). These studies analyze the coordination of a pharmaceutical company with a
There are also papers that investigate bundled payment models in different healthcare settings (Andritsos and Tang, 2015; Adida et al., 2016; Guo et al., 2016). I extend the literature by studying the impact of gain-sharing agreements in a healthcare setting with a bundled payment that ties the profit of the hospital and the provider. This study is motivated by the Centers for Medicare and Medicaid Services (CMS) lower extremity joint replacement (LEJR) bundled payment. This bundled payment model aims to incentivize hospitals to enter into agreements with providers to coordinate care and to ensure that the total treatment cost and care quality meets the bundled payment requirements. I illustrate that a target price bundled payment model would be more effective in healthcare settings where there is more asymmetry between the hospital and the provider’s cost-reduction efficiency. I also show that while using a gain-sharing agreement might be a win-win-win scenario for the payer, the hospital, and the provider, good design of the target price bundled payment is essential to incentivize a hospital to offer gain-sharing agreements to the provider. In addition, the provider prefers a low bundled payment such that the hospital needs the provider to reduce the total cost of treatment.

In the final chapter, I present an overview of the main results and the managerial insights from the analysis of the fragmented supply chain settings discussed in the thesis. I further highlight the policy implications of the implementation of the results.
1.2 References


Chapter 2

2 Optimal Outsourcing Strategies when Capacity is Limited

Outsourcing the production of selected components to competitors is becoming more common among original brand manufacturers (OBM); however, OBMs’ increased attention to outsourcing and the growing demand in many markets can result in capacity allocation conflicts for the contract manufacturers. In this study, I consider a scenario in which the OBM decides whether to outsource to a third-party supplier or to a competitive contract manufacturer (CCM) who has the option of producing a competing product and also has limited capacity. The CCM first chooses the wholesale price and decides whether or not to sell a competing product to the customers. Next, the OBM decides the proportion of its component demand to outsource to the CCM, and then firms set the retail prices. I show that the OBM might multi-source its component demand only when competition in the final-product market is intense. I also find that when CCM’s capacity increases, demand may decrease while the retail price may increase. Moreover, the CCM can be worse off from having more capacity, even when CCM’s capacity is available for free. The results also show that demand may increase when competition in the final-product market becomes more intense. Finally, I find that the value of having a third-party supplier to produce the component decreases amid the intensity of competition in the final-product market.

2.1 Introduction

TPV Technology (TPV), the largest electronic manufacturer of computer monitors, sells monitors under its own brands – AOC and Envision – in the final-product market and acts as a supplier to Philips, which sells monitors under the Philips brand. This arrangement means that Philips competes with TPV’s AOC and Envision brands. The overall demand for monitors is beyond the capacity of TPV, and thus, TPV has decided to reduce the production of its own brands in order to satisfy the outsourcing orders it receives from Philips (Wang, 2008).
Outsourcing the production of certain components to competitive contract manufacturers (CCMs) like TPV is becoming more common among original brand manufacturers (OBMs); however, growing demand often results in capacity-allocation conflicts for these CCMs. For example, Apple outsources its NAND Flash memory requirement to Samsung (Kim, 2012), but as smartphones become more popular, Samsung finds it increasingly difficult to fulfill the demand. Such capacity-allocation conflict “would be bad for Apple if Samsung were forced to choose between Apple and itself in case of a supply shortage at its factories” (Forbes, 2013). In another example, Franz Inc. (Franz) is a contract manufacturer that produces home décor accessories (e.g., tableware, vases and jewellery) for OBMs like Enesco and Lenox. In 2002, Franz started to sell products under its own brand while continuing to supply for the OBMs. The company reached its capacity limit due to increasing orders from OBMs, and eventually, in 2005, Franz decided to prioritize the production of its own brand products ahead of others (Yan, 2013). Tesla Motors was the supplier of battery packs for Mercedes B-class Electric Drive. Despite strong demand for its own automobile, the Tesla Model S, Tesla Motors could not increase the production amount due to its limited supply of batteries (Herron, 2013).

These examples show that when the CCM has a limited capacity, the OBM can influence the CCM’s output to the final-product market by using a portion of the CCM’s capacity, thereby mitigating competition in the final-product market. On the other hand, precisely because of this reason, the CCM sets a higher wholesale price, and thus, the introduction of a capacity constraint adds some interesting trade-offs for the firms. As both cooperation between competitors and capacity shortages begin to occur with increasing frequency, studies of the interaction between these phenomena and the resulting impacts become more relevant.

In this study, I analyze an OBM’s outsourcing strategies when the CCM has limited capacity. In particular, the OBM does not produce a critical component of its product in-house (e.g., the computer monitor in TPV’s example, the NAND Flash memory in Samsung’s example, the home accessories in Franz’s example and the battery pack in Tesla’s example) and must therefore decide the proportion of the production of
the component to be outsourced to the CCM while the remaining production can be outsourced to a third-party supplier. The CCM has limited capacity in the production of the critical component, so it must decide whether or not to sell products to customers under its own brand. Moreover, the CCM must decide the wholesale price of the component in order to compete with other third-party suppliers.

This study considers the following research questions:

1. What are the impacts of the CCM’s capacity on the outsourcing strategy and on the firms’ profitability?
2. How do the competition in the final-product market (between the OBM and the CCM) and the competition in the component market (between the CCM and the third-party supplier) affect the outsourcing strategy and the firms’ profitability?

In response to these questions, I present several interesting findings. First, despite the fact that CCM has the option of not selling final products to customers, but rather using its capacity to act as the OBM’s sole supplier, I find that the firms may forgo this opportunity. In particular, regardless of the capacity limit, the CCM always sells in the final-product market when the competition between the OBM and the CCM in low. However, when the competition in the final-product market is very intense, the CCM might hold a monopoly in the component market, while the OBM holds a monopoly in the final-product market. In order to be the monopoly in the final-product market, the OBM would buy all the CCM’s capacity for relatively high wholesale price and supply the rest of its component demand from the cheaper third-party suppliers. I also show that the OBM only multi-source its component when the CCM does not sell in the final-product market.

Second, when capacity increases, one might expect that demand would increase while retail price would decrease. Surprisingly, I find that this intuition does not hold true when there is a shift in outsourcing strategy. Moreover, even when capacity is available for free, the CCM can be worse off from having more capacity. The impact of “profit decreases in capacity” is always larger when both firms are co-opetitors than when they
are competitors. Third, my results illustrate that the demand for the CCM’s product may increase when the competing products become more substitutable. Furthermore, even though the CCM has less incentive to allocate its capacity to produce components for the OBM when the intensity of competition in the final-product market increases, the value of having a third-party supplier to produce the component decreases as competition intensifies.

This chapter is organized as follows. In the next section, I review the related literature. I then present the mathematical model and the analytical results in Sections 2.3 and 2.4. In Section 2.5, I analyze the value of competition in the component market through a comparison of the basic model using a benchmark model without the third-party supplier. In Section 2.6, I consider a scenario where OBM can order excess quantity (buy-and-hold) to impact CCM’s output to the final-product market. Finally, I provide concluding remarks and managerial insights. I present the details of the derivation of the equilibriums and the proofs of the results in Appendix-A.

2.2 Literature Review

My study relates to two areas of research. The first stream of literature examines the scenario of competition between supply-chain partners. Venkatesh et al. (2006) and Xu et al. (2010) study the optimal strategies of a manufacturer who owns the proprietary component brand. The manufacturer decides whether to use the components exclusively, whether to become a supplier of an OBM, or whether to become a hybrid of both. These two studies show that the proprietary component manufacturer should hold a monopoly in the final-product market only when the two products are almost perfect substitutes. In my study, the component is not of a proprietary nature, so the manufacturer cannot hold a monopoly in the final-product market.

Wang et al. (2013) examine the advantage of being the first mover when the component is not of a proprietary nature. These authors assume that the CCM’s wholesale price should be smaller than the third-party supplier’s wholesale price. However, I show that in a setting with capacity, the OBM would be willing to pay a higher wholesale price in order to reduce competition in the final-product market. Pun
studies how an OBM should outsource its non-proprietary components when firms can exert effort to improve their production process. This author finds that the OBM might be better off outsourcing both the production and the process-improvement effort to the competitor, even when the competitor has a higher cost. Pun (2015) considers the optimal degree of cooperation between two competing manufacturers when the components are not proprietary and finds that competitors can be worse off from more cooperation, even when these competitors have better production capabilities. The studies discussed in this stream of literature do not consider capacity constraints, so I extend this literature by considering suppliers that have limited capacity.

This study examines the ways that capacity constraint affects firms. Osborne and Pitchik (1986) characterize the Nash equilibria in a duopoly that has limited capacity, showing that limited capacity could be beneficial because capacity constraint can be used to mitigate competition. Gupta and Wang (2007) study the capacity-allocation problem of a contract manufacturer that can accept two types of orders: high-volume contractual orders and one-time transactional orders. They find that the threshold acceptance policy is optimal, and a contract manufacturer can be better off serving only transactional orders when capacity is tight. Ülkü et al. (2007) consider the capacity constraint of a contract manufacturer and consider how the risk should be distributed to different OBMs when the OBMs’ demand levels are uncertain. Ozkan and Wu (2009) consider make-to-stock and make-to-order mechanisms for a supplier, seeking the point at which the fixed-capacity allocation level between two different orders becomes optimal. Martínez-de-Albéniz and Talluri (2011) study dynamic price competition under uncertain demand. They provide a characterization of the equilibrium and show that firms may price their product at the reservation value of their competitor. The studies discussed in this paragraph assume that the firms are either supply-chain partners or competitors, and I extend this stream of literature by considering the case where competitors are also supply-chain partners.

Yang et al. (2014) consider the distribution strategies of a manufacturer of proprietary component brands with limited capacity when the OBM and its supplier are competitors in the final-product market. These authors find that the OBM may order
excess inventory, and all firms and their customers can be better off from a system with limited capacity. In my study, I consider the case where the component is of a commodity nature, such that there is more than one supplier for the component. Therefore, I consider competition in the component market, and this competition provides an upper limit to the wholesale price that the OBM is willing to accept.

To the best of my knowledge, I am the first to examine how capacity affects the supply-chain structure in an oligopolistic competition when one firm might outsource a commodity component to its competitor.

2.3 The Model

I consider a scenario where an OBM (firm O) must outsource the production of a critical component. The component can be a monitor for the TPV example, a NAND Flash memory for the Samsung example, or a battery pack in Tesla example. For simplicity, I follow the literature on the supply-chain relationship between competitors (e.g., Venkatesh et al., 2006; Xu et al., 2010; Wang et al., 2013; Pun, 2014; Pun, 2015) by assuming that the final product consists of this component only.

There are two potential suppliers of this component. The first is a CCM (firm C) that also has a competing product. The production cost is normalized to zero, and firm C sells components to firm O at a wholesale price \( w_c \). Moreover, firm C’s production capacity of the component is \( k \).\(^1\) Similar to Gupta and Wang (2007), I assume that firm C produces everything in-house and does not outsource to other third-party suppliers when facing capacity shortage.

The second component supplier is a third-party supplier (firm T) that does not have the option of producing a competitive product under its own brand. Firm T can also

\(^1\) Since there are many industry examples where a firm reveals its production capacity, I assume that firm C’s production capacity is public information. To illustrate, I note that Samsung and TPV Technologies present their capacity information on public websites (www.icinsights.com/news/bulletins/Samsung-TSMC-And-Micron-Top-List-Of-IC-Industry-Capacity-Leaders/ and www.tpv-tech.com/attachment/201504131218064_en.pdf).
be interpreted as the spot market, where the component can be purchased off the shelf.
Similar to other related literature on contract manufacturing (e.g., Jeannet, 2009; Wang et al., 2013), I assume that there are many identical and independent third-party suppliers competing to become firm O’s supplier, and firm T is one such supplier. This assumption is in line with many industry practices. As an example, even though Samsung holds a 30% market share in the NAND Flash memory market, there are many other non-competitive suppliers (e.g., Toshiba, SanDisk and Intel) from which Apple can outsource this critical component (DRAMeXchange, 2014). I incorporate the asymmetry between the component manufacturers by assuming firm C can be more cost-efficient than firm T (Arya et al., 2008; Pun, 2015). Thus, unlike firm C, the production cost of firm T is assumed to be non-negative. Due to the intense competition among these suppliers, firm T’s wholesale price $w_T \geq 0$ is exogenously determined as the equilibrium market price in a competitive market which is simply the production cost (i.e., perfect competition). Note that, when $w_T = 0$ the two component suppliers are symmetric. The reason that, unlike firm T, firm C’s wholesale price is assumed to be endogenous is because firm O might accept a higher wholesale price for the component of firm C (compared to the price offered by other non-competitive suppliers) only because buying from firm C can reduce firm C’s output to the final-product market. Moreover, since firm O can outsource to any of these non-competitive suppliers whenever one of them meets capacity, I do not consider a capacity limit for firm T.

Firm O decides the proportion of its component demand to be allocated to firm C ($\gamma \in [0,1]$), and the remaining component demand ($1 - \gamma$) will be allocated to firm T. There are three outsourcing strategies for firm O: (1) Firm O does not outsource to firm C (i.e., $\gamma = 0$), so firms O and C are pure competitor; (2) Firm O single-sources to firm C (i.e., $\gamma = 1$), and (3) firm O multi-sources to firms C and T (i.e., $0 < \gamma < 1$). The two firms are supply-chain partners and competitors when $\gamma > 0$ if firm C also sells products under its own brand.

The customers are assumed to have a unit reservation price for a product, and the products of firms O and C are located at an exogenously specified distance $M$ apart. When $M$ is small, the two products are more substitutable, and hence competition
becomes intense. The length of the Hotelling line is sufficiently larger than $M$ such that all customers located between the two firms would buy, but not all customers located outside the two firms would buy. Each customer incurs a disutility per unit distance and will buy whichever product provides the higher positive utility. Despite the fact that all the results can be driven for more general form, for expositional convenience, I assume that the disutility per unit distance is equal to 1. Therefore, the total demand depends on the distance $M$ between two products. Customers would have zero utility when not making a purchase. Thus, when buying a product from firm $i \in \{O, C\}$ at retail price $p_i$, a customer that is $d$ away from firm $i$ would have utility:

$$U = 1 - d - p_i$$  \hspace{1cm} (2.1)

**Figure 2.1: Relationship between customer’s utilities and demands**

From the customer’s utility function, it is apparent that firms’ demand is directly proportional to firms’ pricing strategy. The demand of firm $i$ ($D_i$) can be derived from the customer’s utility function by finding the indifferent customers’ location (Venkatesh et al., 2006; Xu et al., 2010). Figure 2.1 illustrates the relationship between the customers’ utilities and the demands; the triangles illustrate the customers’ utility from buying firms’ products (customers at zero distance will have a utility of $1 - p_i$). For each product, I find the location of the customer who is indifferent between buying a product or not (customer with zero utility located in the far-left or far-right of the Hotelling line). Then I find the location of the customer located between the two products who is indifferent between buying either product. The total demand of a product would be the sum of the
distances of the indifferent customers located outside and inside the two firms on the Hotelling line. Consequently, the profit function of firm O is as follows:

$$\pi_o = (p_o - w_C) \gamma D_o + (p_o - w_T) (1 - \gamma) D_o \quad (2.2)$$

The two parts of $\pi_o$ are the profits from selling products containing firm C’s component and from selling products containing firm T’s component. When firm C sells products under its own brand, the profit and the capacity constraint are

$$\pi_c = w_c \gamma D_o + p_c D_c \quad (2.3)$$

$$\gamma D_o + D_c \leq k \quad (2.4)$$

When firm C does not sell products under its own brand, the profit and the capacity constraint of firm C are

$$\pi_c = w_c \gamma D_o \quad (2.5)$$

$$\gamma D_o \leq k \quad (2.6)$$

As commonly used in the related literature (e.g., Cui et al., 2008; Wang et al., 2013) and consistent with many industry practices (e.g., Foxconn, Asustek), I assume that firm C first sets the wholesale price $w_c$, and then firm O decides the proportion of its component demand to be outsourced to each supplier, given the wholesale prices. Therefore, I consider two levels of competition: competition in the component market between firms C and T, and competition in the final-product market between firms O and C. The game sequence is as follows.

1) Firm C decides $w_c$ and chooses whether or not to sell its own product.

2) Firm O decides $\gamma$.

3) Firm O decides $p_o$. If applicable, firm C decides $p_c$.

I use backward induction to find the equilibrium solutions.
2.4 Equilibrium Solution

The competition in the final-product market, that is captured through product substitutability $M$, can affect firms’ strategic choices (equilibrium solution). I will show in Propositions 2.1 and 2.4 that firm O always single sources when the degree of competition is low ($M > \bar{M}$) and might multi-source when competition is intense ($M < \bar{M}$). This is because firm C cannot set an arbitrary high wholesale price due to the competition with firm T in the component market. Moreover, when the competition between firms O and C in the final-product market is small, firm O is not willing to pay a high wholesale price to buy-out firm C’s capacity.

For clarity of expositions, I first focus the analysis to the case when the product competition is low ($M > \bar{M}$). Specifically, I present the equilibrium strategy in Proposition 2.1. I then examine the impact of firm C’s capacity on firms’ price, demand and profit in Proposition 2.2, and the impact of the product substitutability on firms’ demand in Proposition 2.3. Lastly, I expand the analysis to the case when the final product competition is very intense ($M \leq \bar{M}$) in Proposition 2.4.

In order to derive the equilibrium solution, I use the Karush-Kuhn-Tucker conditions to consider firm C’s capacity constraint. I separate the optimization problem into two cases: 1) binding capacity equilibrium, where firm C uses all of its capacity; and 2) non-binding capacity equilibrium, where firm C has some unused capacity. Consequently, starting from the last stage of the game, I derive two different sets of optimal pricing strategies for the firms, depending on the capacity constraint (i.e., binding/non-binding). In the second stage of the game, firm O decides the proportion of its component demand to be allocated to firm C ($\gamma \in [0,1]$), given firms C and T’s wholesale price ($w_c$ and $w_T$, respectively), firm C’s capacity level $k$ and product substitutability $M$. Finally, in the first stage of the game, firm C chooses its wholesale price, anticipating the outsourcing strategy of firm O.

In the equilibrium solution, firm C maximizes its profit function by choosing the wholesale price $w_c$ for the two strategies where it acts as firm O’s supplier (supplier only and co-competition), subject to firm O’s incentive compatibility and participation constraints.
The incentive compatibility constraint ensures that firm O would not deviate to other strategies, given firm C’s wholesale price; the participation constraint makes sure that firm O gets at least as much as its outside option when it outsources to firm T. Finally, knowing the best outcome of each strategy, firm C chooses the equilibrium strategy with its wholesale price, depending on capacity $k$, firm T’s wholesale price $w_T$ and product substitutability $M$. Details of the derivation of the equilibrium are presented in Appendix-A.

First, I present the results for the case when the product competition is low, and then I show the effect of high product competition on the equilibrium solution.

**Lemma 2.1:** Define $\bar{M} = \frac{17}{48}\sqrt{48(1-w_T)^2 + 97k^2 - 2(1-w_T) - \frac{95k}{48}}$. When the competition in the final-product market is low (i.e., $M > \bar{M}$) firm C always sells in the final-product market.

Lemma 2.1 shows that firm C always sells in the final-product market when the competition in the final-product market is low. This is because when competition between the two products is intense, firm O would have more benefit if firm C does not sell product to the customers. However, when the degree of competition is not high, there is not much value for firm O to try to reduce firm C’s output to the final-product market by buying out firm C’s capacity. Therefore, firm C would always sell in the final-product market because of low profits in the component market resulted from lower acceptable wholesale prices by firm O.

Proposition 2.1 and Figure 2.2 present the optimal strategies of the two firms when $M > \bar{M}$. I define four regions: firm O outsources to firm T in Regions I and II, and firm O outsources to firm C in Regions III and IV. Moreover, firm C’s capacity constraint is binding in Regions I and III, and firm C has excess capacity in Regions II and IV. I also define capacity thresholds $k_{12}, k_{23}$ and $k_{34}$ (i.e., $k_{12} < k_{23} < k_{34}$) to describe the locations of the strategy changes in the equilibrium solution. For example, $k_{23}$ is located at the boundary of Regions II and III. I denote the optimal solution with superscript “*”. Firm C’s optimal wholesale price is denoted by $w_c^*$, and $\bar{w}_c$ is the
maximum acceptable wholesale price by firm O such that it is better off to outsource to firm C if and only if \( w_C \leq \bar{w}_C \). See Appendix-A for analytical expressions.

**Proposition 2.1:** When \( M > \bar{M} \), the optimal strategy is such that firm C sells products to customers and

1) If \( k < k_{23} \), firm O outsources to firm T and \( w_C^* > \bar{w}_C \).

2) If \( k \geq k_{23} \), firm O outsources to firm C and \( w_C^* = \bar{w}_C \).

The capacity constraint is binding if and only if \( k \leq k_{12} \) or \( k_{23} \leq k \leq k_{34} \).

When firm C has plenty of capacity (Region IV), it has sufficient capacity to produce for both firms. The literature that studies supply-chain partnerships with a competitor focuses on this region (e.g., Venkatesh et al., 2006; Xu et al., 2010; Wang et al., 2013). However, I show that even in Region IV firm O might accept a wholesale price from firm C that is higher than that of firm T (i.e., \( \bar{w}_C = \frac{17w_T}{16} \)). This is because firm O would be better off when cooperating with its competitor compared to the case where it outsources to firm T.

![Figure 2.2: Optimal outsourcing strategies when \( M > \bar{M} \)](image-url)
Firm C’s capacity is intermediate at Region III. When firm O outsources to firm C, it can reduce the supply of firm C’s product, and the competition in the final-product market can be mitigated. Therefore, firm O would outsource to firm C, and firm C would set a non-negative wholesale price that is higher than what is offered by firm T (i.e., \( w_C^* = \bar{w}_C > w_T \)). The outsourcing strategy in this region could explain how Philips caused TPV to reduce its own brand output by outsourcing the production of its monitors to TPV.

I show in Lemma A1 (in Appendix-A) that \( \bar{w}_C \) weakly decreases in the capacity of firm C. Therefore, one might expect that when firm C has a low capacity (Regions I and II, where firm C can charge a high wholesale price to firm O), instead of using the capacity to produce for its own product, firm C is better off using all capacity to supply to firm O so that firms O and C can hold monopolies in the final-product market and in the component markets, respectively. Interestingly, I find that firm O would not outsource to firm C in these two regions. This is because, on the one hand, firm C can sell its product to the customers at a high retail price, and hence, it would require a high wholesale price if it were to use the capacity to produce for firm O’s product instead of for its own product. On the other hand, competition between suppliers C and T provides a limit in terms of how high a wholesale price firm O is willing to accept. I find that the wholesale price that justifies firm C’s use of its capacity to produce components for firm O’s product is higher than the wholesale price that firm O is willing to accept (i.e., \( w_C^* > \bar{w}_C \)). Therefore, both firms would forgo the opportunity of holding monopolies in the component market and in the final-product market, and both would prefer to act as pure competitors. Moreover, when firm C sells a final product to customers when it has small capacity, it will price its products high, which in turn will allow firm O to price its product higher than its price in the monopoly market. Gelman and Salop (1983) showed similar results, where they found that when the two firms are not supply chain partners and the new entrant had limited capacity, it was not profitable for the incumbent to hold a monopoly in the market. I extend these authors’ results to the case where competitors are supply chain partners.
At Region II, firm O does not outsource to firm C, even though firm C has some unused capacity. This is because, by acting as firm O’s supplier, firm C would reduce its capacity to produce its own products, but the gain from component sales to firm O would not compensate for the loss from the reduction in final product sales. Therefore, firm C would set a high wholesale price (i.e., $w_C^* > \bar{w}_C$) to discourage firm O from outsourcing to firm C. This result can explain the market choices of some competitive CMs like Franz, who prioritize capacity to their own brands and turn down outsourcing contracts when facing capacity-allocation conflicts (Yan, 2013).

Proposition 2.2 presents the impact of capacity to demand, price and profit for the case where $M > \bar{M}$. Figure 2.3 illustrates the impact of capacity on demands and prices (Propositions 2.2a and 2.2b) and Figure 2.4 illustrates the impact of capacity on profit (Proposition 2.2c).

**Proposition 2.2**: Define $k_1 \equiv \frac{7(2+M)+3wr}{24}$, $k_3 \equiv \frac{56(2+M)-51wr}{105}$, $k_{23}^- \equiv k_{23} - \varepsilon$ and $k_{23}^+ \equiv k_{23} + \varepsilon$, where $\varepsilon$ is a small positive number. When $M > \bar{M}$, a. Firm C’s demand and the total demand may decrease in capacity: $D_C^*(k_{23}^-) > D_C^*(k_{23}^+)$ and $D_0^*(k_{23}^-) + D_C^*(k_{23}^-) > D_0^*(k_{23}^+) + D_C^*(k_{23}^+)$. b. Firm i’s price may increase in capacity: $p_i^*(k_{23}^-) < p_i^*(k_{23}^+)$. c. Firm C’s profit may decrease in capacity:

- $\frac{\partial \pi_C^*}{\partial k} < 0 \iff k_1 < k \leq k_{12}$ or $k_3 < k \leq k_{34}$
- \[ \left| \frac{\partial \pi_C^*}{\partial k} \right|_{k=k_1+\Delta} < \left| \frac{\partial \pi_C^*}{\partial k} \right|_{k=k_3+\Delta} \text{ for all } \Delta > 0 \]

When firm C’s capacity increases, one might expect that prices would decrease while demands would increase. I find that this intuition does not hold when the capacity of firm C is around $k_{23}$. This is because, as $k$ increases, the strategy changes from the two firms acting as pure competitors (Region I or Region II) to acting as cooperators (Region III). Therefore, firm C would shift some of its capacity to produce components for firm O, and thus, its demand $D_C^*$ decreases and price $p_C^*$ increases. Venkatesh et al. (2006) in Proposition 3 of their paper find that firms would set higher prices under
coopetition relationships. I extend this finding to a system with limited capacity by showing that firm C sets a higher price because, in addition to the two firms acting as coopetitors, firm C uses some of the capacity to produce for firm O and thus produces fewer units for itself. Firm O also sets a higher retail price $p^*_O$ because it shifts from using a cheaper supplier (firm T) to using a more expensive supplier (firm C), so it sets a higher retail price in order to maintain the margin.

![Figure 2.3: Firms' demand and price as a function of $k$](image)

When firm O outsources to firm C and firm C has sufficient capacity to produce for both firms (Region IV), or when firm O outsources to firm T and firm C has sufficient capacity to produce for itself (Region II), the firms' profits are not affected by the capacity level (cf. Figure 2.4). However, when capacity is binding, firm C may be worse off from having more capacity, even when that capacity can be available for free. This is because, in Regions I and III, the retail prices of both firms would decrease when firm C has more capacity, so competition becomes more intense. I find that the impact of a decrease in retail prices is larger than the impact of an increase in demand, so the profit of firm C decreases. In other words, selling the extra output requires firm C to lower its market price which results in lower overall profit. This can be interpreted as the cost of selling the extra output to the firm. Note that, because firm C’s capacity is public knowledge, firm O will reduce its market price as firm C’s capacity increases. Thus, when its capacity increases, firm C cannot prevent the increasing competition by limiting its output (by hiding its extra capacity from firm O) and thus keeping its market price high. Osborne and Pitchik (1985) also find that having a limited capacity could be
beneficial for the small firm as the reduction in competition due to smaller capacity may offset the lower output levels of the small firm.

Furthermore, the second part of Proposition 2.2c shows that the impact of profit-decreases-in-capacity (Region III) is larger when firms O and C cooperate as supply-chain partners than when firms act as competitors only (Region I). This is because firms would set higher prices under the coopetition scenario than under the competition scenario. When capacity increases, the decrease in price under the coopetition scenario is larger than that under the competition scenario, so the decrease in profit is larger. This finding illustrates the importance of considering the firm’s capacity constraint when competitors cooperate as supply-chain partners. One key takeaway is that firms are not always better off with more capacity, especially when they are coopetitors.

![Figure 2.4](image)

**Figure 2.4: Firms’ profit as a function of k**

The two products are more substitutable when they are located closer to one another (\( M \) decreases), leading to a more intense competition. Then Proposition 2.3 presents the impact of the degree of competition to the demand of firm C when \( M > \bar{M} \).

**Proposition 2.3:** When \( M > \bar{M} \), firm C’s demand may increase in the intensity of competition in the final-product market: \( \frac{\partial D_C^*}{\partial M} < 0 \iff k_{23} \leq k \leq k_{34} \) (entire Region III).
As expected, demand of a product decreases when competition intensify (in Regions I, II and IV). However, Proposition 2.3 shows that when the two firms are supply-chain partners and when firm C has no excess capacity (Region III), the demand for a product increases, even when the two products become more substitutable ($M$ decreases). This is because firm O’s demand decreases when $M$ decreases. Therefore, when capacity level is tight, firm C would allocate less capacity to produce for firm O’s product. Firm C would have more capacity to produce product under its own brand, so it would set a lower price $p_c^*$ for its product, in turn leading to a higher demand $D_c^*$.

In the results presented above, I assume that competition in the final-product market is low (i.e., $M > \bar{M}$). In this case, firm O would never multi-source to both suppliers (i.e., $0 < \gamma < 1$ is never optimal). This is because firm O will outsource as many components as possible to firm C if the wholesale price is less than $\bar{w}_c$. Therefore, the only possible scenario that might lead to a multi-sourcing is when firm C’s capacity is not enough to satisfy all firm O’s demand and firm C sets the wholesale price such that firm O is better off to outsource to firm C. However, in Proposition 2.1 I show that firm C always sells in final-product market even when its capacity is low, so firm C would set a high wholesale price to deter firm O from buying out its capacity. Thus, firm O would never multi-source.

Next, I will illustrate in Proposition 2.4 that firm O may multi-source when the competition between the two products is very intense ($M \leq \bar{M}$). I present the optimal strategy in Figure 2.5, and the analytical definition of the six regions are presented in Appendix-A. I define two new regions for the case where the competition is very intense: firm O multi-sources from firm C and firm T in Region V, and firm O outsources to firm C in Region VI. Moreover, in both Regions V and VI firm O holds a monopoly in the final-product market.
Figure 2.5: Optimal outsourcing strategies when $M \leq \overline{M}$

**Proposition 2.4**: When competition between the two products is very intense ($M \leq \overline{M}$),

a. Firm C does not sell in the final-product market in Regions V and VI.

b. Firm O multi-sources in Region V.

Recall from Proposition 2.1 that when the competition between the two final products is low ($M > \overline{M}$), firm C would always sell to customers in the final-product market. However, when the two products are very substitutable ($M \leq \overline{M}$), Proposition 2.4a shows that firm C might be better off serving as firm O’s supplier and opting not to sell in the final-product market (cf. Regions V and VI in Figure 2.5). There are two different reasons behind this strategy.

First, intuitively when firm T’s wholesale price is high (due to higher production cost), firm C can set higher wholesale prices compared to the cases where firm T has low wholesale price. Then, coupled with the fact that firm C cannot set a high retail price because of the strong degree of substitutability between the two products, firm C can earn more profit from selling components to firm O than from selling products in the final-product market. In these scenarios (Region VI), firm O holds the monopoly in the final-product market, and firm C holds the monopoly in the component market. In this case firm C has enough capacity to fulfill firm O’s component demand. Second, when firm
C’s capacity is low (Region V), firm O would benefit from pushing firm C out of the final-product market to avoid high competition and lower retail prices. This is because when the competition in the final-product market is very intense firm O can increase its retail price significantly if firm C does not sell in the final-product market compared to the case where firms compete in the final-product market.

Proposition 2.4b shows that when competition among final products is very intense there is a region (Region V) where firm O multi-sources its component from firms C and T. The reason for multi-sourcing is that in Region V firm C does not have enough capacity to fulfill firm O’s component demand. Therefore, in this case firm O as the monopoly in the final-product market can expand its demand by supplying the rest of its component demand from firm T. The size of the region where firm C does not sell in the final-product market (Regions V and VI) depends on the degree of competition between the final products. As competition of the two products intensifies (M decreases), the sizes of Regions V and VI increases. This is because firms O and C can avoid this intense competition by being monopolies in the final-product market and in the component markets, respectively.

2.5 Value of Competition in Component Market

Contract manufacturers sometimes have the proprietary rights to produce the component, but after the patent has expired, other suppliers can also produce it. For instance, Qualcomm served as the proprietary supplier of the CDMA chips for cell phone producers, and the expiration of its CDMA patents ended Qualcomm’s control over CDMA (Mock, 2005), resulting in an increase of competition in the cell-phone-chip manufacturing market.

The purpose of this section is to evaluate the impact of competition in the component market. In particular, in the main model, I assume that the component is not proprietary, such that firm O has the option of multi-sourcing from multiple potential suppliers (firms C and T). In this section, I consider a benchmark scenario in which the component is of a proprietary nature, and firm C is the only supplier that can produce the component. Firm C deploys one of the following strategies: (1) monopoly – does not
supply the component to firm O (e.g., sets a very high wholesale price) such that firm C
holds the monopoly in selling the final product, (2) supplier only – acts as a supplier for
firm O but does not enter the final-product market, and (3) coopetitor – supplies
components to firm O and sells final products to customers.

When firm C does not supply components to firm O (monopoly), firm O earns
zero profit, and firm C’s optimization problem is:

\[ \pi_c = p_c D_c \]
\[ \text{s.t. } D_c \leq k \]
(2.7)
(2.8)

When firm C supplies components to firm O and does not sell a final product in the final-
product market (supplier only), firm C’s optimization problem is:

\[ \pi_c = w_c D_o \]
\[ \text{s.t. } D_o \leq k \]
(2.9)
(2.10)

When firm C sells components to firm O and also sells final products in final-product
market (coopetitor), firm C’s profit is:

\[ \pi_c = w_c D_o + p_c D_c \]
\[ \text{s.t. } D_o + D_c \leq k \]
(2.11)
(2.12)

While under the monopoly strategy, firm O’s profit is zero, under the component-supplier
and coopetitor scenarios, firm O’s profit is:

\[ \pi_o = (p_o - w_c) D_o \]
(2.13)

The game sequence under the benchmark is as follows:

1) Firm C decides on its strategy (monopoly, component supplier, coopetitor).
2) If applicable, firm C decides \( w_c \).
3) If applicable, firm C decides \( p_c \) and firm O decides \( p_o \).
I use backward induction to find the equilibrium solutions; the derivation of equilibrium is presented in Appendix-A. I denote the optimal profit of firm O under the benchmark to be \( \pi^B_O \), and I define the value of competition to firm O as \( V_O \equiv \pi^*_O - \pi^B_O \). Recall that \( \pi^*_O \) denotes the equilibrium profit of firm O in the main model. Then, Proposition 2.5 examines how the capacity and competition in the final-product market affect the value of competition in the component market. (The value of competition from the perspective of firm C is the reverse of that from the perspective of firm O.)

**Proposition 2.5:**

a. The value of competition in the component market decreases in capacity: \( \frac{\partial V_O}{\partial k} \leq 0 \).

b. The value of competition in the component market decreases in the intensity of competition in the final-product market: \( \frac{\partial V_O}{\partial M} \geq 0 \).

c. Firm O is always better off if there is more than one supplier: \( V_O > 0 \).

Under the benchmark scenario, when the CCM’s capacity decreases, the wholesale price increases significantly because the CCM holds the monopoly in the component market. On the other hand, the wholesale price would be relatively insensitive to the capacity under the main model because of competition in the component market. Therefore, the value of competition in the component market is large when capacity decreases.

When the competing products are highly substitutable (i.e., small \( M \)), firm C has less incentive to allocate its capacity to firm O, so one might expect the value of competition in the component market to be large. However, I find that the opposite impact holds.

Consider the case where firm C supplies components to firm O under the benchmark. (Otherwise, firm O has zero profit, so the comparison is trivial.) Firm O’s profit is relatively insensitive to the product substitutability under the benchmark because, when it serves as the proprietary component supplier, firm C would set a wholesale price to extract as much profit from firm O as possible. On the other hand, under the main model, the wholesale price would be relatively insensitive to product
substitutability because of competition in the component market. As firm O stands to gain more when the product becomes less substitutable under the main model, the value of competition in the component market increases in $M$.

Despite the fact that the value of competition might decrease for firm O depending on firm C’s capacity or product substitutability, Proposition 2.5c shows that firm O always prefers to have more than one outsourcing option. This preference relates to the fact that the availability of alternative outsourcers would provide firm O with more bargaining power when designing a contract with the supplier.

2.6 Buy-and-hold Option

In the main model, I have assumed that firm O cannot order excess quantity. However, since firm O can manipulate firm C’s output to the final-product market by using a portion of the capacity, thereby mitigating competition in the final-product market, firm O might have the incentive to order more than its demand. Specifically, in the second step of the game, firm O decides how much to order from firm C and how much to order from firm T.

The analysis of this game is similar to the main model, though much more complex due to firm O’s order-quantity constraint. The proofs are available from authors upon request. In summary, it can be shown that the general structure of the equilibrium stays the same. Similar to the findings presented in Proposition 2.4a, firm C is better off not selling in the final-product market since product substitution increases in the presence of the buy-and-hold option. However, when there is a buy-and-hold option, the degree of product substitutability that is required for firm C to prefer to drop out of the final-product market is lower compared to the main model, where firm O does not have the buy-and-hold option. This is because firm O does not need to price its products low when it can have excess order quantity, so it may be better off selling its products for a monopoly price and keeping the extra components at no cost. Consequently, firm O would be willing to pay higher prices for firm C’s component in order to gain a monopoly in the final-product market. In turn, firm C has a larger wholesale price so it would not sell in the final-product market even when the products are less substitutable.
When firm C sells its own product to the end customer, firm O would never utilize the buy-and-hold option. This is because firm C would set a sufficiently high wholesale price when selling to end customers, so firm O would outsource to firm T instead. Moreover, firm O would exercise the buy-and-hold only when firm C has a relatively large capacity (Region VI) because, otherwise, firm O’s demand would always be enough to exhaust firm C’s capacity (Region V). Therefore, the existence of the buy-and-hold option for firm O would largely affect the area under which firm C is the supplier only and firm O is single-sourcing from firm C, thus making this strategy optimal for a larger set of parameters. I show that all of the results in the main model hold in the presence of the buy-and-hold option.

2.7 Conclusion

In this chapter, I examine the impact of capacity on the optimal channel structure when the contract manufacturer may have a competing product. Using a game theoretical approach to study the dynamics of firms’ strategic decisions allows us to better understand the effect of various factors on firms’ decisions – factors that might not be present in single-case or multiple-case study.

I show that capacity limitation, which is a commonly experienced conflict among contract manufacturers, can have nontrivial impacts. In particular, when firms act as supply chain partners, the CCM might reduce its own product output in order to fulfill the OBM’s outsourcing orders. Thus, I present an explanation for both prioritizing capacity to a firm’s own product (such as the strategy used by Franz in the motivating example) and reducing the firm’s own product output to shift profits to component sales to a competitor (such as the motivating example of TPV technologies). However, the former strategic move (by Franz) can be justified whenever a firm has limited capacity that can satisfy its own demand only, and the latter strategy (by TPV technology) becomes more plausible whenever a firm has a lot of excess capacity to accommodate OBM orders as well as most of its own demand. Furthermore, I characterize the conditions under which the OBM might benefit from multi-sourcing its component. I show the competition in the final-product market is very intense the OBM might benefit from buying out the CCM’s capacity while outsourcing the rest of the component demand to a third-party supplier.
I also show that firms’ prices might increase and demand might decrease as capacity increases. Interestingly, I find that the CCM’s profit may decrease in its capacity, and this deterioration becomes more severe when firms are supply-chain partners. For example, since 2014, LG Display produces the OLED screens used by OBMIs like Apple and also by its own firm, LG Electronics. Due to increasing demand in electronic markets, LG Display is planning to expand its production capacity to accommodate both OBMIs (e.g., Apple) and its own brand’s increasing supply needs (Fingas, 2015). However, my results show that it might be better for a contract manufacturer – in this example, LG Electronics – to limit the output of its own product in order to maintain higher market prices and thus reduce market competition. Therefore, in terms of managerial insights and implications, my results suggest that increasing capacity level may be harmful to the industry. Finally, I show that, in the component market, the value of competition to the OBM is small when the two products are highly substitutable.

Moreover, the results of my study provide strategic insights for practicing managers when their firms compete with a supplier that has limited capacity. In particular, when there is a need for capacity expansion to reach economies of scale, the disruption in the industry’s supply/demand balance “often leads to long and recurring periods of overcapacity and price cutting” (Porter, 2008). However, my results show that if the competitors can share the capacity, the increase in capacity levels may result in a more profitable industry rather than a price war. Such insight would increase the desirability of cooperating with competitors when there is a potential to achieve economies of scale through increased capacity levels.

I used a stylized model to study the dynamics of the firms’ optimal decisions, but this model had some limitations. I used a deterministic demand model, so a possible avenue of future research would be to analytically study the impact of demand uncertainty on the capacity-allocation problem of a CCM. Moreover, in my model, I considered only one competitive contract manufacturer that can produce a competitive product. It would be interesting to consider multiple strategic contract manufacturers, each with the option of producing their own brand products.
2.8 References


DRAMeXchange (2014). NAND Flash Brand Supplier Revenue Falls 6.6% in First Quarter:


Forbes (2013). Samsung's Sharp Investment Is Another Thorn In Apple's Side; Accessed on 27/02/2014 from:


Kim, M. (2012). Analysis: Friend and foe; Samsung, Apple won't want to damage parts deal


Chapter 3

3  Contracts to Promote Optimal Use of Optional Diagnostic Tests in Cancer Treatment

In this chapter I examine performance-based payment contracts to promote the optimal use of an optional diagnostic test for newly diagnosed cancer patients. This study is inspired by three trends: tremendous increases in the cost of new, advanced cancer drugs; development of new diagnostic tests to allow physicians to tailor treatment to patients; and changes in healthcare funding models that reward quality care. I model the interaction between two parties—a healthcare payer and an oncologist, in which the oncologist has private information about patients’ characteristics (adverse selection) and the payer does not know whether the oncologist takes the optimal course of action (moral hazard). I show that, in the presence of information asymmetry, an oncologist should never test all patients, even if the diagnostic test is available for free. Moreover, although the oncologist has additional information about a patient’s risk, he cannot always benefit from this private information. I also find that social welfare might increase as a result of an increase in the oncologist’s concerns regarding the health outcome of patients. Finally, I show that it is not always socially optimal to make a diagnostic test compulsory even if such a policy can be implemented for free.

3.1  Introduction

Cancer is the second leading cause of death in the United States, accounting for over 580,000 deaths in 2013 (Center for Disease Control and Prevention 2013). The American Cancer Society estimates that there will be over 1,680,000 new cancer diagnoses in 2016 (American Cancer Society 2016). Cancer treatment typically consists of a combination of surgery, radiation therapy, hormone treatment, and chemotherapy. In recent years, cancer treatment costs have risen dramatically with the development of a number of biologics and targeted treatments. For example, the cost of a course of therapy with trastuzumab (Herceptin) for breast cancer is estimated at $30,000 (Hornberger et al. 2005); the cost of a course of therapy with bevacizumab (Avastin) for colorectal cancer is approximately
$90,000 (Picard 2012), and other drugs have been developed with a wide range of monthly costs per patient (Campbell 2015).

In addition to growth in the number and cost of treatment options available, there has also been growth in the availability and use of personalized medicine, often implemented through genetic and gene-expression profiling tests, which can help guide cancer diagnosis and treatment decisions. In some jurisdictions, a KRAS test is required prior to treatment with cetuximab or panitimuab for colorectal cancer, based on evidence that these drugs are not effective in patients with mutations in the KRAS gene (Díaz-Rubio et al. 2012). The 21-gene assay for breast cancer (Oncotype Dx) generates a recurrence score based on the expression of 21 genes. The score is interpreted as the probability of cancer recurrence within 10 years and is used to guide adjuvant chemotherapy treatment decisions. In particular, women with a high recurrence score are advised to undergo adjuvant chemotherapy, whereas women with a low recurrence score may avoid adjuvant chemotherapy as it provides limited benefits but is associated with harmful side effects. A gene expression profiling test for cancer of unknown primary (the Tissue of Origin test) examines the expression of more than 2,000 genes and produces a set of 15 “similarity scores” that are interpreted as the probability that the tumor is one of 15 common types. This test can help to guide treatment decisions by directing treatment towards the most appropriate agents for a given type of cancer instead of using a general, non-specific treatment regimen. These diagnostic tests are an important element of personalized medicine since individualized treatment cannot be prescribed in the absence of this type of information.

The costs of these tests vary from less than $500 for single-gene tests like KRAS (Behl et al. 2012) to $4,000 or more for multi-gene tests (e.g., Oncotype Dx costs $4,175 (Ray 2011) and Tissue of Origin costs $3,750 (Tansey 2008)). Use of many of these tests is supported by clinical guidelines (Allegra et al. 2009, Carlson and Roth 2013) and a number of health technology assessments (e.g., Hannouf et al. 2012, Nerurkar et al. 2014, Hannouf et al. 2016). However, the uptake of these advanced tests is low despite evidence of benefits. For example, Segurado (2016) writes “despite solid scientific evidence and endorsement by oncology societies for multi-gene tests supporting therapy
decision-making in breast cancer, only a small percentage of women are able to safely skip chemotherapy through personalized medicine.” In the United States, 90 percent of people have insurance plans that cover the Oncotype Dx test, yet only 27 percent of eligible cancer patients used the test in 2010 (Enewold et al. 2015).

To help manage chemotherapy costs and incentivize physicians to provide high-quality care, the Centers for Medicare and Medicaid Services (CMS) has developed a new funding model called the Oncology Care Model (OCM). The OCM payment scheme consists of a “target price” for a six-month episode of treatment that is based on the average treatment cost of patients with similar health conditions. Under OCM, providers receive payments based on a fee-for-service (FFS) model where the providers are compensated based on the volume of services performed. At the end of each treatment episode the providers may receive reward payments, conditional on the patient’s health status, if the total cost of treatment is below the target price (Cms.gov 2015). The OCM thus creates conflicting objectives in managing care: while the target price aims to limit the treatment cost, the reward payments give physicians the flexibility of choosing more expensive treatments if it is believed that they will have a positive impact on the patient’s health status.

The OCM is part of a more general type of payment reform that focuses on performance-based bundled payments. Several other examples have been developed. For instance, CMS introduced a 5-year payment program in April 2016 for lower extremity joint replacement called the Comprehensive Care for Joint Replacement (CJR) Model. Under the CJR model, hospitals will be compensated based on a target price that aims to cover the total cost of an episode of treatment, defined as the period that starts with an admission to a participating hospital and ends 90 days post-discharge. Hospitals may also receive quality incentive payments based on the health outcome of patients. Similarly, the Cardiac Rehabilitation Incentive Payment Model was introduced by the Department of Health and Human Services (HHS) in 2016. Participating hospitals will receive quality incentive payments as well as a bundled payment for a 90-day care period following hospital discharge for patients hospitalized for a heart attack or bypass surgery. Recent studies suggest that there is growing interest in bundled payment models by healthcare
payers (Japsen 2015, Whitman 2016). Bundled payment models accounted for 15 percent of insurers’ spending in 2014 (Evans 2014) and there is evidence that these payment models have resulted in positive outcomes (Korda and Eldridge 2011).

In this essay, I develop a stylized physician compensation model inspired by the growing use of advanced diagnostic testing technologies in cancer care. Although diagnostic tests are not explicitly included in OCM, the growing number of expensive testing technologies and the need to ensure the appropriate use of these tests suggests a need to include them in future funding models. Thus, my research explores what a bundled payment mechanism should look like if diagnostic tests are available. The central problem analyzed in this chapter was identified in a report by consulting firm McKinsey and Co., which described one of the barriers to adopting new testing technologies for personalized medicine as “the difficulty of enforcing standard protocols to ensure that physicians follow through with appropriate patient care based on test results” (Davis et al. 2010).

My model involves interaction between a healthcare payer (payer) who pays for the test and advanced treatment option (e.g., adjuvant chemotherapy), and an oncologist (provider) who makes decisions regarding the use of the test and the advanced treatment for patients. Patients are heterogeneous, with different risks of cancer recurrence. Since the recurrence probability can only be known if the test is used, patients appear to be homogeneous members of different risk groups if the test is not used. The payer offers a payment contract to providers to maximize social welfare, defined as the monetary value of health outcomes minus the total costs of providing care. Depending on the payment structure offered by the payer, the provider chooses one of three treatment options: 1) Treat all patients with an advanced treatment option without ordering the test; 2) Do not order the test and do not treat with an advanced treatment option; or 3) Order the test and only treat those patients with advanced treatment option whose recurrence probability is sufficiently high. While in the main analysis I assume that the provider decides whether to use the test or not, in some cases the test can be made compulsory (e.g., KRAS); therefore, in Section 6 I extend my analysis to the case where the test is compulsory for all patients.
The contract design problem is complicated by information asymmetry in the system. First, there are many patient characteristics that are observed by physicians but are not formally recorded in any clinical or administrative databases held by the payer (adverse selection). Second, the payer also does not know which factors the physician considered when making the treatment decision, so the payer does not know whether the optimal course of action is used by the provider (moral hazard). To incentivize the provider to use the test whenever it is beneficial and to appropriately follow the recommendation of the test results whenever the test is used, the payer offers a menu of contracts consisting of a fixed payment per episode of treatment, and reward payments that are based on patients’ health and on whether or not treatment is applied.

The underlying problem has a structural novelty in that the provider’s decision to use an advanced treatment option is conditional on his decision to order the diagnostic test. Therefore, finding the optimal menu of contracts is a challenging problem as the constraints in the payer’s problem are not continuous. However, I construct the optimal menu of contracts by solving the payer’s problem for each treatment choice of the provider.

My study yields several interesting findings that have policy implications. First, I show how the payer can design contracts such that the provider may not be able to take advantage of his private information. In this optimal contract structure, the payer should incorporate the provider’s treatment choices in the contracts. In particular, the payer should offer reward payments for not treating certain types of patients with an advanced treatment option. This finding underlines empirical evidence on over-prescription of chemotherapy treatment (Smith and Hillner 2010) and suggests that “quality care” does not always depend on using the most expensive technology available. Furthermore, it shows that the advent of new diagnostic tests necessitates new reimbursement models because current payment models do not account for treatment choices made by the providers.

Second, if the test is relatively inexpensive, then it may seem natural that the provider should order the diagnostic test for all patients. Interestingly, when the test is
optional and there is imperfect information flow between the payer and the provider, I find that the payer should motivate the provider to only test certain types of patients, even if the diagnostic test can be available for free. This is in agreement with empirical evidence showing that some new technologies might not be cost-effective even if they can be acquired for free (Davis 2014, Davis and Akehurst 2016). However, under complete information it is optimal to test all patients when the test cost is sufficiently low. This result highlights the importance of information flow in the healthcare system and provides the first step in estimating the value gained by enhancing administrative databases held by the payer.

Third, I find that an increase in reputational concerns and altruistic behavior of a provider may increase social welfare. This result is inline with the inconsistent evidence regarding the impact of malpractice pressure on healthcare systems (Avraham and Schanzenbach 2010, Reyes 2010, Cotet 2012, Lakdawalla and Seabury 2012, Dai et al. 2016, Montanera 2016). Thus, I provide further support for Montanera (2016) who argues that tort reform, even if successful at reducing malpractice pressure, is not a “silver bullet” capable of achieving both cost reductions and quality improvements for all patients.

Finally, I show that it is not always socially optimal to make a diagnostic test compulsory even if there is no cost to implement such a policy. While compulsory testing could eliminate adverse selection, it does cause unnecessary testing of patients who are very likely to benefit (or not benefit) from an advanced treatment option based on their pre-test evaluation. Thus, this finding suggests that even if a diagnostic test is recommended by the guidelines, it might be better to leave the decision about ordering the test to the judgment of physicians instead of making the test compulsory for the whole population.

3.2 Literature Review

My study is primarily related to the stream of literature that considers design of mechanisms (e.g., performance-based payment models) for healthcare systems. For example, there are studies that design performance-based contracts to improve patient
access to care (e.g., Jiang et al. 2012, Jiang et al. 2016, Savva et al. 2016). Jiang et al. (2012) consider contracts between a payer and a healthcare provider that serves a heterogeneous patient pool using an online appointment-scheduling system. They propose a threshold penalty-payment scheme to reach a desired patient waiting time (performance measure). Jiang et al. (2016) and Savva et al. (2016) study performance-based incentives when there is direct competition among providers. In these studies, performance measure is included in the proposed contracts. However, in my study, performance measure that is a patient’s health outcome is probabilistic and cannot be included as a contract term. Fuloria and Zenios (2001) also design performance-based contracts motivated by Medicare’s End Stage Renal Disease (ESRD) program, which was established in 1973 to provide affordable dialysis care for ESRD patients. They assume that all patients are homogeneous with the same probability function of experiencing complications or death after treatment. My work differs in that I study heterogeneous patients that can only be sorted using a diagnostic test. Finally, Gupta and Mehrotra (2015) study the proposer selection problem of CMS for its “bundled payments for care improvement” initiative. Similar to my study, they find coordinating mechanisms to align the incentives of the payer and the provider. However, I focus on the heterogeneous patient pool visiting a provider that is already a participant of a payment program, while they focus on the provider competition and payer’s strategy to get optimal proposals from the providers.

There are also studies in healthcare operations management that analyze a provider’s choice regarding use of diagnostic tests in various healthcare settings. Motivated by the utilization of cancer screening tests, Yaesoubi and Roberts (2011) study optimal reimbursement contracts between a payer and a healthcare provider. The provider is assumed to only care about the cost of treatment, and thus the health outcome of a patient does not affect the treatment choice of the provider. Conversely, I consider the case where the provider is concerned about patients’ health outcome. Moreover, I study the effects of an optional diagnostic test that can help the provider when deciding on treatment, whereas in Yaesoubi and Roberts (2011), the use of a particular test is the provider’s only decision. Dai et al. (2016) study the effect of different service environments (e.g., provider’s misdiagnosis concern, reimbursement ceiling) on the
provider’s test-ordering behavior. Their study is focused on general diagnostic tests, such as MRI or EKG, that are used to identify whether a patient has a disease. In contrast, I am interested in personalized medicine and diagnostic tests that identify if a patient (who is already diagnosed with a disease) can benefit from a targeted therapy. Similar to my approach to modeling the provider’s altruistic behavior and reputational concerns when patients are not properly treated, Dai et al. (2016) explicitly modeled the provider’s misdiagnosis concern and show that both over-testing and under-testing are possible outcomes of these concerns. Finally, Dai et al. (2015) study the use of an optional test on decision-making of a provider regarding the use of percutaneous coronary intervention (PCI) for coronary artery disease. Similar to my approach they assume some degree of altruistic behavior by providers and model it as a conflict of interest between financial gains and patients’ health outcome. They show that the incorporation of optional advanced testing into decision-making might have non-monotonic effects on the use of PCI depending on the amount of the provider’s conflict of interest. Unlike my study, they do not incorporate the payer’s incentives in use of the optional test and do not study the potential payment systems that might align the incentives of the payer and the provider.

Adida et al. (2016) study the effects that FFS and bundled payment models have on patient selection and the treatment intensity decisions made by a provider. Similar to my model they consider a reputational penalty for the provider in case of a treatment failure. They find that the FFS model would result in overtreatment of patients. Guo et al. (2016) also compares the FFS model with a bundled payment model in a setting where patients are sensitive to waiting times. They show that when the patient pool is sufficiently small the bundled payment model dominates the FFS model in all performance measures. Similar to my study they find optimal reimbursement policy of a payer who maximizes the patient welfare. Andritsos and Tang (2015) consider a co-managed situation where both the provider and the patient can exert effort to reduce the readmission rate. They show that bundled payment outperforms FFS by causing the provider and the patient to exert more effort to reduce readmission. These studies investigate incentive problems in healthcare systems in order to compare the outcome of different payment models, but they do not consider performance-based bundled payments. In this chapter, I model newly proposed payment models that reward providers
for higher quality of care and design performance-based payment contracts that align the incentives of the payer and the provider while achieving socially optimal treatment levels. To the best of my knowledge, I am the first to study the use of performance-based contracts to coordinate the incentives of a healthcare payer and a provider while examining the impact of an optional diagnostic test on the provider’s decision-making process.

3.3 The Model

I consider a system that consists of a payer, a provider, and a pool of newly diagnosed patients who are eligible for an optional diagnostic test. I extend the analysis to a case where the diagnostic test is compulsory in Section 6. In this section, I consider a system with adverse selection (the provider has private information about the patients’ risk profile) and moral hazard (the payer does not know whether or not the treatment that is recommended by the test result is followed). The provider makes the testing and treatment decisions, and the payer incentivizes the provider with a payment contract in order to maximize social welfare. Consistent with current practice for the Oncotype Dx and Tissue of Origin tests, I assume that the test result is sent directly to the provider. I assume that the provider incurs the treatment cost (e.g., chemotherapy cost plus costs associated with administering treatment and monitoring patients) and is compensated via his contract with the payer, while the test cost is directly billed to the payer. Throughout the chapter I use the term “treatment” to mean an advanced treatment option, such as targeted therapy, that would be recommended from the results of a diagnostic test. Thus, whenever I say “do not treat a patient” it means that the provider is not treating the patient with a specific treatment (e.g., chemotherapy or a targeted treatment), but the patient will still be treated by other means that are appropriate given her pre-test diagnosis (e.g., hormone therapy or empiric chemotherapy).

The provider’s patients can have different risk levels. The risk profile of the patients is private information to the provider because the provider is able to make several observations about each patient that are not known to the payer. An example of private information is “Has a stable lifestyle”, which can influence many factors such as adherence to medication recommendations and preventive measures to prevent
recurrence, and thus influence the success of a given treatment recommendation. Other examples include detailed family history or various risk factors that are disclosed to a physician but not formally recorded in patient charts. In addition, the results of some lab tests will be sent directly to the physician. Depending on the structure of the payer’s administrative databases, the payer may know that a test has been ordered (because of an associated billing code), but not know the result of the test. A proliferation of new tests, each with their own reporting standards, will likely exacerbate this problem. For instance, the Tissue of Origin test reports results as a set of 15 similarity scores, along with a “most likely tissue of origin” and a “degree of confidence”. Recording the full results of this test would require customization of existing databases, which may not be a priority for the payer, particularly if the tests are used for rare conditions. The extent of this source of information asymmetry will vary among payers depending on the structure of their health records systems and their sophistication in using the information collected—but even for a sophisticated payer, the physician will know more about a patient than the payer does.

For this analysis, without loss of generality, I normalize the number of patients of the provider to one. I assume that patients can be one of two types: high-risk (type-$H$) and low-risk (type-$L$). The recurrence probability (risk score) of type-$i \in \{H, L\}$ patients is characterized by a random variable, $r_i \in [0,1]$, with corresponding distributions $F_H(r)$ and $F_L(r)$ with density functions $f_H(r)$ and $f_L(r)$. Thus, the expected recurrence probability of type-$i$ patients is $E[r_i] = \int_0^1 y f_i(y) \, dy$. I assume that patients’ risk distribution function $F_i(r)$ is continuous and strictly increasing on $[0,1]$. Since type-$H$ patients have a higher expected recurrence probability than type-$L$ patients, I assume that $F_H(r) \leq F_L(r)$ for all $r$. The provider can observe the type but does not know a patient’s recurrence probability $r_i$ unless the patient undergoes the test, and ordering the test reveals this patient’s recurrence probability to the provider. The payer only knows the probability of the patient type to be of type-$i$ with probability $\beta_i$, where $\beta_L + \beta_H = 1$.

I assume that treatment can reduce the recurrence probability to $k r_i$, where $0 < k < 1$ is the treatment’s hazard ratio such that the treatment is more effective when the hazard ratio $k$ is small. This assumption is consistent with the use of the 21-gene assay
for breast cancer where the original validation study found that the difference in the probability of recurrence for women treated with chemotherapy versus hormone therapy alone was increasing in the recurrence score (Paik et al. 2006).

The payer offers a menu of contracts \( \zeta_i = \{w_i, b_{ti}, b_{ni}\} \) that depends on the patient type, \( i \). The contracts consist of the following three terms: 1) A fixed bundled payment for an episode of treatment (\( w_i \)), 2) a reward for stable health of treated patients (\( b_{ti} \)), and 3) a reward for stable health of patients who are not treated (\( b_{ni} \)). The first term of the contract (\( w_i \)) only pays for quantity of treatment and captures the most widely implemented payment method in different healthcare systems. For example, \( w_i \) can be interpreted as FFS payment when it covers the cost of an episode of treatment. However, when \( w_i \) is different from the total cost of the treatment it can be interpreted as a target price (cf. CJR model) – meaning that the provider might receive more or less than the total cost of the treatment. The next two terms are the performance-based payments that depend on the health outcome of patients as well as the treatment choice of the provider. Note that, the payer might be able to benefit from the correlation between the health outcome of the patients and their risk profile by offering contract payments that are in line with the most probable outcome corresponding to the patient type, i.e., including payment terms that reward the recurrence in high-risk patients (Riordan and Sappington 1988). However, these payment structures are impractical and thus are not explored. Figure 3.1 illustrates the sequence of events.

![Figure 3.1: Sequence of events](image-url)

An episode of treatment costs \( c \) for each patient. Every untreated patient with recurrence incurs a welfare loss \( h \) that represents both the opportunity cost and psychological cost of a patient who is not properly treated. Thus, a payer that maximizes social welfare has to account for this welfare loss. I assume that this welfare loss will also
cause a disutility $ah$ for the provider. Existing literature has explained the provider’s disutility as being either a result of a loss of reputation (Adida et al. 2016) or disutility caused by altruism about patients’ health (Montanera 2016). I do not argue for either but instead claim $ah$ to be an aggregate value representing both effects. When $\alpha = 0$, adverse patient outcomes cause no disutility for the provider – i.e., no damage to the provider’s reputation, and no harm associated with the provider’s concerns for the welfare of his patients. I examine the impact of $\alpha$ on social welfare in Section 5.

There is a net monetary benefit $\lambda_s$ associated with a patient with no recurrence (i.e., stable health) and $\lambda_r$ in case of recurrence, $\lambda_r < \lambda_s$. Treatment may be associated with toxicities and adverse events. These cause the payer’s net monetary benefit to decrease by $z$, which is the expected cost of adverse events. I assume that $z$ is independent of $r$ on the grounds that adverse events are driven by specific drug regimens, whereas $r$ is driven by genetic risk factors.

3.3.1 Provider’s problem

For each patient, the provider has to decide whether or not to order the test and whether or not to treat the patient.

**Remark 3.1:** If the provider orders the test, then the provider chooses a threshold $\tau_i$ for each patient type and only treats patients whose recurrence probability is higher than this threshold (i.e., $\tau_i > \tau_i$). This is a result of the recurrence probability being continuous and monotonic, and the treatment effect being proportional to the recurrence probability (i.e., $kr_i$).

Because the risk score of a patient is only revealed if the test is ordered, the treatment threshold ($\tau_i$) decision of the provider is conditional on his test-ordering decision. I use a binary variable $x_i \in \{0,1\}$ to denote whether the provider orders the test ($x_i = 1$) or not ($x_i = 0$) when patients are of type-$i$. Thus, for a given patient type, the provider has three options (Figure 3.2): Option 0 – treat patients without ordering the test; Option 1 – do not order the test and do not treat patients; and Option 2 – order the
test and treat patients only if their recurrence probability is higher than a treatment threshold ($r_i > \tau_i$).

**Figure 3.2: Provider’s treatment choices**

If the provider chooses to treat patients without ordering the test (Option 0), then the provider’s objective function $g_t(x_i, \tau_i)$ is:

$$g_t(0,0) = w_i - c + b_{ti} \int_0^1 \left(1 - ky\right)f_i(y)\,dy$$

The first two terms are the fixed bundled payment less the cost for an episode of treatment and the third term is the reward payment for treated patients with stable health.

If the provider chooses not to order the test and not to treat patients (Option 1), then the provider’s objective function is:

$$g_t(0,1) = b_{ni} \int_0^1 \left(1 - y\right)f_i(y)\,dy - ah \int_0^1 yf_i(y)\,dy$$

The first term is the reward payment for untreated patients with stable health. The second term is the provider’s disutility when untreated patients experience a recurrence. Finally, if the provider orders the test (Option 2) and only treats patients if their recurrence score is above the treatment threshold ($r_i > \tau_i$), then the provider’s objective function is:
The first term is the fixed bundled payment less the cost for an episode of treatment for treated patients. The second and the third terms are the reward payments for untreated and treated patients with stable health, and the last term is the provider’s disutility when untreated patients experience a recurrence. A similar approach has been widely used in the literature to model the provider’s objective function and to incorporate the provider’s concerns regarding the health outcome of the patients (e.g., Dai et al. 2015, Adida et al. 2016, Dai et al. 2016, Montanera 2016).

The provider’s problem is to maximize his objective function, given the menu of contracts offered by the payer \( \zeta_i = \{w_i, b_{ti}, b_{ni}\} \), by choosing the treatment choice for each patient type.

### 3.3.2 Payer’s problem

I assume that the payer’s objective is to maximize social welfare, which is the monetary value of patients’ health outcome minus the total costs of providing care. Thus, assuming the cost of the test is \( B \), the payer’s objective function, when patients are of type-\( i \) is:

\[
v_i(\zeta_i, x_i, \tau_i) = (\lambda_s - b_{ni}) \int_0^{\tau_i} (1 - y) f_i(y) \, dy + (\lambda_s - z - b_{ti}) \int_{\tau_i}^1 (1 - ky) f_i(y) \, dy + (\lambda_r - h) \int_0^{\tau_i} y f_i(y) \, dy + (\lambda_r - z) \int_{\tau_i}^1 ky f_i(y) \, dy - w_i \int_{\tau_i}^1 f_i(y) \, dy - x_i B
\]

(3.4)

The first two terms are the monetary benefit, less the reward payments to the provider, for untreated and treated patients with stable health. The third and the fourth terms are the monetary benefit for untreated and treated patients with recurrence. The fifth and the sixth terms are the payments for an episode of treatment and the test cost. Note that the test cost can only be incurred if the provider orders the test, i.e, setting \( x_i = 1 \) under Option 2. The payer decides on the payment contracts that link the two parties. The payer solves the following program.
The provider’s objective function is the total social welfare in a system with information asymmetry. The \textit{OTC} (optimal treatment choice) constraints are moral hazard incentive constraints to ensure that under the payment contracts \(\{\zeta_L, \zeta_H\} \) the provider’s optimal treatment choice is \(\{x_t^*, \tau_t^*\} \). The next two constraints (\textit{ICL} and \textit{ICH}) are the provider’s adverse selection incentive-compatibility constraints. They ensure that the provider prefers contract \(\zeta_t\) for type-\(i\) patients to the contract for the other patient type and thus prevent patient misrepresentation. Note that, if the provider misrepresents a patient he may also choose to change the treatment choice. This is represented by the provider’s decision variables in the right-hand side of the \textit{ICL} and \textit{ICH} constraints. In particular, \textit{ICL} constraint means that when a provider faces a low-risk patient, he would be better off choosing the contract that is designed for low-risk patients. Similarly, \textit{ICH} constraint ensures that a provider would choose a contract designed for high-risk patients when he faces a high-risk patient. The last two constraints (\textit{IRL} and \textit{IRH}) are the provider’s participation or individual-rationality constraints for each patient type—meaning that the contracts for each patient type result in non-negative payoff for the provider.

I derive the equilibrium solution following the standard approach (Laffont and Martimort 2001; Chapter 7, Bolton and Dewatripont 2005; Chapter 6). A similar
approach is used in other papers that consider both adverse selection and moral hazard (e.g., Yaesoubi and Roberts 2011, Xiao and Xu 2012, Chick et al. 2016). Specifically, first, based on the extended revelation principle (Laffont and Martimort 2001; page 258), I restrict my attention to direct mechanisms where the contracts are a pair of optimal treatment choices for the two patient types (i.e., $[\xi_L, \tau_L, x_L]$ and $[\xi_H, \tau_H, x_H]$). Next, because the provider is exposed to higher risk when treating high-risk patients, the provider’s participation constraints require the contract for high-risk patients to have higher expected value than the contract for low-risk patients. Therefore, for high-risk patients the provider would not deviate to the contract designed for low-risk patients, and thus $ICH$ will not bind at the optimum. Furthermore, the provider has an incentive to misrepresent low-risk patients as high-risk patients, so for low-risk patients the provider can benefit from his private information, and thus $IRL$ will not bind at the optimum. Thus, omitting $ICH$ and $IRL$, the constraints $IRH$ and $ICL$ are binding at the optimum for all treatment choices of the provider. Therefore, the payer’s problem reduces to a maximization problem where the payer chooses optimal contracts subject to the two binding constraints ($ICL$ and $IRH$). Since the provider’s treatment threshold decision is private information (moral hazard), the payer chooses payment terms such that the provider’s best response results in payer’s optimal treatment threshold.

The underlying problem is novel in the sense that the provider has two decisions and the treatment threshold decision is conditional on the test-ordering decision. Moreover, the treatment threshold is the provider’s only hidden action (moral hazard) whereas the test-ordering decision is public information. Therefore, if the provider does not order the test for a patient, the payer knows that the provider has chosen to treat or not treat the patient without knowing patient’s recurrence probability (Option 0 or 1). Given that the treatment threshold decision is conditional on the test-ordering decision, the provider has nine treatment choices because there are two patient types and three

---

2 To relate to the literature in the supply chain coordination (e.g., Çakanyıldırım et al., 2012; Huang et al 2015; Yan et al., 2015), the provider with a high-risk patient corresponds to the supplier with a lower capability (high cost or low quality), and vice versa. Therefore, $IRL$ and $ICH$ are binding in those papers, while $IRH$ and $ICL$ are binding in my study.
treatment options \( \{x_i, \tau_i\} \) per patient type. I solve the payer’s maximization problem for each of the nine treatment choices (cf. Table A.3 in Appendix-B). In the equilibrium solution, given the model parameters the payer offers a menu of contracts that results in one of the nine treatment choices with the highest social welfare. The full details on the derivation of the equilibrium solution are shown in Appendix-B.

### 3.4 System Optimum

In this section, I present the optimal solution for the system with information asymmetry (second-best policy). In Subsection 3.4.1, I analyze a system with full information (first-best policy) to understand the inefficiencies caused by private information of the provider. I also explain how the proposed payment model outperforms current payment models in Subsection 3.4.2.

Define \( \hat{\tau} = \frac{z + c}{(1 + \alpha) h + (1 - k)(\lambda_g - \lambda_r)} \) as the critical treatment threshold; \( \hat{\tau} \) is large when the treatment adverse effect \( (z) \) or the treatment cost \( (c) \) is large, or when the cost associated with the provider’s reputation concerns and altruistic behavior \( (\alpha) \) is small.

Lemma 3.1 presents a property of the optimal treatment threshold choice of the provider.

**Lemma 3.1**: When the provider orders the test \( (x_i^* = 1) \), the optimal treatment threshold is the same for both patient types and is equal to the critical treatment threshold (i.e., \( \tau_i^* = \hat{\tau} \)).

Lemma 3.1 states that in the optimal solution the payer designs contracts such that when the provider orders the test he will choose the same treatment threshold for both patient types that is equal to the critical treatment threshold. This is because the treatment threshold corresponds to the “marginal patient” for whom the provider is indifferent about treatment and the determination of this marginal patient is independent of the full distribution of risks for the patient type \( F_1(r) \). Thus, the change to social welfare from treating or not treating a marginal patient is independent of patients’ risk distribution.
For expositional convenience, I define \( \Gamma(\tau) = \alpha h \frac{\int_0^{\tau} F_L(y) \, dy - \int_0^{\tau} F_H(y) \, dy}{1 - \tau + \int_0^{\tau} F_H(y) \, dy} \) where 

\( \Gamma(\tau) \) can be interpreted as the provider’s relative cost of not treating high-risk patients versus low-risk patients. The value \( \Gamma(\tau) \) is zero when the distributions of risk of the two types are identical; it increases in the difference between the two types of patients and in the treatment threshold \( \tau \). I define the following thresholds:

\[
\bar{B}_L = \frac{z^+ c}{\hat{\tau}} \left\{ \int_0^{\hat{\tau}} F_L(y) \, dy + \min[0, E[r_L] - \hat{\tau}] \right\} \tag{3.6}
\]

\[
\bar{B}_H = \frac{z^+ c}{\hat{\tau}} \left\{ \int_0^{\hat{\tau}} F_H(y) \, dy - \frac{\hat{\tau}}{1 - \beta z + c} \Gamma(\hat{\tau}) + \min\left[0, E[r_H] - \hat{\tau} + \frac{\beta}{1 - \beta z + c} \Gamma(1)\right] \right\} \tag{3.7}
\]

\( \hat{\tau} \) solves 

\[
\hat{\tau} = \left( \frac{\beta}{1 - \beta z + c} \Gamma(\hat{\tau}) + \int_0^{\hat{\tau}} F_L(y) \, dy - \int_0^{\hat{\tau}} F_H(y) \, dy \right) = E[r_L] \tag{3.8}
\]

\[
\hat{\tau}_L = E[r_L] \tag{3.9}
\]

\[
\hat{\tau}_H = \frac{E[r_H]}{1 - \frac{\beta}{1 - \beta z + c} \Gamma(1)} \tag{3.10}
\]

Threshold \( \bar{B}_L \) can be interpreted as the maximum willingness-to-pay for the test. There is a trade-off between the cost of the test and the expected risk to the patients. If the expected risk to the patients is much lower (higher) than the treatment threshold, then there is only a small probability that a patient would (would not) require treatment given the test results. Therefore, as the distance between the expected risk and the treatment threshold \( \hat{\tau} \) gets larger, the value of the test diminishes because the probability of making similar treatment decisions with and without the test gets higher.

**Lemma 3.2:** Define \( \bar{z} = (1 + \alpha)h + (1 - k)(\lambda_s - \lambda_r) \) and \( \bar{c} = (1 + \alpha)h + (1 - k)(\lambda_s - \lambda_r) - z \). If \( z > \bar{z} \) or \( c > \bar{c} \), then the patients are never treated.

Lemma 3.2 shows that if the cost of adverse events is sufficiently large (\( z > \bar{z} \)) or the treatment is sufficiently costly (\( c > \bar{c} \)), then the patients would never be treated. In that case, it is clearly never optimal to order the test. For the remainder of the chapter, I focus on the interesting region where \( z < \bar{z} \) and \( c < \bar{c} \).
Proposition 3.1 presents the optimal solution for the system with information asymmetry (second-best policy), and is illustrated in Figure 3.3.

**Proposition 3.1:** The optimal solution for the system with information asymmetry is one of the following five cases:

- **Case 1:** For both patient types, do not order the test and treat the patients if and only if $B > \bar{B}_L$ and $\hat{t} < \hat{t}_L$

- **Case 2:** For type-L patients, order the test and treat the patients with $r > \hat{t}$; For type-H patients, do not order the test and treat the patients if and only if $B \leq \bar{B}_L$ and $\hat{t} < \bar{t}$

- **Case 3:** For type-L patients, do not order the test and do not treat the patients; For type-H patients, do not order the test and treat the patients if and only if $B > \max\{\bar{B}_L, \bar{B}_H\}$ and $\hat{t} \geq \hat{t}_L$ and $\hat{t} < \hat{t}_H$

- **Case 4:** For type-L patients, do not order the test and do not treat the patients; For type-H patients, order the test and treat the patients with $r > \hat{t}$ if and only if $B \leq \bar{B}_H$ and $\hat{t} \geq \bar{t}$

- **Case 5:** For both patient types, do not order the test and do not treat the patients if and only if $B \geq \bar{B}_H$ and $\hat{t} \geq \hat{t}_H$

![Figure 3.3: Optimal treatment decisions for the system with information asymmetry](image-url)
Consider the case where the cost of the diagnostic test is sufficiently high, such that the payer should not motivate the provider to order the test, regardless of the patient type (Cases 1, 3, 5). In all three cases, the potential benefit of ordering the test—that is, directing treatment to those patients that derive the greatest benefit—does not justify the testing cost. It is analogous to public health and health policy guidelines (e.g., for screening, vaccination, etc.) where the guideline varies in the population prevalence. For example, one analysis of HIV screening concluded that one-time screening in the general population might be cost-effective in areas where HIV prevalence exceeds 3.0 percent (Paltiel et al. 2005).

Next, consider the case where the diagnostic test is relatively inexpensive ($B$ is small), such that social welfare may increase if the provider orders the test for the patients (Cases 2, 4). When the critical treatment threshold $\hat{t}$ is small (Case 2), the provider should test low-risk patients, and treat them when their recurrence probability is larger than $\hat{t}$. However, in Case 2, high-risk patients are expected to have a recurrence risk larger than the threshold $\hat{t}$, so the payer should design a contract that motivates the provider to treat high-risk patients without ordering the test. On the other hand, when the critical treatment threshold is large (Case 4), the provider should test high-risk patients and treat them when their recurrence probability is larger than $\hat{t}$. However, in Case 4, low-risk patients are expected to have a recurrence risk lower than the critical treatment threshold, so the payer should design a contract that motivates the provider to neither test nor treat low-risk patients.

Corollary 3.1 shows how the provider may benefit from his private information.

**Corollary 3.1:** In Cases 1, 2, and 3, the patient type does not affect the provider’s profit. In Cases 4 and 5, the provider has a higher profit from treating low-risk patients than from treating high-risk patients.

In Cases 1, 2, and 3, the provider always treats high-risk patients. As a result, the payer can design a contract for the provider for high-risk patients that (1) does not require any reward payments (i.e., $b_{nH} = 0$ and $b_{tH} = 0$), and (2) sets the fixed per patient treatment bundled payment for an episode of treatment to be the treatment cost (i.e.,
$w_H = c$), leading to the provider having zero profit. Note that this contract structure is independent of the patients’ risk distribution because there is no reward payments, so the provider with low-risk patients would also have the same profit if he were to pick this contract (incentive-compatibility constraint - ICL). As a result, the payer can design a contract such that the provider would also have zero profit from low-risk patients. Thus, patient type does not affect the provider’s profit for these cases.

In Cases 4 and 5, the provider should not always treat high-risk patients, so the payer must design a contract that includes reward payments for not treating high-risk patients (i.e., $b_{nH} > 0$). While treating high-risk patients will have no profit for the provider ($IRH$ is binding), the provider would have a positive profit when picking this contract for low-risk patients. This is because not treating low-risk patients with the advanced treatment option has less risk for the provider. Therefore, to prevent the provider from misrepresenting low-risk patients, the payer has to incentivize the provider by offering a profit larger than that offered for high-risk patients (i.e., information rent). In particular, the provider would have zero profit from high-risk patients in both cases, but the provider would have profit $\Gamma(\hat{\tau})$ in Case 4, and $\Gamma(1)$ in Case 5 from low-risk patients. (Recall that $\Gamma(\hat{\tau})$ is the provider’s relative cost of not treating a patient. The argument of $\Gamma$ is $\hat{\tau}$ in Case 4 because the provider does not treat high-risk patients with recurrence risk smaller than $\hat{\tau}$, and the argument of $\Gamma$ is 1 in Case 5 because the provider never treats high-risk patients, which is equivalent to setting the treatment threshold to one.)

Corollary 3.2 formally defines test cost thresholds beyond which it is never optimal to use the test.

**Corollary 3.2**: In the equilibrium solution, there exists an upper bound for the test cost such that

- if $B > \overline{B}_H|\hat{\tau}_H = \tau_H$ then test is never ordered for high-risk patient type, and
- if $B > \overline{B}_L|\hat{\tau}_L = \tau_L$ then test is never ordered for either patient type, and

\[
\overline{B}_L|\hat{\tau}_L > \overline{B}_H|\hat{\tau}_H.
\]
Corollary 3.2 shows that regardless of the risk level of the patients, there is a level for test cost \( (\bar{B}_L|_{\tau_t = \tau_L}) \) beyond which it never makes sense to use the test. It is intuitive that a threshold for the test cost exists, since there is a trade-off between the cost of the test and the value of the test in terms of improved treatment decisions. However, it may seem counterintuitive that this threshold is higher for low-risk patients (i.e., \( \bar{B}_L|_{\tau_t = \tau_L} > \bar{B}_H|_{\tau_t = \tau_H} \)). This is because in Case 4, the payer incurs information rent when motivating the provider to test high-risk patients (cf. Corollary 3.1). Therefore, ordering the test for high-risk patients will increase social welfare only when the test cost is low. On the other hand, in Case 2, the information rent is zero, and thus it is optimal for the payer to motivate the provider to test low-risk patients even at higher test cost amounts.

Corollary 3.3 presents a property of the optimal contracts.

**Corollary 3.3:** All payment and reward terms are non-negative in the optimal contracts.

The payer can offer a contract with non-negative payment terms as long as the provider gets reward payments for stable health of patients who are not treated (i.e., \( b_{ni} \geq 0 \)). This is because the provider incurs a cost related to his reputation and his altruistic behavior when a recurrence happens to untreated patients, so the provider has an incentive to over-treat patients. The proposed contracts counteract this by compensating the provider for stable health of untreated patients. Thus, this result highlights the necessity of incentivizing the provider based on his treatment choice (such as incentivizing for not prescribing an episode of chemotherapy treatment for breast cancer patients). Having non-negative payments is a desirable feature of contract design in healthcare applications because negative payments might be impossible to implement in practice and could cause additional incentive problems. A similar issue has been noted in other papers on the topic (Yaesoubi and Roberts 2011, Chick et al. 2016).

Note that when the reward payments are zero (i.e., \( b_{ti} = b_{ni} = 0 \)), the resulting contracts are simple bundled payments (i.e., \( w_t \)) that only pay for the quantity of the treatment without any consideration of the quality of the treatment. These types of contracts are similar to the commonly used FFS payment models. I show that when it is
optimal to use the diagnostic test (Cases 2 and 4) the optimal contracts cannot be of this form. Consequently, the advent of new diagnostic tests makes current reimbursement models that only reward the intensity of work sub-optimal and necessitates development of alternative reimbursement models such as performance-based bundled payment models.

**Corollary 3.4:** Under the payment scheme proposed in this section, in the equilibrium of a system with information asymmetry it is never optimal to test patients from both patient risk profiles, even when the test is free (i.e., \( B = 0 \)).

Corollary 3.4 is true for all optimal cases presented in Proposition 3.1. In Case 2, where \( \hat{f} \) is relatively small (Figure 3.3), the cost of adverse events and treatment cost are relatively small compared to the provider’s cost when recurrence happens to untreated patients. In this case, the provider treats high-risk patients without ordering the test; accordingly, as stated in Corollary 3.1, the provider will have zero profit from treating both patient types. If the provider were to test high-risk patients, the test results might suggest that a high-risk patient does not need to be treated; therefore, the provider would expect to have a disutility in case of a recurrence. This will require having higher-valued contract payment terms for high-risk patients if the provider were to order the test in Case 2. Following higher-valued contracts for high-risk patients, to keep the provider from misrepresenting low-risk patients, the payer needs to offer higher-valued contracts for low-risk patients as well. As a result, in order to avoid the additional costs generated through higher-valued contract terms, in Case 2 the payer should not motivate the provider to order the test for high-risk patients.

In Case 4, the cost for a recurrence in untreated patients is relatively small, and the provider neither tests nor treats low-risk patients. If the provider were to test low-risk patients, there is a small probability that the test suggests a low-risk patient to be treated. However, the savings to the cost related to treating low-risk patients (as it might be suggested by the test) does not justify higher payment terms that is required to motivate the provider to test low-risk patients. Thus, in Case 4 the payer should not motivate the provider to test low-risk patients, even when the test is free. In Appendix-B I illustrate
how testing patients from both risk profiles always results in lower value for the payer regardless of the difference between the distributions of the two types of patients.

3.4.1 Full information

To examine whether the finding of Corollary 3.4 is due to decentralized decision-making or information asymmetry, I consider a system where there is a perfect information flow between the payer and the provider (first-best policy). The derivation of equilibrium of this system is presented in Appendix-B, and the equilibrium solution is illustrated in Figure 3.4. The optimal treatment threshold in a system with perfect information is equal to the critical treatment threshold ($\hat{t}$) because it does not depend on the patients’ risk profile and thus is not affected by information asymmetry in the system. Unlike the system with information asymmetry (Proposition 3.1), under complete information there is a new optimal case (Case 6) where the payer should motivate the provider to order the test for both patient types. This occurs when the test cost is sufficiently low. This is because, in the full information scenario, social welfare is only affected by the health outcome of patients and not by private information of the provider. In the full information scenario, only a patient that can benefit from treatment is treated. However, when the provider has private information, not all patients who require treatment will be treated and not all the treated patients will benefit from the treatment. This finding illustrates the importance of information flow in a healthcare system in order to reduce the inefficiencies caused by the information asymmetry.

Remark 2: The highest willingness-to-pay for the test in the system with full information (first-best policy) is independent of patients’ risk profile, and is the same as the highest willingness-to-pay for the test in the system with information asymmetry ($\bar{B}_L|_{T=\hat{t}}$). This is because in Case 2 the provider does not benefit from his private information and thus the payer’s willingness-to-pay for the test is not affected by the information asymmetry.
3.4.2 Alternate payment models

I also compare the proposed payment model with two other arrangements that are similar to currently implemented payment models. The derivation of equilibrium for these payment models are available from authors upon request.

First, I analyze the performance of a single performance-based payment contract that is offered to both patient types (i.e., $\zeta' = \{w, b\}$) in a system with information asymmetry. This is similar to the OCM payment model. Under a single contract the payer cannot motivate the provider to treat each patient type differently. Therefore, the optimal treatment choices are Cases 1, 5 and 6. Note that, unlike in the full information scenario, in the presence of information asymmetry it is costly for the payer to test both patient types (Case 6) and thus this payment structure is always weakly dominated by the proposed payment model (i.e., $\zeta_i = \{w_i, b_{ti}, b_{ni}\}$).

Second, I analyze an alternative model where the payer offers a menu of contracts in which reward payments are only tied to the health outcome of a patient (i.e., $\zeta''_i = \{w_i, b_i\}$). This is similar to the CJR payment model where a risk stratification methodology is used to set different target prices for different patient types. The optimal
treatment choices under this model are qualitatively similar to the 5 cases presented in Figure 3.3. However, the payment model that I propose can better utilize the optional diagnostic test by better managing the inefficiency in the system because by incorporating the treatment choice of the provider into the payment terms the payer can further reduce the information rent in Case 4. Both contract structures analyzed in this section have inferior performance to the proposed payment model, and hence emphasize the importance of development of alternative reimbursement models for different healthcare systems given the availability and use of personalized medicine.

3.5 Comparative Statistics

In this section, I examine how social welfare is affected by treatment characteristics and the provider’s concern for the health outcome of his patients.

For expositional convenience, I define:

\[
\Omega_l = 1 - \hat{\tau} + \int_0^{\hat{\tau}} F_l(y) \, dy
\]  

(3.11)

Proposition 3.2 examines how social welfare is affected when untreated patients experience a recurrence.

**Proposition 3.2:**

a) **Social welfare increases when the welfare loss of an untreated patient with recurrence has higher impact on the provider (i.e., \( \alpha \) increases) if and only if**

\[
\frac{\hat{\tau}^2 (F_l(\hat{\tau}) \Omega_H - F_H(\hat{\tau}) \Omega_L) - (x+c) \Omega_H - ah \hat{\tau}^2 (\Omega_L - \Omega_H)}{(x+c) \Omega_H^2} (1-\beta) \int_0^{\hat{\tau}} y f_H(y) \, dy \leq B_H
\]

and

\[
E[r_L] \leq \frac{ah^2 (F_l(\hat{\tau}) \Omega_H - F_H(\hat{\tau}) \Omega_L) - (x+c) \Omega_H - ah \hat{\tau}^2 (\Omega_L - \Omega_H)}{(x+c) \Omega_H^2} (1-\beta) \int_0^{\hat{\tau}} y f_H(y) \, dy \]

b) **Social welfare always decreases when the welfare loss of an untreated patient with recurrence (\( h \)) increases.**

One might expect that, due to a more restrictive participation constraint and the resulting overtreatment of patients, social welfare cannot increase when the provider is more concerned about his reputation and his patients’ health. However, Proposition 3.2.a formally states that this intuition does not always hold. In particular, when the provider does not treat low-risk patients but uses the test for high-risk patients (Case 4), social
welfare might increase as reputational concern and altruistic behavior of the provider increases. This result arises because when \( \alpha \) increases, the provider treating high-risk patients, in order to decrease his risk exposure, would treat patients with higher probability (i.e., \( \tau^h \) increases), which causes a decrease in the difference in the payments between low-risk and high-risk patient types. This allows the payer to design a contract such that the surplus utility for the provider treating low-risk patients to be smaller, resulting in an increase in the payer’s objective function (maximizing social welfare).

Note that, this is only true when the expected recurrence probability of low-risk patients is sufficiently small that it results in a higher surplus utility for the provider when treating low-risk patients. The higher the surplus utility of the provider, the higher the potential gain for social welfare. Another interpretation for this finding that the social welfare might increase as \( \alpha \) increases could be that when \( \alpha \) is large, the provider becomes completely altruistic or has sufficiently high reputational concerns that there is reduced need to incentivize the provider through the reimbursement mechanism. Proposition 3.2.a is illustrated in Figure 3.5 using parameters listed in Table 3.1.

Proposition 3.2.b shows that I cannot observe the same impact on social welfare when the welfare loss of an untreated patient with recurrence (\( h \)) increases. This is because an increase in the welfare loss of a patient has a direct negative impact on social welfare and cannot be compensated for by reducing the surplus utility of the provider.
Table 3.1: Parameter estimations for numerical examples

<table>
<thead>
<tr>
<th>Model parameter</th>
<th>Estimated value</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost for an episode of treatment (c)</td>
<td>$30,000</td>
<td>Hornberger et al. (2005)*</td>
</tr>
<tr>
<td>Treatment adverse events (z)</td>
<td>$19,000</td>
<td>Hornberger et al. (2005) †</td>
</tr>
<tr>
<td>Monetary benefit from stable health ($\lambda_s$)</td>
<td>$100,000</td>
<td>Shiroiwa et al. (2010) ‡</td>
</tr>
<tr>
<td>Monetary benefit from recurrence ($\lambda_r$)</td>
<td>$60,000</td>
<td>Hornberger et al. (2005) §</td>
</tr>
<tr>
<td>Test cost ($B$)</td>
<td>$4,175</td>
<td>Ray (2011)</td>
</tr>
<tr>
<td>Probability of type-L ($\beta_L$)</td>
<td>0.7</td>
<td>Assumption</td>
</tr>
<tr>
<td>Type-L patient distribution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f_L(r) = 2 - 2r$</td>
<td></td>
<td>Assumption †</td>
</tr>
<tr>
<td>Type-H patient distribution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f_H(r) = 2r$</td>
<td></td>
<td>Assumption †</td>
</tr>
</tbody>
</table>

* Estimates of adjuvant chemotherapy costs vary depending on the type of regimen used. I estimate the cost of a treatment episode per course of trastuzumab.
† Average cost of minor, major, and fatal side effects.
‡ Since health quality that is considered for reward payments is measured for a six-month period after an episode of treatment, I only consider willingness-to-pay for one additional quality-adjusted life-year (QALY) gained.
§ Quality of life would drop to 0.7 when recurrence is stable, and to 0.5 when recurrence is progressive (base value 0.98). I use the average value of 0.6 to represent the average willingness-to-pay for QALY of patients with recurrence.
¶ These distributions are assumed for illustrative purposes. I conducted experiments with other distributions and obtained qualitatively similar results.

Lastly, the impact of changes in treatment adverse effect $z$, treatment cost $c$ and hazard ratio $k$ are intuitive and formally stated in Proposition 3.3.

**Proposition 3.3:** All else equal, the payer is always worse off as the treatment becomes less desirable (i.e., increase in $z$, $c$, or $k$), and this effect of treatment characteristics on the social welfare is always larger when the treatment threshold is relatively small.

### 3.6 Compulsory Testing

There is a wide diversity in guidelines and recommendations for new diagnostic tests. In addition, many major guidelines are only updated every few years meaning that guidelines may lag technical developments. For instance, the Oncotype DX test is recommended by the American Society of Clinical Oncology and the U.S. National Comprehensive Cancer Network guidelines to be part of the routine practice for early stage breast cancer patients but not required as part of treatment (Ragaz 2010 B). Some other tests, such as “Tissue of Origin” test for cancer of unknown primary, currently are
not recognized by any guidelines and the use of these tests depends on the physician’s
decision. Other tests are compulsory; for example, KRAS test must be performed before
treatment with cetuximab or panitimuab for colorectal cancer in Ontario.

In this section, I assume that the payer might make the use of diagnostic tests
compulsory and implement a mechanism to verify test results before reimbursing
treatment. I assume that in making the test compulsory the payer incurs a verification cost
$C_V$ to gather and/or interpret the results – otherwise there would still be a problem with asymmetric information.

The payer’s problem when the test is compulsory is as follows:

$$
\max_{\xi} V^{\text{com}} = v(\xi, 1, \tau) - C_V
$$

\hspace{1cm} (3.12)

\text{s.t} \quad \tau^* = \argmax_{0 \leq \tau \leq 1} g(1, \tau)

\text{ (OTC)}

\hspace{1cm} g(1, \tau^*) \geq 0

\text{ (IR)}

Program (3.12) is the modified version of the program (3.5) in the main model.
The main difference is that there is no adverse selection because the payer acquires
patient information and the test is used for all patients regardless of patients’ risk profile.
However, I assume that there is still moral hazard because the payer cannot impose a
treatment choice, as recommended by the test result, to the provider. Since the test is
compulsory the provider always orders the test (i.e., $x = 1$). Lemma 3.3 presents the
region in which the payer can improve social welfare by making the test compulsory.

**Lemma 3.3:** Define $\bar{B}^{\text{com}} = \frac{\bar{z} + c}{\bar{t}} \left\{ (1 - \beta) \left( \int_{0}^{\xi} F_H(y) \, dy + \min \left[ 0, E[r_H] - \hat{\tau} + \frac{\beta}{1 - \beta} \int_{\bar{z} + c}^{\bar{z} + c} \Gamma(1) \right] \right) + \beta \left( \int_{0}^{\xi} F_L(y) \, dy + E[r_L] - \hat{\tau} \right) \right\}$. Then, $\{V^{\text{com}} > V | C_V = 0\}$if and only if:

- $B \leq \frac{\bar{z} + c}{\bar{t}} \int_{0}^{\xi} F_H(y) \, dy$ and $\hat{\tau} \leq \left( \int_{0}^{\xi} F_H(y) \, dy + \int_{0}^{\xi} F_L(y) \, dy \right) / \left( 2 \int_{0}^{\xi} f_L(y) \, dy - 1 \right)$; 
  \text{Or}

- $B \leq \bar{B}^{\text{com}}$ and $\hat{\tau} > \left( \int_{0}^{\xi} F_H(y) \, dy + \int_{0}^{\xi} F_L(y) \, dy \right) / \left( 2 \int_{0}^{\xi} f_L(y) \, dy - 1 \right)$
**Proposition 3.4:** *It is not always socially optimal to make a diagnostic test compulsory even if such a policy can be implemented for free (i.e., \( C_V = 0 \)).*

![Figure 3.6: Compulsory test versus Optional test when \( C_V = 0 \)](image)

Figure 3.6: Compulsory test versus Optional test when \( C_V = 0 \)

Proposition 3.4 is illustrated in Figure 3.6. The shaded area in Figure 3.6 is the region where social welfare is higher when the test is compulsory. To make it easier to compare with the original case where the test is optional, I superimposed the shaded area on top of the graph from Figure 3.3 using the same scale on the y-axis. Figure 3.6 shows that while making the diagnostic test compulsory removes the adverse selection, even if the payer can implement this policy for free (i.e., \( C_V = 0 \)) it is not always optimal to do so. This is because the test result does not have much value for patients whose expected recurrence probability is higher than the optimal treatment threshold (\( \hat{r} \)). Thus when \( \hat{r} \) is small, if the test is used for all patients, the payer pays for a test that might only change the treatment choice for a small group of patients. However, the compulsory testing policy becomes more effective when the optimal treatment threshold is relatively large, such that there is a higher probability for patients to have recurrence score below the treatment threshold and thus a diagnostic test is needed to eliminate unnecessary treatment of low-risk patients.
Corollary 3.5: There exists an upper bound for the test cost such that if \( B > B^{\text{com}}_{\hat{t} = t_H} \) it is never optimal to make the test compulsory.

Corollary 3.5 formally shows the maximum test cost beyond which it is never optimal to make the test compulsory. However, as the verification cost for implementation of the policy increases, the shaded region in Figure 3.6 will shift down and compulsory testing will be optimal in a smaller region. Therefore, I can argue that while the maximum verification cost that justifies the compulsory testing depends on the test cost (\( B \)), this threshold cannot be larger than the maximum test cost that makes the compulsory test optimal when \( C_V = 0 \) (Lemma 3.3).

3.7 Conclusions

In this chapter I examined optimal design of contracts between a payer and a provider when an optional diagnostic test can be used to predict the benefits of an advanced treatment option. Following the general framework of OCM, the contracts include reward payments based on the health of patients at the end of a review period. The provider has private information about patients’ characteristics (adverse selection), and about whether or not the optimal course of action is used (moral hazard).

My research provides evidence for the necessity of developing alternative payment models; for the importance of enhancing payers’ data bases; on potential impact of tort reform on social welfare; and on the benefits of implementing compulsory testing policy. However, this essay is based on a stylized mathematical model with several limiting assumptions. While it is common to only study two types in systems with adverse selection (e.g., Yaesoubi and Roberts 2011, Çakanyıldırım et al. 2012, Xiao and Xu 2012, Saghafian and Chao 2014, Yan et al. 2015) the analysis for multiple types can be derived using similar methodology (see Bolton and Dewatripont 2005). In this case, there will be multiple perceived risk distributions for patients (i.e., \( F_n(r) < F_{n-1}(r) < \cdots < F_1(r) \) where \( F_1(r) \) is the distribution for the patients with the lowest risk and \( F_n(r) \) is the distribution for the patients with the highest risk). The payer will offer a menu of contracts to the provider by considering the provider’s incentive-compatibility and individual-rationality constraints. The provider will receive a zero profit from treating
patients with highest risk $F_n(r)$, and the provider might have a positive surplus for all other patient types (information rent). The challenging part of the derivation is to reduce the incentive-compatibility constraints because the provider might have an incentive to misrepresent the patients from higher-risk distributions as patients from lower-risk distributions. However, given the structure of my findings, I expect that the main results of my study (such as non-negative contract terms and not using the test for all patient types even if it was free) will continue to hold when there are multiple types. Generalization of the underlying problem to other testing technologies may require explicit modeling of the sensitivity and specificity of the tests. I assumed that there is only one provider with a fixed number of visiting patients. Thus, using the proposed payment model, the payer cannot eliminate the provider’s information rent in a single agent system with private information. However, if patients have multiple options when deciding on which provider to visit, then a fruitful avenue of future research would be to consider competition among multiple providers. In such a competitive setting, the payer might be able to absorb all the information rent of for any treatment choice of the provider (see Cremer and McLean 1988, Demski and Sappington 1984).

3.8 References


payment-loses-some-interest-among-providers/#7466e89011ce [Accessed September 18, 2016].


Chapter 4

4 Care-Coordination: Gain-sharing Agreements in Bundled Payment Models

4.1 Introduction

Hip and knee replacements are the most common inpatient surgery for Medicare patients in the United States. The total cost for surgery, hospitalization, and recovery can vary dramatically among providers. For example, the cost of a total knee replacement procedure could range from $16,772 to $61,585 across different providers (Casey, 2015). The average cost for a total hip replacement procedure follows similar trends. In 2014, there were more than 400,000 hip and knee replacements procedures, costing more than $7 billion in hospitalization expenses alone. These procedures often require lengthy recovery and rehabilitation periods that increase the cost of the procedure. Demand for these procedures is expected to grow significantly over the next 10 years, to over 4 million by 2030—an increase of 174% (Froemke et al., 2015). Despite the high volume of these surgeries, the quality and costs of care still vary greatly among providers. For instance, the rate of complications (e.g., infections, implant failures, etc.) after surgery can be more than three times higher at some facilities compared to others. These complications increase the chances that a patient may be readmitted to the hospital, which will further increase costs and may also reduce quality.

Consequently, the Centers for Medicare and Medicaid Services (CMS) introduced a five-year pay-for-performance program starting in April 2016 for lower extremity joint replacement (LEJR) called the Comprehensive Care for Joint Replacement (CJR) model. Under the CJR model, hospitals will be compensated based on the total cost of an episode of treatment, defined as the period beginning with admission of an LEJR patient to a participant hospital and ending 90 days after discharge. This time period is long enough to cover the complete period of recovery for most patients. Each episode includes all related items and services paid under Medicare’s fee-for-service (FFS) model, such as long-term care hospital services, inpatient rehabilitation facility services, and skilled nursing facility services (Cms.gov 2015).
The CJR model is a retrospective bundled payment model whereby, prior to the start of each performance year, CMS provides participant hospitals with a target price that is the expected sum of expenditures of all parties. Within the performance year, CMS continues to pay all parties involved in the treatment of a patient according to an FFS model, whereby services are paid based on claims from providers. At the end of the performance year, the total expenditure will be adjusted according to the pre-specified target price by penalizing or rewarding the hospital. The target price will be adjusted according to the patient’s condition and historical hospital-specific spending and regional spending for LEJR episodes (The U.S. Department of Health and Human Services (HHS) 2015).

Following completion of a CJR model performance year, participant hospitals that have actual episode spending below the target price and achieve a minimum acceptable quality score will receive a reward payment from Medicare for the difference between the target price and actual episode spending. Hospitals with LEJR episode spending that exceeds the target price will be financially responsible and must pay back the excess expenditure to Medicare. CMS expects the five-year program to yield $153 million in net savings (Dickson and Evans, 2015).

Even though a target price and associated reward or penalty are offered to hospitals, other providers involved in LEJR episodes of care throughout the performance year will be paid under existing Medicare FFS payment systems. The goal is to give hospitals a financial incentive to work with providers to coordinate care (HHS 2015). Therefore, participant hospitals can enter into contracts with collaborating providers and suppliers who provide services to the patient during an episode of treatment. A participant hospital may share with other providers the reward payment received from Medicare as a result of reduced episode spending and the hospital’s internal cost savings. The hospital may also share financial accountability (i.e., penalties) with other providers for increased episode spending (CMS, 2015). The CJR model thus shifts the risk of excess cost to hospitals, and incentivizes hospitals to arrange agreements with other providers to coordinate care. Accordingly, it is hospitals’ responsibility to share the gains
and losses from total cost of treatment with other providers in a way that motivates the providers to reduce their cost while achieving the required quality of care.

In this chapter, I am investigating three questions regarding the impact of a target price bundled payment model on care-coordination. First, from the hospital’s point of view, what is the best way to structure a gain-sharing agreement, whereby each hospital shares a portion of its reward or penalty with the provider(s)? To address this question, I study gain-sharing agreements, which depend on the treatment cost and patients’ quality of care, to coordinate care and reach the desired health quality and cost savings by the hospitals. I model a hospital and a provider that provide care for a specific procedure to patients. The hospital decides the portion of the reward or penalty to be shared with the provider. Then, the hospital and the provider decide on their respective cost-reduction efforts that would reduce the total cost of treatment.

Second, from the payer’s perspective, how does the choice of target price affect each party’s cost-reduction efforts and the total expenditure? To answer this question, I investigate the impact of the gain-sharing agreements on the outcomes of interest when compared to some benchmark scenarios where there is no gain-sharing agreement between the hospital and the provider.

Third, what are the implications of target price bundled payment models for the payer? I provide practical guidance for the payer in anticipation of the percentage of savings and the savings sensitivity to the FFS payments and cost-reduction efficiency of the hospital and the provider.

My research yields several interesting results. First, I show that when there is a gain-sharing agreement between the hospital and the provider, the provider may be better off when the target price is low; this is because when the target price is low, the hospital needs the provider to exert effort to reduce the total treatment cost as it is more efficient for the provider to reduce the cost of its procedure. Furthermore, I show that despite the fact that it is easier for a provider to reduce its costs, the hospital might exert more cost-reduction effort than the provider because the hospital would rather exert some effort in-house than share more of the reward with the provider.
In the CJR model, CMS uses a discount factor, applied to the initial treatment cost, to determine the target price. However, I show that the payer can lower its total spending by imposing a target price on the hospital, even when the payer sets the target price equal to the current treatment cost of a patient under the FFS model, because the potential reward payment would incentivize the hospital and the provider to reduce their respective treatment costs. Further, I demonstrate that, under a target price bundled payment model, using a gain-sharing agreement is a win-win-win scenario for the payer, the hospital and the provider. At the same time, I also note that if the payer does not properly set the target price, it might not incentivize the hospital to arrange gain-sharing agreements with the provider, and to exert all the cost savings in-house instead.

I illustrate that a target price bundled payment model would be more effective in healthcare settings where there is greater asymmetry between the hospital and the provider’s cost-reduction efficiency, because in networks with higher asymmetry in the ability to reduce costs, there is an increased need to align the incentives of healthcare providers to coordinate care. Finally, I show that changing the FFS payments to reach higher savings through target price bundled payment might negatively impact the quality of care.

This chapter is organized as follows. I first present the relevant literature in Section 4.2. Next, I introduce the status quo FFS model in Section 4.3. In Section 4.4, I present the outcome of a bundled payment model where the hospital does not enter into a gain-sharing agreement with the provider. I also introduce a benchmark centralized system where there is no money transfer between the hospital and the provider in section 4.4. I analyze gain-sharing agreements in Section 4.5. To understand the implications of a target price bundled payment model to the payer, I perform numerical analysis on the equilibrium solutions in Section 4.6. Finally, concluding remarks and the study’s managerial insights are found in Section 4.7.

4.2 Literature Review

This study is primarily related to the stream of literature that considers bundled payment models. For example, Gupta and Mehrotra (2015) study CMS’s proposer selection...
problem for its “bundled payments for care improvement” initiative. They find coordinating mechanisms set by the payer to optimally incentivize and select providers to participate in the bundled payment program. However, I focus on hospitals that already participate in CMS’s mandatory CJR bundled payment program. Adida et al. (2016) study the effects that FFS and bundled payment models have on patient selection and the treatment intensity decisions made by a provider. Similar to my model, they compare the outcomes of the FFS model with those of the bundled payment model. They find that the FFS model would result in overtreatment of patients. Guo et al. (2016) also compare the FFS model with a bundled payment model in a setting where patients are sensitive to wait times. They show that when the patient pool is sufficiently small, the bundled payment model is superior to the FFS model in all performance measures. Andritsos and Tang (2015) consider a co-managed situation where both the provider and the patient can exert effort to reduce the readmission rate, and conclude that bundled payment outperforms FFS by causing the provider and the patient to exert more effort to reduce readmission. These studies investigate incentive problems in healthcare systems in order to compare the outcome of different payment models, but they only consider a single provider that will receive the bundled payment. Contrastingly, in this study, I model bundled payments that cover the total treatment cost of patients who require procedures from a hospital and a provider. I investigate how the hospital can use gain-sharing agreements to coordinate care and optimally use the bundled payment to incentivize the provider.

My research closely relates to research on gain-sharing and risk-sharing agreements. For example, Barros (2011) studies the interactions between a pharmaceutical company and a payer with and without a risk-sharing agreement. He assumes patients respond successfully to the drug only with some probability, and the price for the drug is set by the drug manufacturer. He then shows that depending on when a risk-sharing agreement is negotiated (i.e., before or after the price has been set for the drug), the agreement may increase or decrease social welfare. Although I incorporate the probabilistic health outcome of a patient, in my study, a patient’s required health quality depends on both the hospital’s and the provider’s cost-reduction efforts. Zhang et al. (2011) study a price-volume agreement between a drug manufacturer and a payer in the presence of asymmetric information on demand. The authors consider a one-period
problem in which the unit sales price and the rebate rate are offered by the payer to the manufacturer. Mahjoub et al. (2013) study risk-sharing agreements between a drug manufacturer and a healthcare payer to manage uncertainties regarding the cost and effectiveness of drugs. They model a risk-sharing agreement in which a proportion of total sales is rebated to the payer. They investigate the conditions under which the manufacturer will make a profit. These studies analyze the coordination of a pharmaceutical company with a payer where the pharmaceutical company’s revenue is generated through direct sales. In my study, the hospital and the provider can achieve the required cost savings and quality of care only through collective effort.

There are also papers that study supply chain coordination through gain-sharing agreements. For example, Corbett and DeCroix (2001) study a supply chain with a supplier and a customer where the consumption of indirect materials can be saved by both the supplier and the customer. They compare a variety of contracts, including joint investment and shared-savings contracts, in which both parties would benefit from any reduction in consumption. They show that such shared-savings contracts can always lead to higher channel profits, but not necessarily to lower consumption. In Corbett and DeCroix (2001)’s study, as in my study, the total cost depends on the collective effort of the providers to reduce the total cost. However, I study a scenario where both total cost and health quality depend on the collective effort of the hospital and the provider.

Iida (2012) studies a single manufacturer that is exposed to probabilistic demand. The manufacturer decides on the quantity to produce, and the final product consists of \( n \) components that are supplied from \( n \) suppliers. The production cost of the final product depends on an individual supplier’s efforts. Suppliers and the manufacturer can exert effort to reduce the production cost of each component. They develop effort-sharing agreements by which the parties share the benefits of cooperative cost reduction, and effort-compensation agreements that compensate the parties for the costs of their respective efforts. Unlike the system in Iida’s research, where each party receives a reward based on its own cost savings, I analyze a system where the collective cost savings can be distributed among the parties involved.
Bernstein et al. (2015) study a decentralized assembly line consisting of a buyer and many independent suppliers. The components are assembled into a finished product, which is sold to the downstream market with a deterministic demand. Suppliers can invest in process-improvement activities to reduce the fixed production costs. The assembler establishes a knowledge-sharing network that is modeled as a cooperative game. Bernstein et al. (2015) identify and compare various cost allocation mechanisms that are feasible in the cooperative game, and show that the optimal system investment levels can be achieved only when the most efficient supplier receives the incremental benefits of the cost reduction achieved by other suppliers due to the knowledge transfer. Notably, these studies have only one target outcome—that is, the production cost—while in my study, the hospital must achieve a minimum care quality threshold in order to be eligible for the reward payment.

4.3 Fee-for-Service (Strategy S)

First, I analyze a status quo system where there is no mechanism in place to coordinate care. I model a hospital and a provider that receive FFS payments from a payer. I assume that FFS payments cover a profit margin \( \alpha \) on the treatment cost incurred by each party.

There is a pool of patients who will undergo a treatment that requires procedures from both the hospital and the provider. The hospital is responsible for procedure \( H \), and the provider is responsible for procedure \( P \). Let \( c_H(\cdot) \) and \( c_P(\cdot) \) be the cost of the procedures by each party. Without loss of generality, I normalize the number of treated patients in a performance year to one.

Both the hospital and the provider can reduce the cost of their respective procedures by exerting effort. Let \( e_i, \ i \in \{H, P\}, \) denote the effort exerted by each party to decrease the cost of its respective procedure. Therefore, the total cost of treatment for each patient is a function of individual effort (i.e., \( c_H(e_H) + c_P(e_P) \)). I assume cost is linear and decreasing in effort, such that \( c_i(e_i) = \bar{c}_i - b_i e_i \) where \( c_i(0) = \bar{c}_i \) represents the initial cost of the respective procedure for the hospital and the provider before any cost-reduction effort. Because each party undertakes different parts of the treatment, the cost function of the hospital is different from the cost function of the provider, and thus
the effect of cost-reduction effort is different as well. I assume that it is costlier for the hospital (rather than the provider) to reduce its costs (i.e., $b_p \geq b_H$)—a reasonable assumption since the hospital pays for two large and inelastic components of the treatment (the implant and physician salaries), and must also cover significant fixed overhead costs, whereas the provider may reduce its cost by switching between available resources (e.g., using forced walking instead of continuous motion machines for rehab, or changing nursing hours per patient day). Thus, most of the savings can be reached through post-procedure healthcare services, consistent with Saxena (2016). Without loss of generality, I normalize the coefficient for hospital’s effort to one (i.e., $b_H = 1$), which implies $b_p \geq 1$. To incorporate diminishing returns, let $k(e_i) = e_i^2$ be the cost of effort exerted by each party. As a result, under the FFS model, the hospital’s profit can be written as follows:

$$\Pi_H^S(e_H) = \alpha c_H(e_H) - e_H^2$$

Similarly, the provider’s profit can be written as follows:

$$\Pi_P^S(e_p) = \alpha c_p(e_p) - e_p^2$$

The first part of the equation is the total profit from FFS payments, and the second part is the cost of the cost-reduction effort exerted by each party. Both parties’ profit functions are decreasing in the cost-reduction effort because in the FFS model, there is no incentive to reduce the procedure cost. Accordingly, under the FFS model, no cost-reduction effort will be exerted (i.e., $e_H^* = e_p^* = 0$), resulting in $\Pi_H^S(e_H) = \alpha \bar{c}_H$ and $\Pi_P^S(e_p) = \alpha \bar{c}_p$ with the total status quo treatment cost of $c^S = \bar{c}_H + \bar{c}_p$.

## 4.4 Bundled Payment: FFS with Target Price

To reduce the total spending, the payer can impose a bundled payment for total treatment cost to motivate the hospital and the provider to exert cost-reduction effort. I analyze a target price bundled payment model based on the CJR structure. In this scheme, for the treatment of each patient, there is an upper bound $c_t$, representing the target price for the total cost of treatment, above which the hospital is penalized for the extra cost. The
hospital will also receive a reward payment when the treatment cost is below the target price, so I define total savings ($S$) of the system as the difference between the target price and the total cost of treatment (i.e., $S = c_t - (c_H(e_H) + c_P(e_P))$).

In the CJR model, the hospital is eligible for a reward payment only if all providers achieve a desired quality of care. The quality of care from the provider is independent of the quality of care from the hospital, and so the effect of cost-reduction effort on quality of care is separately evaluated, and the hospital will receive the reward payment only if the patient has received adequate care ($q_t$) from both the hospital and the provider (Figure 4.1). Each party may fail to provide the required quality of care. For example, in the CJR model, the major quality measures are based on patient-reported outcomes (PROs) and the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey scores (Terry, 2016). While these scores are based on customer satisfaction, the quality of care provided by each provider in the treatment of a patient’s condition affects these scores. Thus, I assume that the required quality of care is uncertain, and can be achieved with some probability. I further assume that quality decreases in line with an increasing effort to reduce costs (the lower the cost, the lower the quality of care). I assume a linear effort-dependent function such that $p_i(e_i) = 1 - e_i$. To ensure well-defined probabilities, I bound the cost-reduction effort such that $0 \leq e_i \leq 1$.

![Figure 4.1: Payment scheme depending on performance](image)
4.4.1 Centralized system (Strategy C)

As a benchmark scenario for a target price bundled payment model, I analyze a centralized system in which the hospital (central planner) provides all the procedures and care required for the treatment of a patient during an episode of treatment (single provider). In a centralized system, a central planner’s expected profit is as follows:

\[
\Pi^C(e_H, e_P) = \alpha (c_H(e_H) + c_P(e_P)) + \left[ c_t - (c_H(e_H) + c_P(e_P)) \right]^+ p_H(e_H) p_P(e_P) + \left[ c_t - (c_H(e_H) + c_P(e_P)) \right]^- - e_H^2 - e_P^2
\]

The first term is the central planner’s profit from FFS payments. The second term is the central planner’s reward when the required quality is met and the total cost of treatment \(c_H(e_H) + c_P(e_P)\) is lower than the target price \(c_t\). The third term occurs when the total cost of treatment is higher than the target price, and thus the central planner is penalized regardless of the quality of care (Figure 4.1). The fourth and fifth terms are the cost of the cost-reduction effort that the central planner has exerted for each procedure.

The central planner’s problem is to maximize its profit by exerting cost-reduction effort for the two procedures required to treat a patient.

\[
\max_{e_H, e_P} \Pi^C(e_H, e_P)
\]

s.t. \[0 \leq e_i \leq 1 \quad i \in \{H, P\}\]

Because the cost savings are deterministic, the equivalent problem for the central planner is to maximize its expected profit by choosing from the following two strategies.

1) Reward strategy:

\[
\max_{e_H, e_P} \Pi^{cr}(e_H, e_P) = \alpha \left( (\bar{c}_H - e_H) + (\bar{c}_P - b_p e_P) \right) + \left( c_t - (\bar{c}_H - e_H) - (\bar{c}_P - b_p e_P) \right) p_H(e_H) p_P(e_P) - e_H^2 - e_P^2
\]

s.t. \[c_t - (\bar{c}_H - e_H) - (\bar{c}_P - b_p e_P) \geq 0\]

\[0 \leq e_i \leq 1 \quad i \in \{H, P\}\]
2) Penalty strategy:

\[
\max_{e_H, e_p} \Pi^{cp}(e_H, e_p) = \alpha ((\bar{e}_H - e_H) + (\bar{e}_p - b_p e_p)) + c_t - (\bar{e}_H - e_H) - (\bar{e}_p - b_p e_p) - e_H^2 - e_p^2
\]

s.t. \[c_t - (\bar{e}_H - e_H) - (\bar{e}_p - b_p e_p) \leq 0\]

\[0 \leq e_i \leq 1 \quad i \in \{H, P\}\]

Proposition 4.1 illustrates the optimal outcome and effort level of the central planner. Thresholds \(c_t^{CL} < c_t^{CH} < c_t^{CH'} < c_t^{CH''}\) are defined in Appendix-C. The parameter \(\omega\) is also defined in Appendix-C as the critical point of a differentiable function.

**Proposition 4.1:** The central planner will

1) Receive reward payment if and only if \(c_t^{CH} < c_t\). In this case,
   
   \[e_p^* = \omega\]

2) Have zero savings if and only if \(c_t^{CL} \leq c_t \leq c_t^{CH}\). In this case,
   
   \[e_H^* = \frac{\bar{e}_H + \bar{e}_p - c_t}{1 + b_p^2} \quad \text{and} \quad e_p^* = \frac{b_p (\bar{e}_H + \bar{e}_p - c_t)}{1 + b_p^2}\]

3) Be penalized if and only if \(c_t < c_t^{CL}\). In this case,
   
   \[e_H^* = \frac{1-\alpha}{2} \quad \text{and} \quad e_p^* = \frac{b_p (1-\alpha)}{2}\]

Proposition 4.1 is illustrated in **Figure 4.2**.
Figure 4.2: Optimal solution in the centralized system when
\[ \bar{c}_H = \bar{c}_P = 2; b_P = 2; \alpha = 5\% \]

Proposition 4.1 shows how the central planner can exert cost-reduction effort to achieve cost savings. The result is intuitive, in that the central planner should exert more effort through the provider because the provider can reduce the cost of its procedure with less effort. This intuition holds, regardless of the target price \( c_t \), since the provider’s optimal effort always weakly dominates the hospital’s optimal effort (i.e., \( e_H^* \leq e_P^* \) for all cases). Note that both the hospital and the provider’s optimal cost-reduction efforts are weakly decreasing as the target price increases (\( \frac{d e_t^*}{d c_t} \leq 0 \)); this is because (1) exerting cost-reduction effort decreases the FFS payments, and (2) the higher the target price, the easier it is to earn the reward payment. When the target price is small (i.e., \( c_t < c_t^C \)), the central planner cannot achieve cost savings (i.e., \( S^C < 0 \)) and will be penalized. In this region, while the central planner exerts the maximum effort level through the provider, it limits the cost-reduction effort of the hospital as the reduction in the penalty by the hospital does
not compensate for the required investment. Figure 4.2 is illustrative, and the shape and the insights are the same for other parameterizations.

Proposition 4.2 shows how the central planner’s profit in a centralized system with a target price is different from the total profit in the status quo decentralized system with simple FFS payment model. $c_t^c$ is formally defined in Appendix-C.

**Proposition 4.2:** The central planner’s profit is higher than the total profit in the status quo FFS model if the payer sets a target price that is higher than a threshold $c_t^c$. (i.e.,

$$\Pi_t^C \geq \Pi_H^S + \Pi_P^S \Leftrightarrow c_t^C \leq c_t$$

As illustrated in Figure 4.3, the hospital and the provider would prefer a centralized system with the target price bundled payment model to the status quo FFS model only if the payer sets a target price that is higher than the $c_t^c$ threshold. Thus, in a centralized system, $c_t^c$ is the absolute lower bound to the target price that the payer can set.

---

**Figure 4.3:** Profit and savings of the centralized system when

$$\bar{c}_H = \bar{c}_P = 2; b_P = 2; \alpha = 5\%$$
4.4.2 Decentralized system (Strategy D)

In this scenario, I assume that both the hospital and the provider will be compensated based on the FFS model but the payer will impose a target price on the hospital based on the total system cost. Thus, the hospital will be rewarded or penalized depending on the total cost of treatment compared to the target price. Because the provider will only be compensated according to the FFS model, the provider does not exert any cost-reduction effort (i.e., \( e_p^* = 0 \)), resulting in \( \Pi^D_p(e_p) = \alpha \bar{c}_p \).

The hospital’s expected profit can be written as follows:

\[
\Pi^H_h(e_H) = \alpha c_h(e_H) + \left[ c_t - (c_h(e_H) + c_p(e_p)) \right]^+ p_h(e_H) p_p(e_p) + \left[ c_t - (c_h(e_H) + c_p(e_p)) \right]^\sim - e_H^{\frac{2}{2}}
\]

The first term is the hospital’s profit from FFS payments. The second and the third term represent the hospital’s reward and penalty depending on the total cost of treatment. The fourth term is the cost of the cost-reduction effort.

Proposition 4.3 illustrates the optimal outcome and effort level of the hospital in a decentralized system. Thresholds \( c_t^{DL} < c_t^{DH} < c_t^{DH'} \) are defined in Appendix-C.

**Proposition 4.3:** The hospital will

1) Receive reward payment if and only if \( c_t^{DH} < c_t \). In this case,
   i. if \( c_t^{DH} < c_t \leq c_t^{DH'} \), then \( e_H^* = \frac{1}{4}(\bar{c}_h + \bar{c}_p - c_t + 1 - \alpha) \) and \( e_p^* = 0 \)
   ii. if \( c_t^{DH'} < c_t \), then \( e_H^* = 0 \) and \( e_p^* = 0 \)

2) Have zero savings if and only if \( c_t^{DL} \leq c_t \leq c_t^{DH} \). In this case,
   \( e_H^* = \bar{c}_h + \bar{c}_p - c_t \) and \( e_p^* = 0 \)

3) Be penalized if and only if \( c_t < c_t^{DL} \). In this case,
   \( e_H^* = \frac{1 - \alpha}{2} \) and \( e_p^* = 0 \)

Proposition 4.3 is illustrated in Figure 4.4.
It is obvious that the savings under the centralized system are always weakly higher than the savings under the decentralized system because in the decentralized system, where the hospital receives (pays) all the reward (penalty) payments, the provider will not exert any cost-reduction effort. Figure 4.5 illustrates the system savings in the decentralized system using the same scale as in Figure 4.3.

**Proposition 4.4:** In the decentralized system, the hospital’s profit is higher than its profit in the status quo FFS model if the payer sets a target price that is higher than a threshold $c^p_t$ (i.e., \( \exists c^p_t | c^p_t < c^p_f \) s.t. \( \Pi^p_H > \Pi^S_H \iff c^p_t \leq \Pi^c_t \)). The provider is not affected when the target price is only imposed on the hospital (i.e., \( \Pi^p_p = \Pi^S_p = \alpha \Pi_p^p \)).

Proposition 4.4 shows that in a decentralized system with no gain-sharing agreement, the hospital and the provider would participate in a target price bundled payment model only if the payer sets a target price that is higher than the $c^p_t$ threshold. Moreover, this lower bound is always higher than the lower bound for target price in the centralized system (i.e., $c^c_t < c^p_t$).
In this section, I investigate gain-sharing agreements to examine (1) how the hospital can incentivize the provider to exert cost-reduction effort, and (2) whether both parties can be better off under these agreements. To achieve the desired quality of care and the target price, the hospital can offer the provider a gain-sharing agreement where the hospital shares a portion $x_p$ of the reward and a portion $y_p$ of the penalty with the provider, and keep $1 - x_p$ portion of the reward and $1 - y_p$ of the penalty for itself where $x_p \in [0,1]$ and $y_p \in [0,1]$. Thus, the hospital’s expected profit can be written as follows:

$$\Pi_H^G(e_H, x_p, y_p) = \alpha c_h(e_H) + (1 - x_p)[c_t - (c_h(e_H) + c_p(e_p))]^+ p_H(e_H) p_p(e_p)$$

$$+ (1 - y_p)\left[c_t - (c_h(e_H) + c_p(e_p))\right]^+ - e_H^2$$

In this model, if the patient receives the required quality of care (i.e. $p_H(e_H) p_p(e_p)$), then the hospital will keep $1 - x_p$ of the reward. Similarly, if there is a penalty, the provider must pay $y_p$ of the penalty. The provider’s expected profit can be written as follows:
The cost functions are public information, and thus the provider knows the cost incurred by the hospital. Because it is easier for the provider to change the cost of its procedure by switching between its resources, the provider can adjust its cost-reduction effort, given the treatment cost at the hospital. Therefore, I assume that the provider decides on its cost-reduction effort after observing the hospital’s cost-reduction effort. Figure 4.6 illustrates the sequence of events.

**Figure 4.6: Sequence of events**

At the first stage of the game, the hospital decides on the gain-sharing agreement \( \{x_p, y_p\} \) to maximize its expected payoff. I assume that the provider accepts the contract offered by the hospital (I do not impose a participation constraint), because I am interested in the absolute lower bound on the target price (set by the payer) that makes both the hospital and the provider participate in the target price bundled payment model. For these reasons, the focus of the results will be on the non-binding case where the provider is better off when accepting these agreements. Accordingly, the payer’s problem is as follows:

\[
\begin{align*}
\max_{e_H, x_p, y_p} & \quad \Pi_{IH}^G(e_H, x_p, y_p) \\
\text{s.t.} & \quad 0 \leq e_H \leq 1 \\
& \quad 0 \leq x_p \leq 1 \quad \text{and} \quad 0 \leq y_p \leq 1
\end{align*}
\]

Next, the provider, knowing the contract terms, maximizes its expected payoff by investing in cost-reduction effort:
Max \[ \Pi^G_p(e_p) \]

s.t. \[ 0 \leq e_p \leq 1 \]

I solve the problem using backward induction. The full details on the derivation of the equilibrium solution are shown in Appendix-C. Proposition 4.5 illustrates the optimal outcome and effort level of the decentralized system with gain-sharing agreement.

Thresholds \( c^G_t < c^G_H < c^{GH'}_t < c^{DH'}_t \) and expressions \( \phi, e_p(c_t, \phi), \) and \( x_p(c_t, \phi) \) are defined in Appendix-C. Proposition 4.5 is illustrated in **Figure 4.7**.

**Proposition 4.5:** The hospital and the provider will

1) Receive reward payment if and only if \( c^G_H < c_t \). In this case,

i. if \( c^G_H < c_t \leq c^{GH'}_t \), then \( e^*_H = \phi; e^*_p(c_t, \phi); x^*_p(c_t, \phi) \)

ii. if \( c^{GH'}_t < c_t \leq c^{DH'}_t \), then \( e^*_H = \frac{1}{4}(\bar{c}_H + \bar{c}_p - c_t + 1 - \alpha); e^*_p = 0; x^*_p = 0 \)

iii. if \( c^{DH'}_t < c_t \), then \( e^*_H = 0; e^*_p = 0; x^*_p = 0 \)

2) Have zero savings if and only if \( c_t^G \leq c_t \leq c_t^{GH} \). In this case,

\( e^*_H = 0; e^*_p = \frac{\bar{c}_H + \bar{c}_p - c_t}{b_p}; x^*_p = 1 \)

3) Be penalized if and only if \( c_t < c_t^{GL} \), where

\( e^*_H = 0; e^*_p = \frac{1}{2} b_p (1 - \alpha); y^*_p = 1 \)

![Figure 4.7: Optimal solution in the decentralized system with gain-sharing agreement when \( \bar{c}_H = \bar{c}_p = 2; b_p = 2; \alpha = 5\% \)](image-url)
When the target price is small (i.e., \(c_t \leq c_t^{GL}\)), the hospital and the provider cannot achieve cost savings and will be penalized. In this case, the hospital will not exert any cost-reduction effort and will set \(y_p^* = 1\) to minimize its loss. However, when rewards are possible (dark shaded region), the hospital may exert effort to reduce the total treatment cost. In this case, the hospital will get the reward and may share it with the provider \((x_p^* \geq 0)\). The portion of the reward that the hospital shares with the provider decreases as the target price increases because when the target price is high, it is easier to be eligible for reward payment; thus, even if the provider can reduce its cost more efficiently, the hospital does not need to incentivize the provider to exert as much cost-reduction effort.

**Corollary 4.1:** *In the decentralized system with gain-sharing agreements,*

1) *The hospital may exert more cost-reduction effort as the target price increases*

\[ \frac{d e_H^*}{d c_t} \geq 0 \iff c_t^{GH} < c_t \leq c_t^{GH'} \]

2) *The hospital may exert more cost-reduction effort than the provider*

\[ e_H^* \geq e_p^* \iff c_t^{GH'} < c_t \leq c_t^{DH'} \]

Because the provider can reduce its cost more efficiently, intuition may suggest that most of the savings should be generated through the provider (as is the case in Section 4.4.1., “Centralized system”) when the hospital uses gain-sharing agreements. However, Corollary 4.1 shows that this is not always the case when using a gain-sharing agreement. In fact, when the target price is high, in order to have higher probability of qualifying for reward payment, the hospital needs the provider to reach the required quality of care instead of exerting cost-reduction effort. Therefore, to make it unattractive for the provider to exert cost-reduction effort, the hospital will reduce the provider’s share from the reward. To compensate for the increased cost (i.e., because of decreased effort by the provider), the hospital will increase its own cost-reduction effort.

Furthermore, when the target price is too high (i.e., \(c_t^{GH'} < c_t\)), the provider’s effort does not compensate for the cost of the gain-sharing agreement for the hospital, and thus the hospital would rather not use a gain-sharing agreement \((x_p^* = 0)\) and instead switch to Strategy D and exert all the cost-reduction effort in-house. This result highlights the fact
that if the payer does not set the target price properly, the target price bundled payment strategy would not incentivize the hospital to arrange gain-sharing agreements with providers to coordinate care and reduce the treatment cost.

**Corollary 4.2:** *The provider’s profit may decrease as the target price increases.*

One might expect that both the provider and the hospital would be better off as the target price increases. Yet, as illustrated in Figure 4.8, this intuition may not hold when the parties are in the reward zone (dark shaded region) because when the target price is high, the hospital will exert most of the cost savings in-house to avoid sharing the reward with the provider (Corollary 4.1). Hence, the provider may be better off when the target price is low, because in this case, the hospital needs the provider to exert effort to reduce total cost of treatment and thus increase savings.

![Figure 4.8: Profits and savings of the decentralized system with gain-sharing agreement when \( \bar{c}_H = \bar{c}_P = 2; b_P = 2; \alpha = 5\% \)](image)

**Proposition 4.6:** *In the decentralized system with gain-sharing agreement, both the hospital’s and the provider’s profits are higher than their respective profit in the status quo FFS model if the payer sets a target price that is higher than a threshold \( c^G_t \) (i.e.,

\[
\exists c^G_t \mid c^G_t < c^P \quad s.t \quad \Pi^G_H \geq \Pi^*_H \quad \text{and} \quad \Pi^G_P \geq \Pi^*_P \quad \Leftrightarrow \quad c^G_t \leq c_t.
\]

Proposition 4.6 shows the absolute lower bound to the target price that the payer can set in a decentralized system with gain-sharing agreements such that both the hospital
and the provider are better off compared to the status quo FFS model. In the decentralized system, this threshold is always smaller when the hospital uses gain-sharing agreements as opposed to bearing all the reward/penalty (i.e., $c_t^G < c_t^P$).

### 4.6 Implications for the Payer (Numerical Analysis)

In the previous sections, I presented the analytical solution for the equilibrium of each strategy and found the absolute lower bound for the target price that the payer can set in a target price bundled payment model. In this section, I perform numerical analysis on the equilibrium solutions to better understand the implications of the target price bundled payment model to the payer. I evaluate different parameter settings for the provider’s cost-reduction efficiency and the FFS profit margin (i.e., $1 \leq b_p \leq 2$ and $5\% \leq \alpha \leq 20\%$). From the payer’s perspective, the purpose of imposing a target price policy is to minimize its total spending, which can be defined as follows:

$$
\omega(e_H, e_P) = (1 + \alpha)(c_H(e_H) + c_P(e_P)) + \left[ c_t - (c_H(e_H) + c_P(e_P)) \right]^+ p_H(e_H) p_P(e_P) + \left[ c_t - (c_H(e_H) + c_P(e_P)) \right]^-
$$

The first term is the FFS payments plus the profit margin that the hospital and the provider receive on their respective treatment cost. The second and third terms are any potential reward or penalty. **Figure 4.9** illustrates the total spending in each strategy.

![Figure 4.9: Total spending when $\bar{c_H} = \bar{c_P} = 2; b_P = 2; \alpha = 5\%$](image)
Remark 4.1: *The payer can reduce total spending relative to FFS, regardless of the hospital and the provider’s strategy (i.e., C, D, or G), by imposing a target price that is equal to the total cost in the FFS model (i.e., $c_t = c^S$).*

To determine the target price for the CJR model, CMS uses a discount factor on the total cost of treatment under FFS (i.e., it sets a lower target price to achieve cost savings). However, my numerical analysis shows that the lower bounds for the target price are always smaller than the total cost of treatment under FFS (i.e., $c_t^C < c_t^S < c_t^D < c^S$). Thus, the payer can lower its total spending, using a target price bundled payment strategy, even when the payer sets a target price that is equal to the current cost of treatment (i.e., $c_t = c^S$). Nevertheless, in a decentralized system, if the payer knows that there will be a gain-sharing agreement between the hospital and the provider, the payer can minimize the total spending by setting the target price $c_t = c_t^S$; otherwise, the payer can only reduce the target price to $c_t = c_t^D$.

Remark 4.2: *In a decentralized system, both the hospital and the provider prefer to use gain-sharing agreements (i.e., $\Pi_H^G \geq \Pi_H^P \geq \Pi_H^S$ and $\Pi_P^G \geq \Pi_P^P = \Pi_P^S$) if $c_t^C \leq c_t < c_t^{GH'}$.*

Remark 4.2 shows that when the payer sets an acceptable target price (i.e., $c_t^C \leq c_t$), both the hospital and the provider would rather sign a gain-sharing agreement than not sign an agreement. Following Corollary 4.1, the upper bound on the target price is where the provider’s effort does not compensate for the cost of the gain-sharing agreement for the hospital. Thus, I provide a range for the target price such that, by setting a target price in this range, the payer can incentivize the hospital and the provider to enter into gain-sharing agreements.

In addition, my numerical experiments show that the payer’s total spending is always lower when the hospital signs gain-sharing agreements with the provider unless $c_t^{GH'} \leq c_t$ (Proposition 4.5) such that the hospital does not sign a gain-sharing agreement even when it has the option to do so. Hence, in a target price bundled payment model,
signing a gain-sharing agreement would be a “win-win-win” scenario in that it offers higher profit for the hospital and the provider, and lower spending for the payer.

To understand the impact of the provider’s cost-reduction efficiency ($b_p$) and the FFS profit margin ($\alpha$) on the total spending, I show the ratio $\frac{\Omega^G}{\Omega^S}$ for different parameter settings in Figure 4.10.

**Remark 4.3:** The provider’s cost-reduction efficiency has more impact on the payer’s spending compared to the FFS profit margin.

![Figure 4.10: Ratio of $\frac{\Omega^G}{\Omega^S}$ for different values of $c_t$, $b_p$, and $\alpha$](image)

Remark 4.3 implies that the payer might achieve greater savings by applying a target price bundled payment model in settings where the healthcare providers have higher asymmetry in cost-reduction efficiency; this is because higher asymmetry in the ability to reduce cost increases the need to align the incentives of all parties.

Given a cost-reduction efficiency, the payer can always increase the savings by reducing the profit margin, which will incentivize the participants to reduce the total treatment cost to be eligible for a reward payment. However, reducing the total spending might negatively impact the quality of care. To evaluate the potential impact of the quality of care to the payer, I define the cost per quality ratio $\frac{\Omega^G}{\Omega^S p_H(e_H) p_P(e_P^*)}$. Therefore,
as the target price bundled payment model with gain-sharing agreements reduces the total spending ($\Omega^6$) compared to the FFS spending ($\Omega^5$), it will also reduce the probability of achieving the desired quality score (i.e., $p_H(e_H^*) p_P(e_P^*)$), which reduces the cost per quality ratio. Note that in the status quo FFS model, the desired quality score is reached with the probability equal to one. In Figure 4.11 I evaluate the impact of $b_p$ and the $\alpha$ on the cost per quality ratio.

![Figure 4.11](image)

**Figure 4.11**: Ratio of $\frac{\Omega^6}{\Omega^5 p_H(e_H^*) p_P(e_P^*)}$ for different values of $c_t$, $b_p$, and $\alpha$

Figure 4.11 shows that changing the FFS profit margin ($\alpha$) can have an important impact on the payer’s outcome. There is a trade-off for the payer when reducing $\alpha$: while it will incentivize the hospital and the provider to reduce the total treatment cost, it will also negatively impact the quality of care.

### 4.7 Extensions

#### 4.7.1 Separate cost and quality

In the CJR model the CMS aims to coordinate the care and reduce the total spending by imposing a target price bundled payment model on the hospital. However, imposing a separate cost and quality measure for each involved party might have a higher potential impact on the overall outcome. It is obvious that in such a setting hospital and the provider would not enter into gain-sharing agreements as they do not have any common
element in their profit functions. Thus, the hospital’s expected profit can be written as follows.

\[ \Pi_H^T(e_H) = \alpha c_H(e_H) + [c_{tH} - c_H(e_H)]^+ p_H(e_H) + [c_{tH} - c_H(e_H)]^- e_H^2 \]

Note that, both parties are evaluated based on their corresponding cost and quality. Similarly, the provider’s expected profit can be written as follows.

\[ \Pi_P^T(e_p) = \alpha c_p(e_p) + [c_{tp} - c_p(e_p)]^+ p_p(e_p) + [c_{tp} - c_p(e_p)]^- e_p^2 \]

In other to have comparable results with the main model I assume that \( c_t = c_{tH} + c_{tp} \). Figure 4.12 illustrates the total spending in each strategy.

![Figure 4.12: Total spending when \( \bar{c}_H = \bar{c}_P = 2; b_P = 2; \alpha = 5\% \)](image)

Figure 4.12 shows that imposing a separate target price on the hospital and the provider would be more effective in reducing the total spending when compared to a single target price imposed to the hospital. Moreover, this policy would even have a total spending lower than a centralized system in which the desired quality of care is evaluated as an aggregate outcome of both party’s treatment. This is because when evaluated separately each party is responsible for their own cost and quality that is independent of the outcome of the other care provider; however, when the quality of care measured based on the collective effort, each party’s optimal effort would be a function of the care that a patient received elsewhere.
4.7.2 Interdependency in cost functions

In the previous sections I assume that the cost functions of the hospital and the provider are independent. However, when hospital decreases its treatment cost the quality of care may also decrease. As a result, the health condition of a patient that arrives at the provider depends on the services that she received at the hospital. Therefore, cost-reduction efforts at the hospital may negatively impact the treatment cost at the provider.

To incorporate such dependency, I assume that the provider’s cost function is an increasing function in the hospital’s cost-reduction effort such that $c_p(e_p, e_H) = \bar{c}_p - b_p e_p + \theta e_H$ where $0 \leq \theta \leq 1$ is the degree of dependency. The analysis in the previous sections are for the special case where $\theta = 0$. In this section, I analyze the impact of $\theta$ on the gain sharing agreements and its implications for the payer.

Intuitively, the negative impact of the hospital’s cost-reduction effort on the provider’s cost and as a result on the total expenditure would disincentives the hospital from exerting in-house cost-reduction effort. Thus, depending on the degree of dependency ($\theta$) the hospital would use the gain-sharing agreements for a wider set of parameters and rely on provider’s cost-reduction efforts to reduce total treatment cost. Figure 4.13 illustrates the optimal solution in a decentralized system with gain-sharing agreements when $\theta = 0.5$.

Figure 4.13: Optimal solution in decentralized system with gain-sharing agreement when $\bar{c}_H = \bar{c}_p = 2; b_p = 2; \alpha = 5\%; \theta = 0.5$
Figure 4.13 shows that $\theta$ does not have any impact on the penalty and zero savings regions as in those regions hospital does not benefit from any cost-reduction effort even when $\theta = 0$. However, when compared to Figure 4.7, it is clear how the interdependency of the provider’s cost function would result in extended use of gain-sharing agreements.

To understand the impact of the degree of dependency in the cost functions on the payer’s total spending, I show how the ratio $\frac{\Omega^G}{\Omega^S}$ for different values of $\theta$ in Figure 4.14.

As shown in Figure 4.14 the degree of dependency $\theta$, does not have any impact on the outcome when the optimal in-house cost-reduction effort for the hospital is zero (small and large target price). However, when the target price is intermediate and the hospital requires to have some in-house cost savings, a system with lower $\theta$ is always preferred by all parties.

### 4.8 Conclusion

In this chapter, I examine the structure of gain-sharing agreements between a hospital and a provider that provide care for a patient’s condition. Following the general framework of the CJR payment model introduced by CMS, I investigate the effect of gain-sharing
agreements on the performance of a target price bundled payment model where the hospital is responsible for meeting a threshold on the total treatment cost. Both the hospital and the provider can exert cost-reduction effort to reduce their respective cost of treatment, which will, in turn, negatively impact the quality of care.

My research provides practical guidance for the payer regarding the use of such target price bundled payment models to coordinate care. I show that using gain-sharing agreements can be a “win-win-win” scenario for all parties under the target price bundled payment model. However, the choice of the target price is particularly important to incentivize the hospital to enter into a gain-sharing agreement with the provider. The target price should be low enough such that, to be eligible for reward payment, the hospital would need the provider to reduce its cost. I also provide evidence that a target price bundled payment model can be more effective in healthcare settings where the care providers have asymmetric capabilities.

My study is based on a stylized mathematical model with several limiting assumptions, and extending some of these assumptions can lead to interesting future work. I assume that patients receive the required care from a hospital and a single provider; yet, as is the case in joint and hip replacement procedures, there might be multiple providers involved in patients’ treatment. Therefore, a fruitful avenue of research might be to examine care-coordination in multi-agent pay-for-performance agreements. I also assume that the cost functions are public information and the provider can observe the hospital’s cost-reduction effort. A more robust modeling approach would be to model this problem as a multi-agent system with information asymmetry, where each party’s cost function is private information.

4.9 References


Chapter 5

5 Conclusion

In this thesis, I examine incentive problems in different environments, such as a decentralized supply chain setting in a manufacturing industry and a fragmented healthcare system. In the supply chain setting, I extend the literature by examining the impact of capacity on the optimal channel structure when a contract manufacturer can compete in both the component and final-product markets. I use a game theoretical approach to study the dynamics of firms’ strategic decisions under different capacity levels and market conditions.

In the healthcare setting, I examine optimal design of contracts to better incentivize parties with conflicting interests in order to achieve an improved outcome in the delivery of modern healthcare services. In Chapter 3, I model the interaction between a payer and a provider when an optional diagnostic test can be used to predict the benefits of an advanced treatment option. I design performance-based contracts to align the incentives of the payer and the provider to achieve optimal use of the optional diagnostic test. I introduce a new payment scheme in which the reward payments are tied to the treatment decision of the provider, as well as the health outcome of patients. Further, I model the imperfect information flow in fragmented healthcare settings where the provider has private information about patients’ characteristics (adverse selection) and about whether or not the optimal course of action is used (moral hazard).

In Chapter 4, I examine the incentives of a hospital and a provider to reduce their respective procedure costs when providing care for a patient who requires treatment from both the hospital and the provider. This problem is particularly challenging, as the cost-reduction efforts of the hospital and the provider impact both the procedure cost and quality of care. I investigate the effects of different payment structures—such as FFS, bundled payment, and bundled payment with gain-sharing model—on the cost-reduction efforts and profits of the hospital and the provider. I provide guidance, for the payer, regarding the impact of these models on the health outcome of patients and the total spending.
5.1 Managerial Insights

The results of my study in Chapter 2 provide strategic insights for OBM's facing competition from their supplier with limited capacity. I show that sharing capacity through supply chain partnership may result in a more profitable industry, rather than a price war. Such insight would increase the desirability of cooperating with competitors when there is a potential to achieve economies of scale through increased capacity levels. I demonstrate that the OBM only dual-sources its component demand when competition in the final-product market is very intense, such that the OBM buys out the CCM's capacity while outsourcing the rest of the component demand to a third-party supplier. I provide evidence for different resource allocation strategies of a CCM, and present an explanation for both prioritizing capacity to a firm's own product and reducing the firm's own product output to shift profits to component sales to a competitor.

Chapter 3 provides evidence for the necessity of developing alternative payment models as the availability of personalized medicine and the number of treatment options increases. Furthermore, I highlight the importance of information flow in the fragmented systems by showing that in a system with information asymmetry, it is never optimal to use a diagnostic test for patients from all risk groups, even when the test can be available for free. By showing how an increase in reputational concerns and altruistic behavior of a provider may increase or decrease social welfare, I provide further support for the inconsistent evidence regarding the impact of malpractice pressure on healthcare systems. Finally, my findings suggest that it is not always optimal to make a diagnostic test compulsory, even if it is recommended by the guidelines.

Chapter 4 provides guidelines for policymakers regarding the use of target price bundled payment models to coordinate care in fragmented healthcare systems. I show that such payment models can be more effective in healthcare settings where there is more asymmetry between the cost efficiency of the parties involved in the treatment of patients. In addition, I show that the decision on the target price is essential in incentivizing the hospital to arrange gain-sharing agreements because when the target price is low enough, the hospital needs the provider to exert effort to reduce the total treatment cost (as it is more efficient for the provider to reduce the cost of its procedure).
However, if the target price is not set properly, despite the fact that it is easier for a provider to reduce its cost, the hospital might exert more cost-reduction effort than the provider because the hospital would rather exert some effort in-house than share more of the reward with the provider.

While each essay focuses on different settings they share a common insight regarding the fragmented systems with conflicting interests. In all essays, I show that the inefficiency caused by the decentralization may cause firms not to use some resources even if they are efficient or available for free. In the first essay, the CCM might not utilize free capacity. In the second essay, the test should not be used for patients from all risk groups. In the third essay, the hospital may exert costly effort instead of benefiting from the provider’s cost-reduction efficiency.

5.2 Future Research

The studies in this thesis are based on a stylized mathematical models with several limiting assumptions. Therefore, generalization of the underlying problems would be a fruitful avenue for future research on incentive problems of parties with potential conflict of interest in decentralized systems. For example, in a capacitated supply chain setting, incorporating demand uncertainty in a supply chain with multiple strategic contract manufacturers, each with the option of producing their own brand products can lead to interesting future work.

In most circumstances patients have multiple options when deciding on which healthcare provider to visit. Thus, considering the impact of competition among providers would be an interesting research to better understand the impact of provider’s reputational concerns on his test ordering behavior. Furthermore, most of the procedures in fragmented health care systems, as in joint and hip replacement procedures, require patients to receive treatment from multiple providers. Therefore, studying multi-agent pay-for-performance agreements would be an interesting generalization to the care-coordination problem studied in this thesis.
Appendices

Appendix-A

In this section I derive the equilibrium solution for the 3-stage game defined in Section 2.3. From equation (2.1): $D_O = \frac{1}{2} (2 + M - 3p_O + p_C)$; $D_C = \frac{1}{2} (2 + M - 3p_C + p_O)$. I only consider the region where $0 < M < 2 - p_O - p_C$ such that the two products are competing.

1 - Stage 3 of the game

Both firms’ profit function is concave in its price (i.e., $\frac{d^2 \pi_i}{d p_i^2} < 0$), so FOC gives:

$$p_O^\ast(p_C) = \frac{1}{6} (2 + M + p_C + 3w_C \gamma + 3w_T(1 - \gamma)).$$

For firm C, using KKT conditions to consider the capacity constraint, I have two optimal pricing strategies: 1) Non-binding capacity constraint: $p_C^\ast(p_O) = \frac{1}{6} (2 + M + p_O + w_C \gamma)$; 2) Binding capacity constraint:

$$p_C^\ast(p_O) = \frac{2 - 2k + M + p_O + (2 + M - 3p_O) \gamma}{3 - \gamma}.$$ 

I solve for the optimal prices under each capacity condition: 

Case 1) Non-binding capacity constraint: $p_O^\ast = \frac{1}{35} (14 + 7M + 19w_C \gamma + 18w_T(1 - \gamma)), p_C^\ast = \frac{1}{35} (14 + 7M + 9w_C \gamma + 3w_T(1 - \gamma));$ For case 1, I need $\gamma D_O + D_C < k \iff \frac{1}{70} (9w_T + 42(1 + \gamma) + 21M(1 + \gamma) + \gamma(51w_T \gamma - 60w_T - 8w_C(1 + 6\gamma))) < k$. Case 2) Binding capacity constraint: $p_O^\ast = \frac{8 - 2k + 4M + 3(3 - \gamma)(w_C \gamma + w_T(1 - \gamma))}{17 - 3\gamma}$, $p_C^\ast = \frac{7(2 + M) + 3(2 + M + w_C - 4w_T) \gamma - 12k - 9(w_C - w_T) \gamma^2 + 3w_T}{17 - 3\gamma}$.

2 - Stage 2 of the game

Case 1) Firm O’s profit function is strictly convex (i.e., $\frac{d^2 \pi_O}{d \gamma^2} > 0$) which implies that optimal allocation ratio $\gamma^\ast$ is an extreme point. Because in case 1 firm C’s capacity is non-binding the extreme point is either zero or one (i.e., $\gamma^\ast \in \{0, 1\}$). For firm O to set a non-zero ratio (i.e., $\gamma^\ast = 1$), the resulting profit should be higher than its profit when $\gamma = 0$ (i.e., $\frac{3(14 + 7M - 17w_T)^2}{2450} \leq \frac{3(14 + 7M - 16w_C)^2}{2450} \iff 0 < w_T \leq \frac{2(2 + M)}{17}$ and $0 \leq w_C \leq$
firm O will outsource all its component demand to firm C.

Case 2) Firm O’s profit function is neither convex nor concave. Nevertheless, there is only one root to the FOC condition \( \left( \frac{d^2 \pi_0}{d \gamma^2} \right) = 0 \rightarrow \gamma = \frac{4 - k + 2M - 4w_t}{4(w_c - w_T)} \) which is a minimizer (\( \frac{d^2 \pi_0}{d \gamma^2} \mid \gamma = \frac{4 - k + 2M - 4w_t}{4(w_c - w_T)} \geq 0 \)). However, depending on the values of the parameters, this critical point may not be in the feasible region \( 0 \leq \gamma \leq 1 \). Though, since there is only one stationary point, I can claim that the profit function is maximized at an extreme point: 1) if the stationary point is in the feasible region, or 2) if the stationary point is not in the feasible region and the profit function is a decreasing/increasing function of \( \gamma \) in the feasible region. Therefore, for firm O to set a non-zero ratio the resulting profit should be higher than the profit when \( \gamma = 0 \) (i.e., \( \frac{6}{289}(k - 4 - 2M + 4w_T)^2 \leq \frac{6(k - 4 - 2M + 4w_T + 4\gamma(w_c - w_T) - (17 - 3\gamma)^2)}{(17 - 3\gamma)^2} \Leftrightarrow (0 \leq w_c \leq \frac{1}{68}(12 - 3k + 6M) \quad and \quad 0 \leq w_T \leq \frac{136 - 34k + 68M - 12\gamma + 3ky - 6My - 68w_c\gamma}{136 - 80\gamma} \quad OR \quad \left( \frac{1}{68}(12 - 3k + 6M) < w_c < \frac{1}{4}(4 - k + 2M) \right) \quad and \quad \frac{1}{56}(3k - 12 - 6M + 68w_c) \leq w_T \leq \frac{136 - 34k + 68M - 12\gamma + 3ky - 6My - 68w_c\gamma}{136 - 80\gamma}). \)

In the binding capacity case, in order to have non-negative prices and profits, firm C’s capacity has to be between \( 0 \leq k \leq 4 + 2M \). Because \( \gamma^* \) is always and extreme point multi-sourcing can only happen when firm C does not have enough capacity to fulfill firm O’s component demand in which case firm O is the monopoly in the final-product market and firm C acts as the supplier only. Consequently, I separate the game into three competing strategies: 1) \( \gamma^* = 1 \), Coopetition (Superscript C); 2) \( \gamma^* = 0 \), Competition (Superscript T); and 3) \( 0 < \gamma^* \leq 1 \), Supplier only (Superscript S).

I can express the optimal prices and capacity conditions for each strategy as follows. The optimal prices for the coopetition strategy (C) are: Case 1) \( p_0^c^* = \frac{1}{35}(14 + 7M + 19w_c) \), \( p_c^* = \frac{1}{35}(14 + 7M + 9w_c) \) \iff \( k > \frac{1}{5}(6 + 3M - 4w_c) \); Case 2) \( p_0^c^* = \frac{1}{7}(4 - k + 2M + 3w_c), p_c^* = \frac{1}{7}(10 - 6k + 5M - 3w_c) \) \iff \( k \leq \frac{1}{5}(6 + 3M - 4w_c) \).
prices for the competition strategy (T) are: Case 1) $p^*_O(T) = \frac{1}{35}(18w_T + 7M + 14)$, $p^*_C(T) = \frac{1}{35}(3w_T + 7M + 14)$ \(\text{iff} k > \frac{3}{70}(14 + 3w_T + 7M) \text{ and } 0 \leq w_T \leq \frac{7(2+M)}{17} \); Case 2) $p^*_O(T) = \frac{1}{17}(9w_T - 2k + 4M + 8)$, $p^*_C(T) = \frac{1}{17}(3w_T - 12k + 7M + 14)$ \(\text{iff} k \leq \frac{3}{70}(14 + 3w_T + 7M) \text{ and } 0 \leq w_T \leq \frac{1}{4}(4 - k + 2M)\). The limits on $w_T$ ensures that firm T’s component is competitive in the component market – that is firm O would get a non-negative profit if it uses firm T’s component. When firm O buys out firm C’s component capacity, firm C will have the supplier only role. In this case, firm O is the monopoly in the final product market that is only feasible when $0 \leq k \leq 1$ and $0 \leq w_T \leq 1 - k$ with $D^*_O = 2 - 2p^*_O(S)$, $p^*_O(S) = \frac{1}{2}(1 + w_C\gamma + w_T(1 - \gamma))$. Firm O’s allocation ratio in the supplier only strategy is such that firm O buys all firm C’s capacity ($\gamma D_O = k \iff \gamma^* = \frac{1-w_T+\sqrt{(1-w_T)^2-4k(w_C-w_T)}}{2(w_C-w_T)}$) and if necessary firm O multi-sources the rest of its component demand from firm T.

3 - Stage 1 of the game

In the equilibrium, firm C maximizes its profit function by choosing its wholesale price for each strategy that it acts as the supplier (i.e., S and C) subject to firm O’s incentive compatibility and participation constraints. The incentive compatibility constraint ensures that firm O would not deviate to other strategies given firm C’s wholesale price. The participation constraint makes sure that firm O gets at least as much as its outside option strategy T. Finally, knowing the best outcome of each strategy, firm C chooses the equilibrium strategy with its wholesale price for each capacity $k$, competition level $M$, and firm T’s wholesale price $w_T$.

Knowing the optimal set of prices in each strategy, firm C’s optimization problem for the competition scenario with binding capacity constraint pricing strategy (case 2, represented by superscript $B$) is:
Superscript \(N\) is used to represent nonbinding capacity case (case 1). I solve the problem using KKT conditions. Note that, the two IR constraints are exclusive and thus I solve two similar optimization problems for each. Firm C’s optimal wholesale prices are such that firm O’s participation constraint binds. Therefore, these wholesale prices represent the maximum wholesale price before firm O outsources to firm T (\(\bar{w}_C\)). The optimal wholesale price for collectively exhaustive and individually exclusive regions in case 2 of coopetition scenario are:

\[
\bar{w}_C = \frac{1}{10} (10 - 6k + 5M) \iff \left( \frac{49(2+M)}{179} < w_T \leq \frac{7(2+M)}{17} \quad \text{and} \quad \frac{20}{27} \left(2 + M - 2w_T\right) \leq k \leq \frac{1}{70} (21M + 9w_T + 42) \right) \quad \text{OR} \quad \left( \frac{7(2+M)}{17} < w_T \leq \frac{2+M}{2} \quad \text{and} \quad \frac{20}{27} \left(2 + M - 2w_T\right) \leq k \leq 2M + 4(1 - w_T) \right) \quad \text{OR} \left( \frac{56(2+M)}{221} \leq w_T \leq \frac{49(2+M)}{179} \quad \text{and} \quad \frac{1}{7} (7M - 17w_T + 14) \leq k \leq \frac{1}{13} (10 + 5M) \right) \quad \text{OR} \left( \frac{49(2+M)}{179} < w_T \leq \frac{7(2+M)}{17} \quad \text{and} \quad \frac{1}{70} (21M + 9w_T + 42) \leq k \leq \frac{1}{13} (10 + 5M) \right) ;
\]

\[
\bar{w}_C = \frac{1}{68} (12 - 3k + 6M + 56w_T) \iff \left( 0 \leq w_T \leq \frac{49(2+M)}{179} \quad \text{and} \quad 0 \leq k \leq \frac{1}{70} (21M + 9w_T + 42) \right) \quad \text{OR} \left( \frac{49(2+M)}{179} < w_T \leq \frac{2+M}{2} \quad \text{and} \quad 0 \leq k \leq \frac{20}{27} \left(2 + M - 2w_T\right) \right) ;
\]
\( \bar{w}_c = \frac{1}{20} (6 - 5k + 3M + 17w_T) \iff \left( 0 \leq w_T < \frac{56(2+M)}{221} \text{ and } \frac{1}{70} (21M + 9w_T + 42) \leq k < \frac{1}{20} (24 + 12M - 17w_T) \right) \ OR \left( \frac{56(2+M)}{221} \leq w_T \leq \frac{49(2+M)}{179} \text{ and } \frac{1}{70} (21M + 9w_T + 42) \leq k < \frac{1}{7} (7M - 17w_T + 14) \right) ;

\( \bar{w}_c = \frac{1}{4} (6 - 5k + 3M) \iff \left( 0 \leq w_T \leq \frac{56(2+M)}{221} \text{ and } \frac{1}{20} (24 + 12M - 17w_T) \leq k < \frac{1}{6} (24 + 12M - 17w_T) \right) \ OR \left( \frac{56(2+M)}{221} < w_T \leq \frac{7(2+M)}{17} \text{ and } \frac{1}{13} (10 + 5M) \leq k < \frac{1}{6} (17w_T - 4 - 2M) \right) ;

Firm C’s optimization problem for non-binding case (case 1) of coopetition scenario is:

\[
\max_{\bar{w}_c} \pi^C_{CN} \\
\text{s.t.} \quad \pi^C_{CN} \geq \pi^T_{O} \quad \text{if} \quad k \leq \frac{3}{70} (14 + 3w_T + 7M) \text{ and } 0 \leq w_T \leq \frac{1}{4} (4 - k + 2M) \quad (IR) \\
\pi^C_{CN} \geq \pi^T_{O} \quad \text{if} \quad k > \frac{3}{70} (14 + 3w_T + 7M) \text{ and } 0 \leq w_T \leq \frac{7(2+M)}{17} \quad (IR) \\
\pi^C_{CN} \geq \pi^S_{O} \quad \text{if} \quad 0 \leq k \leq 1 \text{ and } 0 \leq w_T \leq 1 - k \quad (IC) \\
\text{andr} \quad k > \frac{1}{5} (6 + 3M - 4\bar{w}_C)
\]

KKT conditions would result in collectively exhaustive and individually exclusive regions with optimal wholesale prices as follows.

\( \bar{w}_c = \frac{217(2+M)}{876} \iff \left( \frac{868(2+M)}{3723} \leq w_T \leq \frac{7(2+M)}{17} \text{ and } k > \frac{1}{219} (176 + 88M) \right) ;

\( \bar{w}_c = \frac{17w_T}{16} \iff \left( 0 \leq w_T < \frac{868(2+M)}{3723} \text{ and } k > \frac{1}{20} (24 + 12M - 17w_T) \right) .

\textbf{Lemma A1:} The maximum acceptable wholesale price by firm O such that it is better off to outsource to firm C weakly decreases in capacity of firm C (i.e., \( \frac{d\bar{w}_C}{dk} \leq 0 \)).
Proof of Lemma 2.1, Propositions 2.1, and Proposition 2.4: The Equilibrium Solution

Despite the fact that each case (binding and non-binding) has wholesale prices for exclusive regions there are overlaps between two scenarios’ feasible regions. Firm C chooses the scenario with higher profits considering firm O’s incentive compatibility constraint. For example, if firm C prefers binding capacity case it has to make sure that \( \{ \pi^B_C \geq \pi^N_C | w_C \} \). Moreover, firm C can choose to set wholesale price high enough so to deter firm O from outsourcing to firm C (violating IR constraints). Firm C would set high wholesale prices when firm Cs profit from competition strategy is more than its profit from coopetition strategy (i.e., \( \pi^T_C \geq \pi^C_C \)). Replacing the optimal pricing and wholesale price of each region I can show that without considering the supplier only strategy, firms’ optimal outsourcing strategies is as follows.

**Competition Strategy: (outsource to firm T)**

**I. Binding Capacity Case**

\[ 0 \leq w_T < \frac{2 + M}{11} \text{ and } 0 \leq k < \frac{1}{70} (42 + 21M + 9w_T) \]

\[ \text{OR} \left( \frac{2 + M}{11} < w_T < \frac{2 + M}{2} \text{ and } 0 \leq k < \frac{8}{21} (2 + M - 2w_T) \right) \]

**II. Non-Binding Capacity Case**

\[ 0 \leq w_T < \frac{2 + M}{11} \text{ and } \frac{1}{70} (42 + 21M + 9w_T) \leq k < \frac{1}{105} \left( 56(2 + M) - 51w_T - \sqrt{49(2 + M)^2 - 10566w_T^2 + 4242w_T(2 + M)} \right) \]

**Coopetition Strategy: (outsource to firm C)**

**III. Binding Capacity Case**

\[ \left( \frac{49(2 + M)}{179} < w_T \leq \frac{7(2 + M)}{17} \text{ and } \frac{20}{27} (2 + M - 2w_T) \leq k \leq \frac{1}{70} (42 + 21M + 9w_T) \right) \]

\[ \text{OR} \left( \frac{7(2 + M)}{17} < w_T \leq \frac{2 + M}{2} \text{ and } \frac{20}{27} (2 + M - 2w_T) \leq k \leq 2M + 4(1 - \right) \]
\[ \left( \frac{56(2+M)}{221} \leq w_T \leq \frac{49(2+M)}{179} \text{ and } \frac{1}{7} (7M - 17w_T + 14) \leq k \leq \frac{1}{13} (10 + 5M) \right) \]

\[ \Rightarrow w_c = \frac{1}{10} (10 - 6k + 5M) \]

\[ \left( 0 \leq w_T < \frac{2+M}{11} \text{ and } \frac{1}{105} (56(2 + M) - 51w_T - \sqrt{49(2 + M)^2 - 10566w_T^2 + 4242w_T(2 + M)}) \leq k \leq \frac{1}{20} (24 + 12M - 17w_T) \right) \]

\[ \Rightarrow w_c = \frac{1}{20} (6 - 5k + 3M + 17w_T) \]

\[ \left( \frac{868(2+M)}{3723} < w_T \leq \frac{56(2+M)}{221} \text{ and } \frac{1}{20} (24 + 12M - 17w_T) \leq k \leq \frac{1}{219} (176 + 88M) \right) \]

\[ \Rightarrow w_c = \frac{1}{4} (6 - 5k + 3M) \]

\[ \left( \frac{2+M}{11} \leq w_T \leq \frac{49(2+M)}{179} \text{ and } \frac{8}{21} (2 + M - 2w_T) \leq k \leq \frac{1}{70} (42 + 21M + 9w_T) \right) \]

\[ \Rightarrow w_c = \frac{1}{68} (12 - 3k + 6M + 56w_T) \]

**IV. Non – Binding Capacity Case**

\[ \frac{868(2 + M)}{3723} \leq w_T \leq \frac{7(2 + M)}{17} \text{ and } k > \frac{1}{219} (176 + 88M) \Rightarrow w_c = \frac{217(2 + M)}{876} \]

\[ 0 \leq w_T < \frac{868(2 + M)}{3723} \text{ and } k > \frac{1}{20} (24 + 12M - 17w_T) \Rightarrow w_c = \frac{17w_T}{16} \]
Figure 2.2 illustrates the four-region strategies that are outlined above at $M = 1$ where $\bar{w}_T = \frac{2+M}{11}$ is the maximum wholesale price of firm T in Region II. For expositional convenience define capacity thresholds that separate the four regions:

$$k_{12} \equiv \frac{3}{70} (14 + 3w_T + 7M)$$

$$k_{23} \equiv$$

$$\begin{cases} 
\frac{1}{105} (56(2 + M) - 51w_T - \sqrt{49(2 + M)^2 - 10566w_T^2 + 4242w_T(2 + M)}) \text{ if } 0 \leq w_T < \frac{2+M}{11} \\
\frac{8}{21} (2 + M - 2w_T) \text{ if } \frac{2+M}{11} \leq w_T < \frac{2+M}{2}
\end{cases}$$

$$k_{34} \equiv$$

$$\begin{cases} 
\frac{1}{20} (24 + 12M - 17w_T) \text{ if } 0 \leq w_T < \frac{868(2+M)}{3723} \\
\frac{1}{219} (176 + 88M) \text{ if } \frac{868(2+M)}{3723} \leq w_T \leq \frac{7(2+M)}{17}
\end{cases}$$

Next, I analyze supplier only option of firm C. Firm C’s profit function in supplier only strategy is an increasing function in wholesale price (i.e., $\pi_C = w_C D_0$); therefore, firm C would choose a wholesale price that is incentive compatible with firm O.

$$\max_{w_C} \pi_C^S = w_C D_0$$

s.t.  $\pi_C^S \geq \pi_O^{TB}$ if $k \leq \frac{3}{70} (14 + 3w_T + 7M)$ and $0 \leq w_T \leq \frac{1}{4} (4 - k + 2M)$ (IR)

$$\pi_C^S \geq \pi_O^{TN}$ if $k > \frac{3}{70} (14 + 3w_T + 7M)$ and $0 \leq w_T \leq \frac{7(2+M)}{17}$ (IR)

$$\pi_C^S \geq \pi_O^{CB}$ if $k \leq \frac{1}{5} (6 + 3M - 4w_C)$ (IC)

$$\pi_C^S \geq \pi_O^{CN}$ if $k > \frac{1}{5} (6 + 3M - 4w_C)$ (IC)

$0 \leq k \leq 1$ and $0 \leq w_T \leq 1 - k$

In the above maximization problem the two IC (IR) constraints are exclusive and do not appear together. First, I find the minimum wholesale price that is required for firm C to be better off being supplier only compared to its profit in the four-region strategies.
outlined above. Then, I find the wholesale price range, for which IR and IC constraints are satisfied while keeping wholesale price higher than the minimum required for firm C to prefer supplier only strategy. Because firm O’s profit in the supplier only strategy is a decreasing function in wholesale price, the wholesale price that maximizes firm C’s profit in the supplier only strategy will force either the IC or IR constraint to bind. Therefore, I find the wholesale price that binds each constraint. Then, whichever wholesale price that is smaller would be the maximum wholesale price that firm C can charge in the supplier only strategy for its corresponding region depending on $k, w_T$ and $M$. If this maximum acceptable wholesale price is larger than the minimum required wholesale price for firm C to prefer the supplier only strategy, firm C would apply the maximum acceptable wholesale price and thus incentivize firm O to buyout firm C’s capacity (supplier only strategy). For illustration purposes, I show the analysis only for the Region IV where $w_C = \frac{217(2+M)}{876}$. In this region the optimal strategy, without consideration of supplier only strategy, is coopetition strategy with non-binding capacity constraint. Therefore, firm C, in order to prefer the supplier only strategy, should have

$$\pi_C^S \geq \pi_C^{CN}_{|w_C=\frac{217(2+M)}{876}} \iff w_C D_O \geq \frac{91}{876} (2 + M)^2;$$

this would result in a minimum required wholesale price of $w_C \geq \frac{364+364M+91M^2}{876k}$ for firm C to prefer the supplier only strategy to coopetition strategy. Because in this region (i.e., $\frac{868(2+M)}{3723} \leq w_T \leq \frac{7(2+M)}{17}$ and $k > \frac{1}{219} (176 + 88M)$) capacity is always larger than $k > \frac{1}{70} (42 + 21M + 9w_T)$) I only need to consider $\pi_C^S \geq \pi_C^{TN} (IR)$. Moreover, with a minimum wholesale price of $w \geq \frac{364+364M+91M^2}{876k}$ the capacity in this region is always larger than $k > \frac{1}{5} (6 + 3M - 4w_C)$ which leads us to only consider $\pi_C^S \geq \pi_C^{CN} (IC)$. As mentioned, I only need to consider the wholesale prices that make the IC and IR constraint binding.

$$\pi_C^S = \pi_C^{TN} \rightarrow w_C = \frac{1}{2450k} (49(25(2-k)k - 3(2 + M)^2) + 714(2 + M)w_T - 867w_T^2);$$

$$\pi_C^S = \pi_C^{CN} \rightarrow \frac{7}{768} (96 - 175k + 48M + 5\sqrt{192k + 457k^2} - 672kM) (I \text{ only consider the positive root}).$$
In the sequential equilibrium the supplier only strategy is optimal if and only if
\[\min\left\{ \frac{1}{2450k} (49(25(2 - k)k - 3(2 + M)^2) + 714(2 + M)w_T - 867w_T^2), \frac{7}{768} (96 - 175k + 48M + 5\sqrt{192k + 457k^2 - 672kM}) \right\} \geq \frac{364 + 364M + 91M^2}{876k}; \]
That is
\[
0 < M \leq \frac{1922}{37583} \quad \text{and} \quad \frac{868(2+M)}{3723} \leq w_T < \frac{7(-1222-611M+5\sqrt{168410-147059M+2+M})}{22338} \quad \text{and} \quad \frac{176+88M}{219} < k < \frac{1}{560} \left( 560 - 595w_T + \sqrt{12544(13 - 3M(4 + M)) - 3808(79 - 48M)w_T + 132073w_T^2} \right)
\]
\[
\left\{ \frac{1922}{37583} < M \leq \frac{14527654}{225927373} \quad \text{and} \quad \frac{868(2+M)}{3723} \leq w_T \leq \frac{7(2628+1314M+5\sqrt{6254-35417M+2+M})}{22338} \quad \text{and} \quad 1 - \right.
\]
\[
\frac{1}{35} \sqrt{102(20 - 7M)w_T - 867w_T^2 - \frac{1}{438} (434950 + 7M(11140 + 25123M))} \leq k < \frac{1}{560} \left( 560 - 595w_T + \sqrt{12544(13 - 3M(4 + M)) - 3808(79 - 48M)w_T + 132073w_T^2} \right)
\]
\[
\left. \frac{1922}{37583} < M \leq \frac{14527654}{225927373} \quad \text{and} \quad \frac{7(2628+1314M+5\sqrt{6254-35417M+2+M})}{22338} < w_T < \right.
\]
\[
\frac{7(-1222-611M+5\sqrt{168410-147059M+2+M})}{22338} \quad \text{and} \quad \frac{176+88M}{219} < k < \frac{1}{560} \left( 560 - 595w_T + \sqrt{12544(13 - 3M(4 + M)) - 3808(79 - 48M)w_T + 132073w_T^2} \right)
\]

The analysis for the rest of the regions is omitted due to redundancy. Figure 2.5 shows the final equilibrium when product substitutability is \( M = 0.01 \). Regions V and VI are the supplier only strategy region. From the derivation of the above equilibrium I find that the supplier only strategy is only optimal in the sequential equilibrium if and only if \( M \leq \)
\[
\bar{M} = \frac{17}{48} \sqrt{48(1-w_T)^2 + 97k^2 - 2(1-w_T) - \frac{95k}{48}}.
\]
Otherwise, the four-region equilibrium presented above is the equilibrium solution of the game. Consequently, I claim that the four-region equilibrium conditions presented above is the unique equilibrium of this sequential game when \( M > \bar{M} \) and Proposition 2.1 is a different representation of the above-mentioned equilibrium regions. Table A.1 presents the optimal prices, demands and profits of the firms in the four-region equilibrium when \( w_T \leq \bar{w}_T = \frac{2+M}{11} \).

**Proof of Proposition 2.2: Sensitivity Analysis**

**a.** From Table A.1 I can replace the corresponding demands: \( D_C^*(k_{23}^-) > D_C^*(k_{23}^+) \)
\[
\Rightarrow \frac{3}{70} (14 + 7M + 3w_T) > k - \frac{3}{70} (14 + 7M - 17w_T). \]
This inequality is always true as long as \( k < k_{34} \), which is always larger than \( k_{23} \). Since firm O’s demand does not change around \( k_{23} \), the second expression (i.e., \( D_O^*(k_{23}^-) + D_C^*(k_{23}^-) > D_O^*(k_{23}^+) + D_C^*(k_{23}^+) \)) is the immediate result of the previous one.

**b.** This result is immediate from Table A.1 as long as \( k < k_{34} \) which is always larger than \( k_{23} \).

**c.** Profit function of firm C in Regions I and III is a concave function in its capacity where \( k_1 \) and \( k_3 \) are the maximizers of its profit in these regions respectively. These maximum points are situated within the regions (e.g., \( k_{23} \leq k_3 \leq k_{34} \)). Therefore, firm C’s profit will decrease beyond these points: \( \text{if } k_3 < k < k_{34} \Rightarrow \frac{d\pi_C}{dk} = \frac{1}{980} (14(56(2 + M) - 51w_T) - 1470k) < 0; \text{ if } k_1 < k < k_{12} \Rightarrow \frac{d\pi_C}{dk} = \frac{1}{17} (14 - 24k + 7M + 3w_T) < 0. \) Moreover, I show that the absolute value of the slope in Region III is larger than that of Region I:
\[
\left| \frac{d\pi_C}{dk} \right|_{k=k_1+\Delta} < \left| \frac{d\pi_C}{dk} \right|_{k=k_3+\Delta} \Rightarrow \left| \frac{1}{980} (14(56(2 + M) - 51w_T) - 1470k) \right|_{k=k_3+\Delta} < \left| \frac{1}{17} (14 - 24k + 7M + 3w_T) \right|_{k=k_3+\Delta}.
\]
Proof of Proposition 2.3: Sensitivity Analysis

This proposition is an immediate result of Table A.1.

Derivation of equilibrium when there is no competition in component market

In this section I investigate the equilibrium conditions of the scenario presented in Section 2.5 where firm C is the proprietary component manufacturer. I use the same logic as the main model to derive the equilibrium. The optimal prices are: Case 1) \( p_o^* = \frac{1}{35} (14 + 7M + 19w_C) \), \( p_c^* = \frac{1}{35} (14 + 7M + 9w_C) \) if and only if \( \frac{1}{5} (6 + 3M - 4w_C) < k \) (Condition 1); Case 2) \( p_o^* = \frac{1}{7} (4 - k + 2M + 3w_C), p_c^* = \frac{1}{7} (10 - 6k + 5M - 3w_C) \).

Knowing the best response optimal prices firm C will choose the wholesale price: Case 1) Firm C’s profit function is strictly concave \( \left( \frac{d^2 \pi_C}{d w_C^2} > 0 \right) \) and has a maximum \( w_C = \frac{217(2+M)}{876} \) at \( \frac{d \pi_C}{d w_C} = 0 \); Case 2) Firm C’s profit function is strictly concave \( \left( \frac{d^2 \pi_C}{d w_C^2} > 0 \right) \) and has a maximum \( w_C = \frac{1}{10} (10 - 6k + 5M) \) at \( \frac{d \pi_C}{d w_C} = 0 \). Note that, in binding case firm C’s capacity has to be small enough so that firm C can use all of its capacity while maintaining positive prices and profits (i.e., \( 0 \leq k \leq \frac{1}{6} (10 + 5M) \)).

Similar to the main model, there is an overlap between the capacity conditions. Since firm C is the first mover, considering the incentive compatibility conditions of firm O, it decides which of the optimal wholesale prices to choose when capacity is in the overlapping region. Consequently, firm C’s optimal action in each region is:

A) Binding Capacity Case: \( k \leq \frac{5}{6} + \frac{5M}{12} + \frac{1}{12} \sqrt{\frac{5}{73} (2 + M)} \Rightarrow w_C = \frac{1}{10} (10 - 6k + 5M) \); B) Non – Binding Capacity Case: \( k > \frac{5}{6} + \frac{5M}{12} + \frac{1}{12} \sqrt{\frac{5}{73} (2 + M)} \Rightarrow w_C = \frac{217(2+M)}{876} \).

Before claiming the above regions to be the equilibrium of the game I also consider a case where firm C chooses to be a monopoly. In this scenario, firm C faces a capacity constraint (i.e., \( D_C < k \)) and only decides on its market price. Solving for the optimal
price considering the price and demand non-negativity, I have: if \( k \geq 1 \Rightarrow p_c = \frac{1}{2} \)

\[ \Rightarrow \pi_c = \frac{1}{2} OR \text{ if } k < 1 \Rightarrow p_c = \frac{1}{2} + \frac{1}{2} \sqrt{1 - 2k + k^2} \Rightarrow \pi_c = \frac{1}{2} (2k - k^2). \]

Comparing with the cooperation scenario, I show that there are some cases that firm C prefers to be monopoly in the final product market and thus results in four different regions of equilibrium depending on firm C’s capacity and product substitutability \( M \):

**A. Binding Capacity Case**

1. Coopetition: if \( k \leq \frac{1}{876} (365 + \sqrt{365})(2 + M) \) and \( M \geq \frac{k}{5} \)

2. Monopoly:
   - if \( 0.9612 < k \leq 1 \) and \( M < \frac{1}{91} (\sqrt{39858\sqrt{(2 - k)k - 182}}) \)
   - if \( k \leq 0.9612 \) and \( M < \frac{k}{5} \)

**B. Non – Binding Capacity Case**

3. Coopetition:
   \[
   \begin{cases} 
   if \ k \geq 1 
   \quad \text{and } M > 0.1939 \\
   if \ \frac{1}{876} (365 + \sqrt{365})(2 + M) < k < 1 \quad \text{and } M \geq \frac{1}{91} (\sqrt{39858\sqrt{(2 - k)k - 182}}) 
   \end{cases}
   \]

4. Monopoly: if \( k \geq 1 \) and \( M \leq 0.1939 \)

Finally, I consider a case where firm C chooses to be supplier only. In this case, firm C chooses the wholesale price \( w_C \) and only after that firm O will decide on the market price of its product. Firm O’s profit function is strictly concave \( (\frac{d^2 \pi_o}{dp_o^2} > 0) \). So, FOC gives:

\( p_o^* = \frac{1 + w_c}{2} \). Then, firm C chooses its optimal wholesale price. Having a strictly concave profit function \( (\frac{d^2 \pi_p}{dw_c^2} > 0) \): if \( k \geq \frac{1}{2} \Rightarrow w_c = \frac{1}{2} \Rightarrow \pi_c = \frac{1}{4} OR \text{ if } k < \frac{1}{2} \Rightarrow w_c = \frac{1}{2} + \frac{1}{2} \sqrt{1 + 4(k^2 - k)} \Rightarrow \pi_c = k - k^2 \). Firm C’s profit when only a supplier is always dominated by its profit in the monopoly or coopetition scenarios presented in the four-region equilibrium above. Consequently, I claim that the four-region equilibrium is the unique equilibrium of this sequential game. Table A.2 presents the optimal prices, demands and profits of the firms in each region of the equilibrium.
Proof of Proposition 2.5: Value of competition

From Table A.1 and A.2 I can find $V_o$ for any given capacity $k$, product substitutability $M$ and firm $T$’s wholesale price $w_T$ (e.g., $Non-binding$ cooperation: $V_o = \left[\frac{3(14 + 7M - 17w_T)^2}{2450}\right] - \left[\frac{361}{31974}(2 + M)^2\right]$). There are in total 11 different values of competition outcomes for firm O depending on capacity level and product substitutability when $w_T \leq \hat{w}_T$. Knowing the $V_o(k, M, w_T)$ functions I can derive the results in Proposition 2.5. The results for firm C can be driven in the same way.

### Table A.1: Equilibrium profits, prices, and demands of the firms when $w_T \leq \hat{w}_T$

<table>
<thead>
<tr>
<th>Equilibrium Region</th>
<th>$w_c^*$</th>
<th>Firm C</th>
<th>Firm O</th>
</tr>
</thead>
<tbody>
<tr>
<td>I) Competition $\gamma^* = 0$ Binding $\bar{w} &lt; w_c &lt; 2w_c$</td>
<td>$\pi_c = \frac{k(14 - 12k + 7M + 3w_T)}{17}$</td>
<td>$\pi_o = \frac{6(k - 2M + 4w_T - 4)^2}{289}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p_c = \frac{14 - 12k + 7M + 3w_T}{17}$</td>
<td>$p_o = \frac{8 - 2k + 4M + 9w_T}{17}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$D_c = k$</td>
<td>$D_o = \frac{3(4k + 2M - 4w_T)}{17}$</td>
<td></td>
</tr>
<tr>
<td>II) Competition $\gamma^* = 0$ Non-Binding $\bar{w} &lt; w_c &lt; 2w_c$</td>
<td>$\pi_c = \frac{3(14 + 7M + 3w_T)^2}{2450}$</td>
<td>$\pi_o = \frac{3(17w_T - 7M - 14)^2}{2450}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p_c = \frac{14 + 7M + 3w_T}{35}$</td>
<td>$p_o = \frac{7M + 18w_T + 14}{35}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$D_c = \frac{3(14 + 7M + 3w_T)}{70}$</td>
<td>$D_o = \frac{3(7M - 17w_T + 14)}{70}$</td>
<td></td>
</tr>
<tr>
<td>III) Coopetition $\gamma^* = 1$ Binding $2w_c &lt; w_c &lt; 3w_c + 5 + \frac{1}{3}k$</td>
<td>$\pi_c = \frac{1}{980}(784k(2 + M) - 3(14 + 7M - 17w_T)^2 - 714k w_T - 735k^2)$</td>
<td>$\pi_o = \frac{3(17w_T - 7M - 14)^2}{2450}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p_c = \frac{91(2 + M) - 51w_T - 105k}{140}$</td>
<td>$p_o = \frac{98 - 35k + 49M + 51w_T}{140}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$D_c = \frac{70k - 21(2 + M) + 51w_T}{70}$</td>
<td>$D_o = \frac{3(7M - 17w_T + 14)}{70}$</td>
<td></td>
</tr>
<tr>
<td>IV) Coopetition $\gamma^* = 1$ Non-Binding $\frac{1}{16}w_c &lt; w_c &lt; \frac{1}{16}(2M + 3\bar{w} + 5 + \frac{1}{3}k)$</td>
<td>$\pi_c = \frac{4704(2 + M)^2 + 29512(2 + M)w_T - 63291w_T^2}{78400}$</td>
<td>$\pi_o = \frac{3(17w_T - 7M - 14)^2}{2450}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p_c = \frac{112(2 + M) + 153w_T}{560}$</td>
<td>$p_o = \frac{224 + 112M + 323w_T}{560}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$D_c = \frac{84 + 62M - 17w_T}{140}$</td>
<td>$D_o = \frac{3(7M - 17w_T + 14)}{70}$</td>
<td></td>
</tr>
</tbody>
</table>
Table A.2: Equilibrium profits, prices, and demands of the firms when there is no competition in the component market

<table>
<thead>
<tr>
<th>Equilibrium Region</th>
<th>Firm C</th>
<th>Firm O</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Coopetition Binding</td>
<td>(\pi_c = \frac{1}{10} \left(10k - 6k^2 + 5kM\right))</td>
<td>(\pi_0 = \frac{3k^2}{50})</td>
</tr>
<tr>
<td></td>
<td>(p_c = \frac{1}{10} \left(10 - 6k + 5M\right))</td>
<td>(p_0 = \frac{1}{10} \left(10 - 4k + 5M\right))</td>
</tr>
<tr>
<td></td>
<td>(D_c = \frac{7k}{10})</td>
<td>(D_0 = \frac{3k}{10})</td>
</tr>
<tr>
<td>2) Monopoly Binding</td>
<td>(\pi_c = \frac{1}{2} \left(2k - k^2\right))</td>
<td>(\pi_0 = 0)</td>
</tr>
<tr>
<td></td>
<td>(p_c = \frac{1}{2} + \frac{1}{2} \sqrt{1 - 2k + k^2})</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(D_c = k)</td>
<td>-</td>
</tr>
<tr>
<td>3) Coopetition Non—Binding</td>
<td>(\pi_c = \frac{91}{876} (2 + M)^2)</td>
<td>(\pi_0 = \frac{361}{31974} (2 + M)^2)</td>
</tr>
<tr>
<td></td>
<td>(p_c = \frac{77}{292} (2 + M))</td>
<td>(p_0 = \frac{293}{876} (2 + M))</td>
</tr>
<tr>
<td></td>
<td>(D_c = \frac{119}{438} (2 + M))</td>
<td>(D_0 = \frac{19}{146} (2 + M))</td>
</tr>
<tr>
<td>4) Monopoly Non—Binding</td>
<td>(\pi_c = \frac{1}{7})</td>
<td>(\pi_0 = 0)</td>
</tr>
<tr>
<td></td>
<td>(p_c = \frac{1}{2})</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(D_c = 1)</td>
<td>-</td>
</tr>
</tbody>
</table>
Appendix-B

**Proof of Lemma 3.1:** Applying the first-order condition (FOC) to the provider’s objective function gives \( \tau^*_i = \frac{c + b_{ni} - b_{li} - w_i}{ah + b_{ni} - kb_{li}} \), which is independent of patients’ risk distribution \( F_l(r) \). This function is concave at \( \tau^*_i \) if and only if \( \frac{d^2 g_i(\tau)}{d \tau^2} < 0 \Leftrightarrow 0 \leq b_{li} < \frac{ah + b_{ni}}{k} \); otherwise, \( \tau^*_i \) is the minimum and the provider would either treat or don’t treat patients (i.e., \( \tau^*_i \in \{0,1\} \)). \( \tau^*_i = \hat{\ell} \) follows from the proof of the Proposition 3.1. \( \Box \)

**Proof of Lemma 3.2:** Since \( r \in [0,1] \), I focus on \( 0 < \hat{\ell} < 1 \Leftrightarrow 0 < z < z \Leftrightarrow 0 \leq c < \bar{c} \) where \( \hat{\ell} = \frac{z + c}{(1 + a)h + (1 - k)(\lambda_H - \lambda_L)} \). \( \Box \)

**Proof of Proposition 3.1:** Applying the extended revelation principle to program (3.5), I restrict my attention to direct mechanisms where the contracts are a pair of optimal treatment choices for the two patient types (i.e., \( \{\zeta_L, \tau_L, x_L\} \) and \( \{\zeta_H, \tau_H, x_H\} \)). Because treating high-risk patients has relatively higher risk, the payer needs higher-valued contract payment terms to ensure the provider’s participation \( (IRH \geq 0) \) – compared to the case where patients are low-risk. Thus, I omit the ICH constraint and check if it holds after designing optimal contracts. Furthermore, the provider has incentive to misrepresent low-risk patients, so IRL always holds. I prove this conjecture by showing that IRL > 0 given the payment contracts. Thus, omitting ICH and IRL, in the equilibrium IRH and ICL has to bind for all treatment choices. I solve the payer’s constrained maximization problem for the nine treatment choices given in Table A.3. Note that the provider may choose to change the treatment choice when misrepresenting a patient (c.f. ICL and ICH). Following Lemma 3.1 I know that the optimal treatment threshold decision of the provider does not depend on the patient risk (i.e., \( \tau^*_i = \frac{c + b_{ni} - b_{li} - w_i}{ah + b_{ni} - kb_{li}} \)) but does depend on payment terms. Therefore, if the provider misrepresents his patient and chooses a contract that is designed for other patient type, then the optimal treatment threshold will also change for the patient because of the change in contract terms. For example, when treatment option 2 is selected for both patient types (i.e., \( \{x^*_i, \tau^*_i\} = \{1, \tau^*_i\} \)) the payer’s problem is:
\[
\max_{\xi_L, \xi_H, \tau_L, \tau_H} V = \sum_{i \in \{L, H\}} \beta_i v_i(\xi_L, 1, \tau_i)
\]
\[
\text{s.t. } \tau_i^* = \arg\max_{0 \leq \tau \leq 1} g_i(1, \tau_i | \xi_i) \tag{OTC}
\]
\[
(w_L - c) \left( \int_{\tau_L^*}^1 f_L(y) dy \right) + b_{t_L} \left( \int_{\tau_L^*}^1 (1 - ky) f_L(y) dy \right) + b_{n_L} \left( \int_0^{\tau_L^*} (1 - y) f_L(y) dy \right) - ah \left( \int_0^{\tau_L^*} y f_L(y) dy \right)
\]
\[
= (w_H - c) \left( \int_{\tau_H^*}^1 f_L(y) dy \right) + b_{t_H} \left( \int_{\tau_H^*}^1 (1 - ky) f_L(y) dy \right) + b_{n_H} \left( \int_0^{\tau_H^*} (1 - y) f_L(y) dy \right) - ah \left( \int_0^{\tau_H^*} y f_L(y) dy \right) \tag{ICL}
\]
\[
(w_H - c) \left( \int_{\tau_H^*}^1 f_H(y) dy \right) + b_{t_H} \left( \int_{\tau_H^*}^1 (1 - ky) f_H(y) dy \right) + b_{n_H} \left( \int_0^{\tau_H^*} (1 - y) f_H(y) dy \right) - ah \left( \int_0^{\tau_H^*} y f_H(y) dy \right) = 0 \tag{IRH}
\]

Payer’s problem for other treatment choices can be presented similarly. When the test is ordered, the payer’s value function can either be concave or convex in threshold level, so the optimal treatment threshold might be an interior point or an extreme point. Note that if FOC gives a minimum \( \left( \frac{d^2 V}{d \tau^2} \right) > 0 \) then the underlying treatment choice will be dominated by other treatment choices that consider the extreme points (i.e., \( \tau_i^* = 0 \) or \( \tau_i^* = 1 \). Thus, if the test is ordered \( \frac{d V}{d \tau} = 0 \iff \tau_i = \frac{w_i + b_{t_i} - b_{n_i} + z}{h + k b_{t_i} - b_{n_i} + (1-k)(\lambda_3 - \lambda_4)} \). From Lemma 3.1 the payer knows the best response function of the provider for optimal treatment threshold of each patient type. Therefore, the payer would choose payment terms such that the provider’s best response (i.e., \( \tau_i^* = \frac{c + b_{n_i} - b_{t_i} - w_i}{ah + b_{n_i} - kb_{t_i}} \)) results in payer’s optimal treatment threshold.

From ICL, IRH and \( \frac{w_i + b_{t_i} - b_{n_i} + z}{h + k b_{t_i} - b_{n_i} + (1-k)(\lambda_3 - \lambda_4)} = \frac{c + b_{n_i} - b_{t_i} - w_i}{ah + b_{n_i} - kb_{t_i}}, \) I get \( \xi_L = \{w_L(b_{t_i}), b_{n_L}(b_{t_i}), b_{t_L}\} \) and \( \xi_H = \{w_H(b_{t_i}), b_{n_H}(b_{t_i}), b_{t_H}\}. \) Replacing the payment terms in the provider’s optimal treatment threshold I get \( \tau_i^* = \hat{\tau} \) for all treatment choices with \( x_i^* \neq 0. \) Thus, regardless of the slack variables (e.g., \( b_{t_i} \)), the provider would choose the same treatment threshold for both patient types. The payer would choose the remaining
two terms to maximize its value function. Replacing payment terms into payer’s value function, I show that when the test is used for both patient types (i.e., \(\{\tau^*_U, \tau^*_R\} = \{1, \tau^*_U\}\)), while the value function is independent of changes in \(b_{UL}(i.e., \frac{dV(\cdot)}{db_{UL}} < 0)\); therefore, \(b_{UL} = 0\). I have: \(w_H = (\left(c \int_0^1 f_H(y) dy + ah \int_0^y yf_H(y) dy\right)\left(z + c - (1 + \alpha)h - (1 - k)(\lambda_s - \lambda_r)\right) + \left(f_0^1 f_H(y) dy - \int_0^y yf_H(y) dy\right)\left(ahz - c(1 - k)(\lambda_s - \lambda_r)\right) / \left((z + c) \int_0^1 f_H(y) dy - (1 - \int_0^y yf_H(y) dy\right)\left((1 + \alpha)h + (1 - k)(\lambda_s - \lambda_r)\right)\right), b_{nH} = \left((z + c)ah \int_0^1 f_H(y) dy + \int_0^y yf_H(y) dy\right)\left((1 + \alpha)h + (1 - k)(\lambda_s - \lambda_r)\right) / \left((1 - \int_0^y yf_H(y) dy\right)\left((1 + \alpha)h + (1 - k)(\lambda_s - \lambda_r)\right) - (z + c) \int_0^1 f_H(y) dy\right) / \left((1 - \int_0^y yf_H(y) dy\right)\left((1 + \alpha)h + (1 - k)(\lambda_s - \lambda_r)\right) - (z + c) \int_0^1 f_H(y) dy\right) / \left((1 - \int_0^y yf_H(y) dy\right)\left((1 + \alpha)h + (1 - k)(\lambda_s - \lambda_r)\right) - (z + c) \int_0^1 f_H(y) dy\right).

Payment terms for the other treatment choices are derived using a similar approach. Next, I check if \(IRL\) and \(ICH\) hold for all nine treatment choices. I show that when the provider treats high-risk patients (i.e., \(\tau^*_H = 0\)) then \(IRL = 0\). Otherwise,
\[
f_0^\hat{t} yf_H(y) \, dy \left((1 + \alpha) h + (1 - k)(\lambda_s - \lambda_r)\right) - (z + c) \int_0^1 f_H(y) \, dy > 0.
\]
Using integration by parts I have: \(\int_0^\hat{t} yf_H(y) \, dy = [y F_H(y)]_0^\hat{t} - \int_0^\hat{t} F_H(y) \, dy = \hat{t} F_H(\hat{t}) - \int_0^\hat{t} F_H(y) \, dy\). This can be replicated for the risk distribution of low-risk patients, so I get \(I_{RL} = \Gamma(\hat{t}) = a h \int_0^{\hat{t}} F_L(y) \, dy - \int_0^{\hat{t}} F_H(y) \, dy\). \(\Gamma(\hat{t}) > 0\) because \(F_L(y) \leq F_H(y)\) and \(1 - \tau + \int_0^{\hat{t}} F_H(y) \, dy > 0\) because \(0 \leq \hat{t} \leq 1\). Similar logic can be used to show \(ICH > 0\).

To find the equilibrium treatment decisions, I examine the conditions for \(\max[V(0), V(1)] > V(\hat{t})\) such that the payer would skip the test. Proposition 3.1 illustrates the equilibrium solution resulted from the comparison of the outcomes from all treatment choices. □

**Proof of Corollary 3.1**: Cases 1, 2, and 3 correspond to treatment choices with \(\tau^*_H = 0\), where \(I_{RL} = 0\). Case 4 and 5 correspond to treatment choices with \(\tau^*_H \neq 0\), where \(I_{RL} = \Gamma(\hat{t})\), and \(\hat{t} = 1\) in case 5. □

**Proof of Corollary 3.2**: The maximum value for \(\bar{B}_L\) and \(\bar{B}_H\) is at \(\hat{t}_L = E[r_L]\) and \(\hat{t}_H = \frac{E[r_H]}{1 - \beta E[r(1)]}\) respectively. \(\bar{B}_L|_{\hat{t}=\hat{t}_L} > \bar{B}_H|_{\hat{t}=\hat{t}_H}\) because \(\int_0^{\hat{t}} F_L(y) \, dy \geq \int_0^{\hat{t}} F_H(y) \, dy\) and

\[
\frac{\beta}{1 - \beta z + c} \Gamma(\hat{t}) > 0.
\]

**Proof of Corollary 3.3**: Contract payment terms for all subgames have similar format as in subgame 1 that is presented above. It can be shown that \(w_H \geq 0\) and \(b_{nH} \geq 0\). Moreover, \(b_{nL} = b_{nH} + \varphi b_{tl} \geq 0\) because \(\varphi > 0\). Lastly, \(w_L = w_H - \psi b_{tl} \geq 0\) (where \(\psi > 0\)) when \(0 \leq b_{tl} < \frac{\alpha h + b_{nL}}{k}\). □

**Proof of Corollary 3.4**: Corollary 3.4 is a direct result of the Proposition 3.1. □
To illustrate how testing only one patient type always results in higher value for the payer when compared to testing patients from both risk profiles, I draw the payer’s willingness-to-pay for the test for three different triangular distributions.

**Figure A.1:** Willingness-to-pay when \( f_L(r) = 2 - 2r \) and \( f_H(r) = 2r \)

**Figure A.2:** Willingness-to-pay when

\[
\begin{align*}
    f_L(r) = \begin{cases} 5r & 0 \leq r < 0.4 \\ 2(1 - r)/0.6 & 0.4 \leq r \leq 1 \end{cases} \\
    f_H(r) = \begin{cases} 2r/0.6 & 0 \leq r < 0.6 \\ 2(1 - r)/0.4 & 0.6 \leq r \leq 1 \end{cases}
\end{align*}
\]
As the two distributions get closer to each other the difference between testing both patient types versus only one patient type gets smaller. We know that the information rent in the optimal treatment strategies decreases as the two distributions get closer to each other (i.e., $\Gamma(\tau)$). In addition, the value gained from testing patients from both patient risk profiles in Case 6 does not outweigh the high information rent in this Case. Thus, Cases 2 and 4 remain the best choices of the payer even when the two distributions are similar.

**Derivation for decentralized system with full information:** The payer offers a type specific contract to the provider and extracts the entire provider surplus. $g_i(.)$ and $v_i(.)$ are the same as in Equations (3.3) and (3.4). The optimization problem for each patient type is:

\[
\max_{\xi_i} \bar{V}_i = v_i(\xi_i, x_i, \tau_i) \quad \text{for} \quad i \in \{L, H\}
\]

s.t. \( \{x_i^*, \tau_i^*\} = \arg\max_{x_i \in \{0,1\}, \tau_i \in [0,1]} g_i(x_i, \tau_i | \xi_i) \)
The treatment threshold is \( \tau_i^* = \frac{c + b_{ti} - b_{ti} - w_{li}}{ah + b_{ni} - kb_{ti}} \) (cf. Lemma 3.1). The payer’s optimization problem is to satisfy the following two conditions for each of the three options:

Conditions 1A) ensures that the provider’s best response (i.e., \( \tau_i^* = \frac{c + b_{ti} - b_{ti} - w_{li}}{ah + b_{ni} - kb_{ti}} \)) results in payer’s optimal treatment threshold, \( w_i((1 + \alpha)h + (1 - k)(\lambda_s - \lambda_r) + b_{ti}((1 + \alpha)h - (z + c)k + (1 - k)(\lambda_s - \lambda_r)) + b_{ni}(z + c - (1 + \alpha)h - (1 - k)(\lambda_s - \lambda_r)) + ahz - ch - c(1 - k)(\lambda_s - \lambda_r) = 0); \) Conditions 1B) is the individual-rationality constraint, \( (w_i - c) \int_{\tau_i}^{1} f_i(y) dy + b_{ti} \int_{\tau_i}^{1} (1 - ky) f_i(y) dy + b_{ni} \int_{0}^{\tau_i} (1 - y) f_i(y) dy - ah \int_{0}^{\tau_i} y f_i(y) dy = 0. \)

Conditions 1A and 1B give the payment contracts for each option as well as the \( \hat{\tau} \) that maximizes social welfare when \( x_i = 1 \). When \( x_i = 0 \), then, the provider should treat patients if and only if, \( \check{V}_i(\hat{\tau} = 0) > \check{V}_i(\hat{\tau} = 1) \iff \int_{0}^{1} y f_i(y) dy > \frac{z + c}{(1 + \alpha)h + (1 - k)(\lambda_s - \lambda_r)} \iff E[r_{li}] > \hat{\tau} \). The payer compares social welfare \( \check{V}_i(\hat{\tau}) \) when \( x_i = 1 \) with \( \max[\check{V}_i(\hat{\tau} = 0), \check{V}_i(\hat{\tau} = 1)] \) when \( x_i = 0 \). Then \( \max[\check{V}_i(\hat{\tau} = 0), \check{V}_i(\hat{\tau} = 1)] > \check{V}_i(\hat{\tau}) \iff B > \hat{B}_i = \frac{z + c}{\hat{\tau}} \left\{ \int_{0}^{\hat{\tau}} f_i(y) dy + \min(0, E[r_{li}] - \hat{\tau}) \right\}. \)

**Proof of Proposition 3.2:** The payer’s value function under case 4 is \( V = \lambda_s + (1 - \beta) \left( -k \left( \int_{0}^{1} y f_H(y) dy \right) (\lambda_s - \lambda_r) - (z + c) \int_{\hat{\tau}}^{1} f_H(y) dy \right) - \left( \int_{0}^{\hat{\tau}} y f_L(y) dy \right) \left( (1 + \alpha)h + (1 - k)(\lambda_s - \lambda_r) - B \right) + \beta \left( - \left( \int_{0}^{1} y f_L(y) dy \right) (h(1 + \alpha) + \lambda_s - \lambda_r) - \Gamma(\hat{\tau}) \right). \)

Using implicit function theorem, I have

\[
\frac{d V(\cdot)}{d \alpha} > 0 \iff E[r_{li}] \leq \frac{a h \hat{\tau}^2 (\Omega_H - F_H(\hat{\tau}) \Omega_H) - (z + c) \Omega_H - ah \hat{\tau}^2 (\Omega_L - \Omega_H)}{(z + c) \Omega_H^2} - \frac{(1 - \beta) \left( \int_{0}^{\hat{\tau}} y f_H(y) dy \right)}{\beta}.
\]

I find the necessary conditions for Proposition 3.2.a. by including the optimality conditions of case 4 from Proposition 3.1. Similar approach can be used for Proposition 3.2.b. \( \square \)

**Proof of Proposition 3.3:** The proof is similar to the logic presented in Proposition 3.2. \( \square \)
Proof of Lemma 3.3 and Proposition 3.4: The program (3.12) can be solved similar to the system with full information where the payer offers a type specific contract to the provider and extracts the entire provider surplus. The payer’s value function is independent of changes in $b_t$ when

$$w = \frac{\int_0^\varphi yf(y)\,dy - \int_0^\varphi f(y)\,dy}{(z+c)\int_0^\varphi f(y)\,dy - \int_0^\varphi yf(y)\,dy - (1+\alpha)h + (1-k)(\lambda_s-\lambda_r)}$$

$$b_n = \frac{\alpha (z+c)\int_0^\varphi f(y)\,dy + \int_0^\varphi yf(y)\,dy}{(1+\alpha)h + (1-k)(\lambda_s-\lambda_r) - (z+c)\int_0^\varphi f(y)\,dy}$$

The ex-ante outcome of the payer when the test is compulsory is:

$$V^{com} = \beta \left( \lambda_s \int_0^\varphi f_L(y)\,dy - (1+\alpha)h + \lambda_s - \lambda_r \right) \int_0^\varphi yf_L(y)\,dy - (z+c - \lambda_s) \int_0^\varphi f_L(y)\,dy - k(\lambda_s - \lambda_r) \int_0^\varphi yf_H(y)\,dy + (1-\beta) \left( \lambda_s \int_0^\varphi f_H(y)\,dy - (1+\alpha)h + \lambda_s - \lambda_r \right) \int_0^1 yf_H(y)\,dy - (z+c - \lambda_s) \int_0^1 f_H(y)\,dy - k(\lambda_s - \lambda_r) \int_0^1 yf_H(y)\,dy - B - C_V$$

In order to find when it is optimal to make the test compulsory, I examine the conditions for $V^{com}(\hat{\tau}) > \max[V(0), V(1), V(\hat{\tau})]$ such that social welfare is higher when the test is compulsory. Lemma 3.3 illustrates the region where compulsory testing is optimal when $C_V = 0$. □

Proof of Corollary 3.5: The maximum value for the function $\bar{B}^{com}$ is at $\tau_H$. □
Table A.3: Optimization problem for 9 subgames

<table>
<thead>
<tr>
<th>Treatment choice</th>
<th>Type-(L)</th>
<th>Type-(H)</th>
<th>Type-(H)</th>
</tr>
</thead>
<tbody>
<tr>
<td>({x^<em>_L, \tau^</em>_L} = {0, 0}) (Option 0)</td>
<td>({x^<em>_L, \tau^</em>_L} = {0, 1}) (Option 1)</td>
<td>({x^<em>_L, \tau^</em>_L} = {1, \tau^*_L}) (Option 2)</td>
<td></td>
</tr>
<tr>
<td>({x^<em>_H, \tau^</em>_H} = {0, 0}) (Option 0)</td>
<td>({x^<em>_H, \tau^</em>_H} = {0, 1}) (Option 1)</td>
<td>({x^<em>_H, \tau^</em>_H} = {1, \tau^*_H}) (Option 2)</td>
<td></td>
</tr>
<tr>
<td>({x^<em>_L, \tau^</em>_L} = {0, 0}) (Option 0)</td>
<td>({x^<em>_L, \tau^</em>_L} = {0, 1}) (Option 1)</td>
<td>({x^<em>_L, \tau^</em>_L} = {1, \tau^*_L}) (Option 2)</td>
<td></td>
</tr>
<tr>
<td>({x^<em>_H, \tau^</em>_H} = {0, 0}) (Option 0)</td>
<td>({x^<em>_H, \tau^</em>_H} = {0, 1}) (Option 1)</td>
<td>({x^<em>_H, \tau^</em>_H} = {1, \tau^*_H}) (Option 2)</td>
<td></td>
</tr>
<tr>
<td>({x^<em>_L, \tau^</em>_L} = {0, 0}) (Option 0)</td>
<td>({x^<em>_L, \tau^</em>_L} = {0, 1}) (Option 1)</td>
<td>({x^<em>_L, \tau^</em>_L} = {1, \tau^*_L}) (Option 2)</td>
<td></td>
</tr>
<tr>
<td>({x^<em>_H, \tau^</em>_H} = {0, 0}) (Option 0)</td>
<td>({x^<em>_H, \tau^</em>_H} = {0, 1}) (Option 1)</td>
<td>({x^<em>_H, \tau^</em>_H} = {1, \tau^*_H}) (Option 2)</td>
<td></td>
</tr>
<tr>
<td>({x^<em>_L, \tau^</em>_L} = {0, 0}) (Option 0)</td>
<td>({x^<em>_L, \tau^</em>_L} = {0, 1}) (Option 1)</td>
<td>({x^<em>_L, \tau^</em>_L} = {1, \tau^*_L}) (Option 2)</td>
<td></td>
</tr>
<tr>
<td>({x^<em>_H, \tau^</em>_H} = {0, 0}) (Option 0)</td>
<td>({x^<em>_H, \tau^</em>_H} = {0, 1}) (Option 1)</td>
<td>({x^<em>_H, \tau^</em>_H} = {1, \tau^*_H}) (Option 2)</td>
<td></td>
</tr>
<tr>
<td>({x^<em>_L, \tau^</em>_L} = {0, 0}) (Option 0)</td>
<td>({x^<em>_L, \tau^</em>_L} = {0, 1}) (Option 1)</td>
<td>({x^<em>_L, \tau^</em>_L} = {1, \tau^*_L}) (Option 2)</td>
<td></td>
</tr>
<tr>
<td>({x^<em>_H, \tau^</em>_H} = {0, 0}) (Option 0)</td>
<td>({x^<em>_H, \tau^</em>_H} = {0, 1}) (Option 1)</td>
<td>({x^<em>_H, \tau^</em>_H} = {1, \tau^*_H}) (Option 2)</td>
<td></td>
</tr>
<tr>
<td>({x^<em>_L, \tau^</em>_L} = {0, 0}) (Option 0)</td>
<td>({x^<em>_L, \tau^</em>_L} = {0, 1}) (Option 1)</td>
<td>({x^<em>_L, \tau^</em>_L} = {1, \tau^*_L}) (Option 2)</td>
<td></td>
</tr>
<tr>
<td>({x^<em>_H, \tau^</em>_H} = {0, 0}) (Option 0)</td>
<td>({x^<em>_H, \tau^</em>_H} = {0, 1}) (Option 1)</td>
<td>({x^<em>_H, \tau^</em>_H} = {1, \tau^*_H}) (Option 2)</td>
<td></td>
</tr>
</tbody>
</table>
Appendix-C

**Proof of Proposition 4.1:** The central planner’s profit function under reward strategy is concave in efforts (i.e., $\frac{d^2 \Pi_{cr}}{d e_H^2} < 0$; $\frac{d^2 \Pi_{cr}}{d e_p^2} < 0$). Considering the non-negative savings constraint, I have two optimal set of efforts:

1) Non-binding savings constraint: $e_{H*}^* = \frac{b_p \omega^2 - (\delta H + \overline{c}_p - c_t + b_p + 1) \omega + \delta H + \overline{c}_p - \alpha + 1}{2(2-\omega)}$; $e_p^* = \omega$

where $\omega$ solves $3b_p^2 + (12b_p^2 - b_p(4(c_t - \overline{c}_H - \overline{c}_p) - 12)) - 8\omega^3 + (13b_p^2 - b_p(16(c_t - \overline{c}_H - \overline{c}_p) - 2\alpha + 20) + \overline{c}_p^2 + 2\overline{c}_H(c_p - c_t - 1) - 2\overline{c}_p(1 + c_t) + c_t(2 + c_t - 31)\omega^2 + \left(-4b_p^2 + b_p(16(c_t - \overline{c}_H - \overline{c}_p) - 8\alpha + 32) - 4(\overline{c}_H^2 + \overline{c}_p^2 + 2\overline{c}_H(c_p - c_t - 1) - 2\overline{c}_p(1 + c_t) + c_t(2 + c_t - 7))\right)$\omega - $b_p(4(c_t - \overline{c}_H - \overline{c}_p) + 12(1 - \omega)) + (c_t - \overline{c}_H - \overline{c}_p + \alpha + 3)(c_t - \overline{c}_H - \overline{c}_p + 1 - \alpha) = 0$. Efforts are decreasing functions in target price (i.e., $\frac{d e_p^*}{d c_t} < 0$; $\frac{d e_H^*}{d c_t} < 0$) and $e_{H*}^* \leq e_p^*$. Given the optimal effort levels in the reward strategy, the savings constraint is non-binding if and only if $c_{t}^{CH} < c_t$ where $c_{t}^{CH} =\frac{-b_p^3 - 3b_p^2 + b_p(2(\overline{c}_H + \overline{c}_p - 1) - 1 + (1 + b_p^2)(\overline{c}_H + \overline{c}_p - 1) + 2(2 + \alpha + b_p^2)\sqrt{b_p^2 + b_p(2 + \alpha + 9)}}{2b_p}$. Thus, $c_{t}^{CH'} =\frac{-b_p^2 + b_p(\overline{c}_H + \overline{c}_p - \alpha - 1) + 2(\overline{c}_H + \overline{c}_p - 1 + \sqrt{b_p^4 + 2b_p^2(\overline{c}_H + \overline{c}_p - 1 + \alpha + b_p^2(\overline{c}_H + \overline{c}_p - 1 + \alpha) + 2b_p(9 - \alpha) + 9 - \alpha)}}{2b_p}$. If $c_{t}^{CH'} < c_t$ then $\frac{d \Pi_{cr}}{d e_p} |_{e_{H*}=0} \Rightarrow e_p^* = \frac{b_p(1 - \alpha) + \overline{c}_H + \overline{c}_p - c_t}{2(1 + b_p)}$. Furthermore, $b_p(1 - \alpha) + \overline{c}_H + \overline{c}_p - c_t = 0 \Rightarrow c_{t}^{CH''} = b_p(1 - \alpha) + \overline{c}_H + \overline{c}_p$ such that if $c_{t}^{CH''} < c_t$ then $e_{H*}^* = 0; e_p^* = 0$.

2) Binding savings constraint: If $c_t \leq c_{t}^{CH}$ then $e_{H*}^* = \frac{\overline{c}_H + \overline{c}_p - c_t}{1 + b_p^2}; e_p^* = \frac{b_p(\overline{c}_H + \overline{c}_p - c_t)}{1 + b_p^2}$. $0 \leq e_i \leq 1 \Leftrightarrow c_t \geq \max[\frac{b_p^2 \overline{c}_H - \overline{c}_p}{b_p^2}, \frac{b_p(\overline{c}_H + \overline{c}_p) - b_p^2 - 1}{b_p}].$
The central planner’s profit function under penalty strategy is concave in efforts (i.e.,
\[
\frac{d^2 \Pi_{CP}^{*}}{d e_{H}^2} < 0; \quad \frac{d^2 \Pi_{CP}^{*}}{d e_{P}^2} < 0.
\]
The first-order condition (FOC) gives
\[
e_{H}^{*} = \frac{1-\alpha}{2} \quad \text{and} \quad e_{P}^{*} = \frac{b_{p}(1-\alpha)}{2}.
\]
The savings are negative if and only if \( c_{t} < c_{t}^{CL} \) where
\[
c_{t}^{CL} = \frac{1}{2} (2(\bar{e}_{H} + \bar{e}_{P}) + \alpha - 1) - \frac{1}{2} b_{p}^2 (1 - \alpha) \quad \text{such that} \quad 0 \leq e_{t} \leq 1, \quad \text{and} \quad c_{t} \geq \max \left[ \frac{b_{p}^2 \bar{e}_{H} - \bar{e}_{p}}{b_{p}^2}, \frac{b_{p}(\bar{e}_{H} + \bar{e}_{P}) - b_{p}^2 - 1}{b_{p}} \right]
\]
and
\[
\Pi_{CP}^{*}\big|_{e_{H}^{*} = \frac{1-\alpha}{2}, e_{P}^{*} = \frac{b_{p}(1-\alpha)}{2}} > \Pi_{CR}^{*}\big|_{e_{H}^{*} = \frac{\bar{e}_{H} + \bar{e}_{P} - c_{t}}{1+b_{p}^2}, e_{P}^{*} = \frac{b_{p}(\bar{e}_{H} + \bar{e}_{P} - c_{t})}{1+b_{p}^2}}
\]
when \( c_{t} < c_{t}^{CL} \); thus, the central planner would rather be in the penalty strategy instead of increasing efforts to bind the savings constraint. \( \Box \)

**Proof of Proposition 4.2:** \( \Pi_{H}^{S}\big|_{e_{H}^{*} = \frac{1-\alpha}{2}, e_{P}^{*} = \frac{b_{p}(1-\alpha)}{2}} = \alpha (\bar{e}_{H} + \bar{e}_{P}) \). The central planner’s profit in penalty and zero savings cases (cases 2 and 3 of Proposition 4.1) is always smaller than the total status quo FFS profit:

Case 3) \( \Pi_{CP}^{*} = \alpha (\bar{e}_{H} + \bar{e}_{P}) + \frac{1}{4} (1 - 4(\bar{e}_{H} + \bar{e}_{P}) + 4c_{t} + b_{p}^2 (1 - \alpha)^2 + \alpha^2 - 2\alpha) \).

Thus, \( \Pi_{CP}^{*} < \Pi_{H}^{S} + \Pi_{P}^{S} \) because \( \frac{1}{4} (1 - 4(\bar{e}_{H} + \bar{e}_{P}) + 4c_{t} + b_{p}^2 (1 - \alpha)^2 + \alpha^2 - 2\alpha) < 0 \) when \( c_{t} < c_{t}^{CL} \).

Case 2) \( \Pi_{CR}^{*}\big|_{e_{H}^{*} = \frac{\bar{e}_{H} + \bar{e}_{P} - c_{t}}{1+b_{p}^2}, e_{P}^{*} = \frac{b_{p}(\bar{e}_{H} + \bar{e}_{P} - c_{t})}{1+b_{p}^2}} = \frac{(1 + b_{p}^2)c_{t}\alpha - (\bar{e}_{H} + \bar{e}_{P} - c_{t})^2}{1 + b_{p}^2} \).

Thus, \( \Pi_{CR}^{*} < \Pi_{H}^{S} + \Pi_{P}^{S} \) because \( (1 + b_{p}^2)(\bar{e}_{H} + \bar{e}_{P} - c_{t}) (\bar{e}_{H} + \bar{e}_{P} - c_{t} + \alpha (1 + b_{p}^2)) < 0 \) when \( c_{t}^{CL} \leq c_{t} \leq c_{t}^{CH} \).

In case 1 where the central planner receives reward payment the central planner may be better off compared to the total profit in the status quo FFS model.

Case 1.iii) \( \Pi_{CR}^{*}\big|_{e_{H}^{*} = 0, e_{P}^{*} = 0} = \alpha (\bar{e}_{H} + \bar{e}_{P}) + c_{t} - \bar{e}_{H} - \bar{e}_{P} \).

Thus, \( \Pi_{CR}^{*} > \Pi_{H}^{S} + \Pi_{P}^{S} \) because \( c_{t} - \bar{e}_{H} - \bar{e}_{P} > 0 \) when \( c_{t}^{CH''} < c_{t} \).

Case 1.ii)
Thus, has a unique target price threshold and is satisfied if and only if (1) \( b_p > \kappa \) and \( c_t > c_t^{C1} \) (where \( c_t^{C1} = \tilde{c}_H + \tilde{c}_p - (1 + \alpha)b_p - 2 + \frac{b_p}{2(1+b_p)} \)).

Therefore, the absolute lower bound threshold for the target price is located in Case 1.ii and otherwise in Case 1.i.

**Case 1.i)**

\[
\frac{d^2 \Pi^{cr\ast}}{d c_t^2} \bigg|_{e_H^\ast = 0, e_p^\ast = \frac{b_p(1-\alpha)+\tilde{c}_H+\tilde{c}_p-c_t}{2(1+b_p)}} > 0 \quad \text{and} \quad \frac{d \Pi^{cr\ast}}{d c_t} \bigg|_{e_H^\ast = 0, e_p^\ast = \frac{b_p(1-\alpha)+\tilde{c}_H+\tilde{c}_p-c_t}{2(1+b_p)}} \bigg|_{c_t^{CH'} \leq c_t} > 0.
\]

Thus, \( \Pi^{cr\ast} > \Pi_{H}^{S\ast} + \Pi_{P}^{S\ast} \) has a unique target price threshold and is satisfied if and only if \( b_p \leq \kappa \) and \( c_t > c_t^{C2} \) such that \( c_t^{C2} \) is the unique threshold that solves \( \Pi^{cr\ast} = \Pi_{H}^{S\ast} + \Pi_{P}^{S\ast} \).

Therefore, the absolute lower bound of the target price in centralized system is as follows.

\[
c_t^C = \begin{cases} 
  c_t^{C1} & \text{if } b_p > \kappa \\
  c_t^{C2} & \text{if } b_p \leq \kappa 
\end{cases}
\]

**Proof of Proposition 4.3:** In this strategy, because the provider will only be compensated according to the FFS model, the provider does not exert any cost-reduction effort (i.e., \( e_p^\ast = 0 \)); resulting in \( \Pi_P^{cr\ast} (e_p) = \alpha \tilde{c}_p \). The hospital’s profit function under reward strategy is concave in efforts (i.e., \( \frac{d^2 \Pi_P^{cr\ast}}{d e_H^2} < 0 \)). Considering the non-negative savings constraint, I have two optimal set of efforts:
1) Non-binding savings constraint: \( e_H^* = \frac{1}{4} (\bar{c}_H + \bar{c}_p - c_t + 1 - \alpha) \). Given the optimal effort levels in the reward strategy, the savings constraint is non-binding if and only if \( c_t^D < c_t \) where \( c_t^D = \bar{c}_H + \bar{c}_p - \frac{1}{3} (1 - \alpha) \). The hospital’s Effort is decreasing functions in target price (i.e., \( \frac{d e_H^*}{d c_t} < 0 \)); thus, to ensure \( 0 \leq e_t \) I solve \( e_H^* = 0 \) that gives \( c_t^D = \bar{c}_H + \bar{c}_p + 1 - \alpha \). If \( c_t^D < c_t \) then \( e_p^* = 0 \).

2) Binding savings constraint: If \( c_t \leq c_t^D \) then \( e_H^* = \bar{c}_H + \bar{c}_p - c_t \). \( 0 \leq e_t \leq 1 \Leftrightarrow c_t \geq \bar{c}_H + \bar{c}_p - 1 \).

The hospital’s profit function under penalty strategy is concave in cost-reduction effort (i.e., \( \frac{d^2 \Pi_H^{DP}}{d e_H^2} < 0 \)). The first-order condition (FOC) gives \( e_H^* = \frac{1-\alpha}{2} \). The savings are negative if and only if \( c_t < c_t^{DL} \) where \( c_t^{DL} = \bar{c}_H + \bar{c}_p - \frac{1}{2} (1 - \alpha) \). \( c_t^{DL} \geq \bar{c}_H + \bar{c}_p - 1 \) and \( \Pi_H^{DP} |_{e_H^* = \frac{1-\alpha}{2}; e_p^* = 0} > \Pi_H^{Dr} |_{e_H^* = \bar{c}_H + \bar{c}_p - c_t; e_p^* = 0} \) when \( c_t < c_t^{DL} \); thus, the hospital would rather be in the penalty strategy instead of increasing efforts to bind the savings constraint. □

**Proof of Proposition 4.4:** The hospital’s profit in penalty and zero savings cases (cases 2 and 3 of Proposition 4.3) is always smaller than the total status quo FFS profit:

Case 3) \( \Pi_H^{DP^*} = c_t - \bar{c}_p - \bar{c}_H (1 - \alpha) + \frac{1}{4} (1 - \alpha)^2 \). Thus, \( \Pi_H^{DP^*} < \Pi_H^{S^*} \) because \( \frac{1}{4} - \bar{c}_p - \bar{c}_H + c_t - \frac{\alpha}{2} + \frac{\alpha^2}{4} < 0 \) when \( c_t < c_t^{DL} \).

Case 2) \( \Pi_H^{Dr^*} |_{e_H^* = \bar{c}_H + \bar{c}_p - c_t; e_p^* = 0} = (c_t - \bar{c}_p) \alpha - (\bar{c}_H + \bar{c}_p - c_t)^2 \). Thus, \( \Pi_H^{Dr^*} < \Pi_H^{S^*} \) because \( -(\bar{c}_p + \bar{c}_H - c_t)^2 - (\bar{c}_p + \bar{c}_H - c_t) \alpha < 0 \) when \( c_t^{DL} \leq c_t \leq c_t^{DH} \).

In case 1 where the hospital receives reward payment the hospital may be better off compared to the status quo FFS model.

Case 1.ii) \( \Pi_H^{Dr^*} |_{e_H^* = 0; e_p^* = 0} = c_t + \bar{c}_H (1 - \alpha) - \bar{c}_p \). Thus, \( \Pi_H^{Dr^*} > \Pi_H^{S^*} \) because \( c_t - \bar{c}_H - \bar{c}_p > 0 \) when \( c_t^{DH^*} < c_t \).
Case 1.i)

\[
\Pi_H^{pr^*}\big|_{e_H^*=\frac{1}{4}(\tilde{c}_H+c_p-c_t+1-\alpha);e_p^*=0} = \frac{1}{8}\left(1 + \tilde{c}_H^2 - 6\tilde{c}_p + 6c_t - 2\alpha + (c_t - \tilde{c}_p + \alpha)^2 + 2\tilde{c}_H(c_p - c_t + 3\alpha - 3)\right)\frac{d^2 \Pi_H^{pr^*}}{d c_t^2}\bigg|_{e_H^*=\frac{1}{4}(\tilde{c}_H+c_p-c_t+1-\alpha);e_p^*=0} > 0 \text{ and } \\
\frac{d \Pi_H^{pr^*}}{d c_t}\bigg|_{e_H^*=\frac{1}{4}(\tilde{c}_H+c_p-c_t+1-\alpha);e_p^*=0} |_{c_t^P < c_t} > 0. \text{ Thus, } \Pi_H^{pr^*} > \Pi_H^s^* \text{ has a unique target price threshold and is satisfied if and only if } c_t > \tilde{c}_H + \tilde{c}_p - 3 - \alpha + 2\sqrt{2}\sqrt{1 + \alpha}. \text{ Therefore, the absolute lower bound of the target price in decentralized system with no gain-sharing agreement is } c_t^D = \tilde{c}_H + \tilde{c}_p - 3 - \alpha + 2\sqrt{2}\sqrt{1 + \alpha} \text{ where } c_t^c < c_t^D. \quad \square
\]

**Proof of Proposition 4.5:** The provider’s profit function under reward strategy is concave in cost-reduction effort (i.e., \(\frac{d^2 \Pi_H^{gr^*}}{d e_p^2} < 0\)). Considering the non-negative savings constraint, I have two optimal set of efforts:

1) Non-binding savings constraint: \(e_p^* = \frac{xp(\tilde{c}_H+c_p-c_t+e_H)(1-e_H)-abp}{2+2xbp(1-e_H)}\). The hospital’s profit function is concave in \(e_H\) and \(x_p\) (i.e., \(\frac{d^2 \Pi_H^{gr^*}}{d e_H^2} < 0; \frac{d^2 \Pi_H^{gr^*}}{d x_p^2} < 0\)). Define \(c^{ex} = c_t - \tilde{c}_H - \tilde{c}_p\). From FOC I have \(e_H^* = \phi\) and \(x_p^* = z\) where \(z\) solves \(-2b_p^2 + 4c^{ex} - 4b_p c^{ex} - 2c^{ex^2} - 2b_p^2 \alpha - 4b_p^3 \alpha + 2b_p c^{ex} \alpha + 4b_p^2 c^{ex} \alpha - b_p^3 \alpha^2 - 2b_p^4 \alpha^2 - 10\phi + 6c^{ex} \phi - c^{ex^2} \phi - 26\alpha \phi + 8c^{ex} \alpha \phi - 16\alpha^2 \phi + 5\phi^2 - 2c^{ex} \phi^2 + 8\alpha \phi^2 - \phi^3 + (4b_p^2 + 4b_p c^{ex} + 4c^{ex^2} + 2b_p^3 \alpha - 2b_p^2 c^{ex} \alpha + b_p^4 \alpha^2 + 4\phi + 2c^{ex^2} \phi + 12\alpha \phi - 4c^{ex} \alpha \phi + 8\alpha^2 \phi + 2\phi^2 + 4c^{ex} \phi^2 - 4\alpha \phi^2 + 2\phi^3)z + (3b_p^3 + 6b_p^2 c^{ex} + 3b_p c^{ex^2} + 12\phi + 18c^{ex} \phi + 6c^{ex^2} \phi + 9 \phi^2 - 3c^{ex^2} \phi^2 - 6\phi^3 - 6c^{ex} \phi^3 - 3\phi^4)z^2 + (b_p^4 + 2b_p^3 c^{ex} + b_p^2 c^{ex^2} + 16\phi + 20c^{ex} \phi + 6c^{ex^2} \phi - 2\phi^2 - 12c^{ex} \phi^2 - 6c^{ex^2} \phi^2 - 10\phi^3 - 4c^{ex} \phi^3 - 2c^{ex^2} \phi^3 + 2\phi^4 + 4c^{ex} \phi^4 + 2\phi^5)z^3 = 0; \) and \(\phi\) solves

\[
32 + 32b_p - 4b_p^2 - 4b_p^3 - b_p^4 - 16c^{ex} - 4b_p c^{ex} - 10b_p^2 c^{ex} - 2b_p^3 c^{ex} - 16c^{ex^2} - 6b_p c^{ex^2} - b_p^2 c^{ex^2} - 64\alpha - 80b_p \alpha - 24b_p^2 \alpha - 16b_p^3 \alpha - 16b_p c^{ex} \alpha + 8b_p^2 c^{ex} \alpha + 8b_p^3 \alpha^2 - 4b_p^4 \alpha^2 + (-176 - 260b_p - 126b_p^2 - 14b_p^3 + 3b_p^4 - 64c^{ex} - 32b_p c^{ex} + 14b_p^2 c^{ex} + 6b_p^3 c^{ex} + 16c^{ex^2} + 12b_p c^{ex^2} + 3b_p^2 c^{ex^2} + 3b_p^3 c^{ex^2} + \cdots
\]
\[64b_p \alpha + 88b_p^2 \alpha + 32b_p^3 \alpha - 8b_p^2 c^{ex} \alpha + 4b_p^4 \alpha^2)\phi + \left(-48 + 198b_p + 243b_p^2 + 60b_p^3 - 3b_p^4 + 48c^{ex} + 60b_p c^{ex} + 6b_p^2 c^{ex} - 6b_p^3 c^{ex} - 6b_p c^{ex^2} - 3b_p^2 c^{ex^2} - 48b_p^2 \alpha - 24b_p^3 \alpha\right)\phi^2 + \left(32 + 48b_p - 105b_p^2 - 62b_p^3 + b_p^4 - 24b_p c^{ex} - 14b_p^2 c^{ex} + 2b_p^3 c^{ex} + b_p^2 c^{ex^2} + 8b_p^3 \alpha\right)\phi^3 + \left(-18b_p - 11b_p^2 + 20b_p^3 + 4b_p^2 c^{ex}\right)\phi^4 + 3b_p^2 \phi^5 = 0\] such that \(0 \leq \phi \leq 1\).

Thus, by replacing \(e_H^*\) and \(x_P^*\) I get \(e_H^*(c_t, \phi)\). Given the optimal effort levels in the reward strategy, the savings constraint is non-binding if and only if \(c_t^{GH} < c_t\) where

\[c_t^{GH} = \frac{2(\bar{e}_H + \bar{e}_P + b_p(\bar{e}_H + \bar{e}_P) - b_p^2(1-\alpha)}{2 + b_p}.\]

The hospital must be better off when offering a gain-sharing agreement compared to the decentralized system with no gain-sharing agreement (i.e., \(\Pi_t^{*} \geq \Pi_t^{D*}\)). \(\Pi_t^{*} \geq \Pi_t^{D*} \iff c_t \leq c_t^{GH}\) where \(c_t^{GH} = \bar{c}_H + \bar{c}_P + 4\alpha - \phi + 5 + \frac{b_p(2 + \alpha + 2b_p(1-\alpha))}{2 + \phi} \]

\[+ \left(4(2 + b_p(2 + \alpha) + (3 + 4\alpha - \phi)\phi + 2\alpha b_p^2)^2 - 4(2 + \phi)(10\phi + 2\alpha(13 + 8\alpha)\phi^2 + \phi^3 + b_p^2(2 + \alpha(1 + 2b_p)(2 + \alpha(5 + 6b_P)))\right)^{1/2} / (2(2 + \phi)).\]

2) Binding savings constraint: If \(c_t \leq c_t^{GH}\) then \(e_H^* = 0\); \(e_p^* = \frac{\bar{e}_H + \bar{e}_P - c_t}{b_p}\); \(x_p^* = 1\). \(0 \leq e_i \leq 1 \iff c_t \geq \bar{c}_H + \bar{c}_P - b_p\).

The provider’s profit function under penalty strategy is concave in effort (i.e., \(\frac{d^2 \Pi_P^{GP}}{d e_p^2} < 0\)). The first-order condition (FOC) gives \(e_p^* = \frac{1}{2} b_p (1 - \alpha)\). In the penalty case the hospital will push all the penalty to the provider by setting \(e_H^* = 0\); \(y_p^* = 1\). The savings are negative if and only if \(c_t < c_t^{GL}\) where \(c_t^{GL} = \frac{1}{2} \left(2(\bar{c}_H + \bar{c}_P) - b_p^2 (1-\alpha)\right)\) such that \(0 \leq e_i \leq 1\). When \(1 < b_p < \frac{2}{1-\alpha} c_t^{GL} \geq \bar{c}_H + \bar{c}_P - b_p\) and \(\Pi_t^{GP} \mid_{e_H^* = 0; e_p^* = \frac{\bar{e}_H + \bar{e}_P - c_t}{b_p}} > \Pi_t^{GL} \mid_{e_H^* = 0; e_p^* = \frac{\bar{e}_H + \bar{e}_P - c_t}{b_p}}\) when \(c_t < c_t^{GL}\); thus, the hospital would rather be in the penalty strategy instead of zero savings case. □

**Proof of Corollary 4.1:** Corollary 4.1 is a direct result of the Proposition 4.5. □
Proof of Corollary 4.2: Following proposition 4.5, the hospital sets $x_p^* = 0$ if $c_t > c_t^{GH'}$. This implies that $c_t > c_t^{GH'} \Rightarrow \Pi_p^* = \Pi_{p}^{b*} = \alpha \bar{c}_p$. The provider’s profit decreases in target price because $\Pi_p^{GH'} |_{c_t=c_t^{GH'}} < \alpha \bar{c}_p$. □

Proof of Proposition 4.6: Under the gain sharing agreements hospital is always better off compared to the status quo FFS model. The provider’s profit in penalty and zero savings cases is always smaller than the total status quo FFS profit. In case 1 where the provider may be better off compared to the status quo FFS model: $\Pi_p^{gr*} > \Pi_p^{s*}$ if $c_t^G < c_t$ where $c_t^G$ is unique because $\frac{d^2 \Pi_p^{gr}}{d c_t^2} > 0$ and $\frac{d \Pi_p^{gr}}{d c_t} |_{c_t^{GH} < c_t} > 0$. $c_t^G$ solves ($\bar{c}_p - b_p e_p^*(c_t, \phi) \alpha + (1 - e_p^*(c_t, \phi)) z (1 - \phi)(c_t - \bar{c}_H - \bar{c}_p + b_p e_p^*(c_t, \phi) + \phi) - e_p^*(c_t, \phi)^2 - \alpha \bar{c}_p = 0$. □
Curriculum Vitae

Name: Salar Ghamat

Post-secondary Education and Degrees:

Tabriz University
Tabriz, Iran

Bilkent University
Ankara, Turkey
2008-2011 M.B.A.

Tilburg University
Tilburg, The Netherlands
2010-2011 M.Sc.

Western University
London, Ontario, Canada
2012-2017 Ph.D.

Honours and Awards:

Richard Ivey Plan for Excellence Doctoral Fellowship, 2012-2016

Province of Ontario Graduate Scholarship, 2015 – 2016

John F. Rankin Doctoral Scholarship, 2015-2016

Al Mikalachki PhD Research Grant, 2015-2016

Ivey International Centre for Health Innovation Research Grant, 2016-2017

Related Work Experience:

Lecturer
Western University
2015-2016

Publications: