

Wilfrid Laurier University

Scholars Commons @ Laurier

Theses and Dissertations (Comprehensive)

2014

Modeling and Analysis of Value-Based Healthcare Delivery

Tannaz Mahootchi

Wilfrid Laurier University, hoss7980@mylaurier.ca

Follow this and additional works at: <https://scholars.wlu.ca/etd>



Part of the [Operations and Supply Chain Management Commons](#)

Recommended Citation

Mahootchi, Tannaz, "Modeling and Analysis of Value-Based Healthcare Delivery" (2014). *Theses and Dissertations (Comprehensive)*. 1664.

<https://scholars.wlu.ca/etd/1664>

This Dissertation is brought to you for free and open access by Scholars Commons @ Laurier. It has been accepted for inclusion in Theses and Dissertations (Comprehensive) by an authorized administrator of Scholars Commons @ Laurier. For more information, please contact scholarscommons@wlu.ca.

WILFRID LAURIER UNIVERSITY

DOCTORAL THESIS

**Modeling and Analysis of
Value-Based Healthcare Delivery**

Author:

Tannaz MAHOOTCHI

A thesis submitted in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy in Operations and Supply Chain Management

at

School of Business and Economics

August 2014

© Copyright by
Tannaz MAHOOTCHI
2014

Abstract

Modeling and Analysis of Value-Based Healthcare Delivery

by Tannaz MAHOOTCHI

Doctor of Philosophy in Operations and Supply Chain Management

WILFRID LAURIER UNIVERSITY

Dr. Ignacio CASTILLO, Supervisor

Dr. Logan MCLEOD, Co-Supervisor

Healthcare reforms are emerging in order to control the increasing healthcare expenditures and to improve the health outcomes. In the context of the “Value-based Healthcare Delivery” reform, Michael Porter defines value as a patient’s health outcome per dollar spent. Porter’s proposal is comprised of organizing care around a medical condition (or around patient segments for primary care). Specifically, care will be provided by a dedicated, multidisciplinary team of providers, an Integrated Practice Unit (IPU). The IPU is jointly accountable for the health outcomes of patients and the costs of providing care during the full cycle of care.

The main objective of this dissertation is to use analytics to determine enabling factors for the successful implementation of the value-based healthcare delivery reform. This dissertation consists of three core chapters.

Chapter 2 draws insights on the effects of current payment schemes, including fee-for-service, capitation, and pay-for-performance, in fulfilling the objectives of value-based healthcare delivery. Particularly, a mathematical representation of healthcare delivery is proposed to assess if any of the existing payment systems can incentivize providers to improve the quality and integrate the care simultaneously. The results provide insights on strengths, shortcomings, and applicability of each payment system in fulfilling value-based healthcare delivery objectives.

Chapter 3 determines the optimal payment system between the healthcare purchaser and the IPU. The current payment systems do not pay for health outcomes. Most importantly, they do not consider health outcomes over the care cycle and fail to provide dynamic incentives for the providers. This study investigates the

contract that can coordinate the healthcare purchaser-IPU relationship over the care cycle.

Chapter 4 studies the effects of different contractual arrangements on collaboration dynamics among the providers involved in an IPU. A mathematical representation that characterizes the relationship between the providers throughout the care cycle is proposed. When efforts are not contractible, the contractual agreement will determine the dynamics of the collaboration. Aside from characterizing the first-best solution, the effects of reward-sharing and relational contracts, together with traditional schemes, such as capitation, are studied in this chapter.

The results of this dissertation shed light on the enablers of the value-based healthcare delivery reform. This dissertation is the first to design a dynamic incentives contract between the healthcare purchaser and the IPU, who is accountable for the health outcomes of a patient over the care cycle. The optimal contract can coordinate the objectives of the purchaser and the IPU and maximize social welfare. In addition, this is the first study to characterize the collaboration dynamics among the IPU members under different contractual agreements. The insights from this study can strengthen the work relationship of the providers within an IPU.

To my mother, Shahnaz

Acknowledgements

Finishing my PhD studies has been, by far, one of the most unique experiences I have been through. Moments of joy, despair, smile, tears, ups, and downs were part of the every day challenge in this long journey, which makes the milestone of completing my PhD more meaningful. I would not have been able to accomplish this without all those who supported me in this journey. First and foremost, I would like to thank Dr. Ignacio Castillo. Ignacio gave me invaluable advice not only on my research but also he helped me build my academic career and character. In every moment of my PhD studies, Ignacio was the one who always believed in me. His optimism and patience kept me going even when I was facing failures. He kept repeating: This is research, Tannaz. Just keep up. It is very hard to put everything in words, but I have learnt from Ignacio that the things that matter are endurance, professionalism, right attitude, and good heart. I am truly indebted to Ignacio for his dedication, priceless coaching, and for my pleasant PhD experience.

During the past two years, I interacted with Dr. Logan McLeod and benefited from his advice on my research. I am thankful to Logan for his advice and support. I would like to thank my dissertation committee members: Dr. Sapna Isotupa and Dr. Mike Pavlin, with whom I have always felt comfortable to share my thoughts.

I would like to acknowledge all the students I had the opportunity to connect with while teaching at Laurier. Laurier students are a set of unique, talented, and congenial individuals. The joyous moments in the classroom as well as positive feedback from the students for the past two years of my PhD motivated me during challenging times (#TeamTannaz). The teaching opportunity reaffirmed my purpose of being in academia, to contribute to others lives, for which I started my PhD. I am also thankful to everyone at Laurier, especially to my friends and colleagues at Wilfrid Laurier Graduate Students Association, I loved the experience and am looking forward to living the life of meaning and purpose.

While living far away from my family, my friends made my PhD years joyous. I would like to acknowledge the presence and support of my friends during this journey. Firstly, I want to thank my high school friend Glareh Amirjamshidi. Our friendship started in high school, sitting on the same bench for four years, and

extended throughout years. We immigrated to Canada the same year, started our PhD together, and now finishing it together. We have wonderful and unforgettable memories. But most important of all, she has always been there for me, in midst of desperations, illnesses, and tears. Her friendship and presence is priceless to me. I cannot thank her enough. Secondly, I want to extend my special thanks to the friends who helped me when I suddenly found out that I was suffering from retinal detachment on the first year of my PhD studies. I am not sure, how I would be able to make it without their help. Special thanks to Dr. Hoda Parvin who drove from the US to London, ON to be there for my surgery, to her kind family for taking care of us when we were in London, to my dear friend Dr. Mehdi Amoui who has always been there for me for the past 10 years and also drove me there while he had a deadline to submit his PhD papers, to Dr. Saiedeh Razavi for being a cheerful companion during the stressful moments before the surgery, to Dr. Sahar Karimi for nursing me at her place, and to my most precious woman, my mother, Shahnaz Behnam, who made it to Canada urgently and took care of me for the two months that I was not able to see with my right eye. Finally, I would like to thank every single of my friends here in Waterloo, in Toronto, in Iran, and around the globe, whom their presence and inspiring lives have made me a better person, I am so grateful for all of them. Undoubtedly, they are my most valuable assets in life.

Greatest, sincerest, and dearest thanks to my mother, Shahnaz Behnam. I have no words that can describe her, and what she has done for my sister and I. I owe a debt of gratitude to Shahnaz khanoom! She is my role model in persistence, optimism, and unconditional love. All the credit for what I have done goes to my parents and how they cared about our education, for all the sacrifices they made for us to be in the best schools with elite students. My dad, Karim Mahootchi, with his sportsmanship spirit, has always encouraged me in my life to strive for the best.

Anyone who knows me is aware of how much my sister, Farnaz, means to me. She has been my best friend and companion in my lifetime. I absolutely adore her. Farnaz has been my true source of inspiration, to know that one can be bold and feminine, in love and serious, calm and accountable, forgiving and strong at the same time. Although just 15 months older than me, she has always been the

mature one. She has always been the first one to ride on the long and dark roads, lightening the road, and standing the gusty moments for me. There is not a way that I can thank her enough for her encouragement, for her friendship, and for her love and courage.

During my life, I wanted to have a brother. Ata, my brother-in-law, for the past six years has been like a brother to me, supportive, kind, serious, and objective. His unlimited love for my sister and for my family is one of the most pleasant experiences in my life. His passion for tennis brought us closer together and made me gradually like Rafael Nadal too. However, I still like and adore Roger Federer, hands down!

Lastly, I want to thank the objects/things that helped me go through my PhD studies. I would like to extend my sincere thanks to PhD Comics website and Jorge Cham. Witty comics and a daily dose of making fun of PhD life were helpful in keeping things light while working in my academic bubble. The PhD Comics website is a clear indication that the misery of PhD students is universal and mostly existential. I am very grateful to few musicians, including Chaartaar, Homayoun Shajarian, and Ludovico Einaudi, for their genius music that helped me take my mind off of earthly material and concentrate on the important heavenly stuff, aka. my Ph.D. Dissertation.

Contents

Abstract	ii
Acknowledgements	v
List of Figures	x
List of Tables	xi
1 Introduction	1
2 Value-Based Healthcare Delivery: A Principal-Agent Model	7
2.1 Introduction	7
2.2 Literature Review	14
2.3 The Model	19
2.4 The First-Best Solution	22
2.5 Hidden Action	25
2.5.1 Fee-for-Service	27
2.5.2 Capitation	29
2.5.2.1 Partial Accountability	30
2.5.2.2 Full Accountability	31
2.5.3 Performance-Based	33
2.6 Conclusions	36
2.A Appendix: Chapter 2 Proofs	38
2.A.1 The First-Best Solution	38
2.A.2 Fee-for-Service	40
2.A.3 Capitation	42
2.A.4 Performance-Based	44
3 Coordinating Contracts in the Value-Based Healthcare Delivery; Effect of Integration and Dynamic Incentives	47
3.1 Introduction	47
3.2 Literature Review	51

3.3	The Model	55
3.3.1	Objective Functions and the Contract Space	61
3.3.2	Incentive Compatibility Constraint	61
3.4	The Coordinating Contract	64
3.4.1	Properties of the Optimal Contract	65
3.5	Conclusions	70
3.A	Appendix: Chapter 3 Proofs	73
4	Healthcare Providers in an Integrated Practice Unit: Modeling Partnerships	76
4.1	Introduction	76
4.2	Literature Review	81
4.3	Partnership Modeling	85
4.4	The First-Best Solution	88
4.5	Double-Sided Hidden Action	90
4.5.1	Contractible Reward	91
4.5.2	Noncontractible Reward	96
4.6	Conclusions	102
4.A	Appendix: Chapter 4 Proofs	104
4.A.1	The First-Best Solution	104
4.A.2	Contractible Reward	108
4.A.3	Noncontractible Reward	113
5	Conclusions	120
	Bibliography	125

List of Figures

3.1	Problem state: patient status (W or U), number of failures (m), and number of successes (n); probability of failure is denoted by p^i .	57
4.1	Problem state: patient status “well” or “unwell” (W or U), number of failures (m), and number of successes (n); probability of failure p_j	85
4.2	State dynamics in working mode $j = \{H, L^p, L^\ell\}$	86
4.3	Dynamics of collaboration among the providers in an IPU; the first-best Solution	89
4.4	Dynamics of collaboration among providers in an IPU in a reward-sharing contract	95
4.5	Dynamics of collaboration among providers in an IPU in a relational contract	101

List of Tables

2.1	Quality of service and integration efforts and the working outcomes	20
2.2	Total surplus in each working mode	23
2.3	Providers' payoffs with fee-for-service payment scheme	28
2.4	Providers' payoffs with capitation payment scheme	31
2.5	Providers' payoff with capitation payment scheme and full accountability for complications	32
2.6	Payoff for the providers with performance-based payment scheme in each working mode	34
4.1	Providers' efforts and working relationship outcomes	86

Chapter 1

Introduction

Healthcare systems are organizations of people, resources, and institutions designed to provide healthcare services to meet the health needs of populations. Canadian healthcare expenditures accounted for 11.4% of the gross domestic product (GDP) in 2011 and were expected to reach an all-time high of \$211 billion in 2013 [1]. With growing expenditures, governments and policy makers are paying closer attention to health system performance measurement.

Healthcare systems are generally organized to deliver specific healthcare services from a provider to the population. For example, a patient with diabetes might go through different providers during his or her disease cycle, including a primary care physician, nutritionist, diabetes nurse, endocrinologist, nephrologist, ophthalmologist, psychiatrist/psychologist, social worker, laboratory, and dialysis provider. The providers then receive payment for the specific service they provide.

Current care delivery is generally structured around specialties. Each healthcare provider treats patients whom need care from the provider, sometimes with different medical conditions. Providers have traditionally been paid for the specific

services they provide, known as fee-for-service payment. The fee-for-service rates are usually set above marginal cost, intended to cover fixed costs through volume of services. By covering all expenses and rewarding providers based on volume, providers have little incentive to consider costs when treating their patients. Furthermore, by rewarding volume instead of better health outcomes for patients, the fee-for-service system might punish the efficient use of care [2].

To control the increasing healthcare expenditures and merely marginal health improvements, healthcare reforms are emerging [3]. Michael Porter has introduced the “Value-based Healthcare Delivery” reform, in which he defines value as a patient’s health outcome per dollar spent [4]. Porter argues that value is the only goal that can unite all stakeholders’ interests, because when providers succeed in delivering higher value, everyone (including patients, employers, health plans, and governments) will win through better health outcomes at lower costs [5]. Porter’s proposal is comprised of organizing care around a medical condition (or around patient segments for primary care). Specifically, care will be provided by a dedicated, multidisciplinary team who takes responsibility for the full cycle of care for the condition, including outpatient, inpatient, and rehabilitative care, as well as supporting services. This dedicated team of providers is called Integrated Practice Unit (IPU) and is jointly accountable for the health outcomes of patients and the costs of providing care. Costs will be measured around the patient and will be aggregated around the full cycle of care for the patient’s medical condition. As a result, the cost depends on the actual use of resources in a patient’s care process.

Michael Porter defines health outcomes as survival, extent of recovery or functionality restored, complications, recovery time, patient’s experience, and other

aspects of a patient's health. Accordingly, the payment system for the value-based healthcare delivery should be modified to pay for health outcomes over the care cycle.

It is worth noting that health is both demanded and produced by patients, therefore it depends on many factors including patients' compliance with the medical interventions, education and income level, and the quality of healthcare delivery [6]. Healthcare delivery is one of many inputs into the health production process. However, integrated care delivery including supporting services can arguably influence patient's behavior and compliance with the care process.

Several payment systems have been studied and implemented in the healthcare context [7–9]. Among these are fee-for-service, capitation, and pay-for-performance. As mentioned earlier, fee-for-service mechanism covers all costs borne by the provider plus a margin for each service they provide. Consequently, a fee-for-service system can encourage the overprovision of healthcare services [10]. In a capitation payment system, providers are paid an up-front fixed amount for each patient enrolled under their care for a set of services for a period of time. Capitation payment mechanism encourages providers to keep their patients healthy, but yet it creates incentives for physicians to systematically select patients who are healthier and require less care in the future, known as cream skimming. Further, capitation payments mainly cover primary care services and exclude specialty or hospital care. As a result, if a primary care physician refers a patient to a specialist or to a hospital, the primary care physician can keep the capitation fee and not need to provide the care. In summary, capitation payment could result in underprovision of care [11]. Neither fee-for-service nor capitation payment systems can incentivize the providers to improve health outcomes.

In response to quality issues, changes to the payment systems have been proposed to reward appropriate and high-quality care. This type of payment system is known as pay-for-performance. Proposals for implementing pay-for-performance payment systems vary from rewarding the providers for their processes (how things are done) to the patients' health outcomes (the effectiveness of treatments) [2]. Despite promising principles of pay-for-performance systems, their implementation faces several significant challenges [12]. Among these challenges is the problem of "multitasking," that is if the providers face several tasks and their resources are limited, then their effort will be allocated toward explicitly rewarded tasks. Tying remuneration to processes is administratively easy to implement, but might create unwanted results by encouraging the providers to concentrate on the processes that are incentivized under pay-for-performance and ignore the processes that are not. Additionally, pay-for-performance systems that pay for outcomes are challenging to implement in the current structure: when healthcare providers work individually, they cannot be held accountable for the health outcomes since each provider can argue the other providers involved in the well-being of a patient have not provided good quality of care [2].

The main objective of this dissertation is to use analytics to determine enabling factors for the successful implementation of a value-based healthcare delivery reform. This dissertation consists of three core chapters. Chapter 2 draws insights on the effects of current payment systems in fulfilling the objectives of value-based healthcare delivery. Chapter 3 determines the optimal payment system between the healthcare purchaser and the IPU. Chapter 4 studies the effects of different contractual arrangements on collaboration dynamics among the providers involved in an IPU.

Chapter 2 studies the effect of current payment schemes, including fee-for-service, capitation, and performance-based, on care provision. The effort each provider exerts on patients has two dimensions: quality and integration of care. Quality of care is defined as any aspect of care provision that improves the health outcomes for the patient. High quality care needs more resources, thus it is costly. Integration of care is to optimize and streamline the care process to ensure efficiency and effectiveness. Highly integrated care leverages the information about the patient's interrelated conditions, shares resources, and eventually results in cost reduction for the providers involved.

A mathematical representation of the healthcare delivery has been offered to assess if any of the existing payment systems can incentivize providers to improve the quality and integrate the care simultaneously. The results provide insights on strengths, shortcomings, and applicability of each payment system in fulfilling value-based healthcare delivery objectives.

Chapter 3 determines the optimal contract between the healthcare purchaser and the IPU during the care cycle. The current payment systems are not paying for health outcomes. Most importantly, they do not consider health outcomes over the care cycle and fail to provide dynamic incentives for the providers. This study investigates the contract that can coordinate the healthcare purchaser-IPU relationship over the care cycle. A dynamic continuous-time principal-agent model has been employed to characterize the optimal contract. The model considers the interactions between treatment strategy and health outcomes, where there is a single entity responsible for the health outcomes of a patient during care cycle.

Chapter 4 studies the potential partnership models among the healthcare providers within an IPU. Particularly, this chapter investigates the effects of contractual agreements on collaboration dynamics among the providers within an

IPU. The relationship between the providers throughout the care cycle is modeled using a finite-horizon dynamic game. When efforts are not contractible, the contractual agreement will determine the dynamics of the collaboration. Aside from characterizing the first-best solution, the effect of reward-sharing and relational contracts, together with traditional schemes, such as capitation is studied in this chapter. A relational contract is a combination of formal and informal contracts. The formal contract promises a fixed amount to the provider no matter what the health outcomes are. The informal contract includes a discretionary payment to the provider after observing the health outcome in each period.

The results of this dissertation shed light on the enablers of the value-based healthcare delivery reform. This dissertation is the first to design a dynamic incentives contract between the healthcare purchaser and the IPU who is accountable for the health outcomes of a patient over the care cycle. Current payment systems fail to incentivize the IPU to be accountable for the health outcomes during the care cycle. The optimal contract can coordinate the objectives of the purchaser and the IPU and maximize the social welfare. In addition, this is the first study to characterize the collaboration dynamics among the IPU members under different contractual agreements. The insights from this study can strengthen the work relationship of the providers within an IPU.

Chapter 2

Value-Based Healthcare Delivery: A Principal-Agent Model

2.1 Introduction

Healthcare costs have increased in Canada and most developed countries. In the latest report by the Canadian Institute for Health Information (CIHI), the total health spending in 2013 was projected to an all-time high of \$211 billion or \$5,988 per person [1]. Total health expenditure in Canada was 11.4% of gross domestic product (GDP) in 2011. It is forecast to be 11.3% in 2012 and 11.2% in 2013. The rate of growth in health spending is slowing for the first time in nearly 15 years. Healthcare costs are growing in many other developed countries as well. The rates of increase in spending for all the Organisation for Economic Co-operation and Development (OECD) countries, except for Iceland, were above the rates at which their respective economies grew from 2000 to 2011 [1].

Furthermore, costs are not the only concern in healthcare delivery systems. Disintegration in information flow and continuity of care, resulting in inconsistent patient services and unacceptable waiting times for patients, remains as a challenge for healthcare systems [13]. Other than advances and innovations in healthcare treatments and tools, improvements in healthcare delivery process is necessary as well. Optimizing and streamlining the delivery processes, efficient and effective decision making could result in better health outcomes.

In another report, CIHI examines the contributing factors to the growing expenditures and focuses on the boom period, between 1998 and 2008, when the annual health spending in Canada more than doubled [14]. This report studies the major cost drivers of public-sector healthcare spending, which accounts for 70% of the total health care costs in Canada over the last decade. The main contributing factors identified as compensation of healthcare providers, increased use of services, and an evolution of types of services provided and used. Albeit the shift in demographics, the aging of baby boomers only accounts for a modest portion in healthcare spending increases, i.e. 0.8% per year from 1998 to 2008.

CIHI's data demonstrates that the compensation paid to the healthcare providers is one of the most significant portions of public-sector healthcare spending, Hospitals (29.2%), drugs (15.9%), and physician services (14.4%) continue to account for the largest shares of health dollars [1]. Furthermore, in Canada, the decisions physicians make directly affect the use of drugs, hospital care, and diagnostic tests. Since the healthcare delivery structure affects the expenditures, policy makers are required to reform how healthcare is provided in order to ensure an efficient and effective system. CIHI recommends to increase the efficiency and effectiveness of the healthcare system, including the introduction of inter-professional collaboration to provide team-based care, increased focus on patient-centered care, and providing

incentives to healthcare providers to meet the needs of their patient population [14].

In Canada primary healthcare reform, undertaken by provinces, is calling for change in the structure of healthcare delivery [15]. The main concern of primary care reform is to improve health outcomes and access to primary care for patients, promote coordination among providers, and increase efficiency while improving patient and provider satisfaction. The shift is toward team-based primary care, in which the providers are accountable for providing comprehensive services to their clients.

In addition, the necessity of restructuring the healthcare system has been recognized in other countries. In a report by the American Institute of Medicine, the chasm between quality the U.S. healthcare system is capable of providing and the care patients are receiving has been documented [16] and can be attributed to the way healthcare is delivered and organized. Among other efforts, the U.S. has employed different programs and mechanisms to impact the cost and quality of care by using financial incentives to alter patient and provider behavior [17]. Examples include mechanisms and programs like pay for performance and disease management, which concentrate on the quality of the care and use of tools to improve patient health. Other stream of programs focuses on cost containment and increases patient's share in total cost of care, i.e. health savings account and consumer-driven health plans. The drawback of cost containment programs is where they raise conflicts. For example, when the patients are required to pay more for their health, they tend to use less of necessary care, and therefore the quality of care will be compromised [18].

One of the proposed reforms that can address the quest for both lower costs and better health outcomes, is Michael Porter's *value-based healthcare delivery* [4].

Porter argues health outcomes cannot be improved efficiently, unless the value of healthcare delivery is improved. Value in the healthcare context has been defined as patient health outcomes per dollar spent. Health outcome has been defined as survival, prevention of illness, early detection, right diagnosis, right treatment to the right patient, fewer avoidable complications, greater functionality, slower disease progression, and less care induced illnesses [5]. In order to enable improved value for the patients, changes should take place in different pillars of the system, including organizational structure, outcome and cost measurement, and reimbursement systems.

Currently, the healthcare system is organized based on specialties and reimbursement is being made to individual providers, that is “specialized healthcare delivery”. The other delivery system that supports value-based healthcare reform is called *Integrated Practice Unit* (IPU). IPU is a team of providers formed around medical conditions to provide all necessary care and is accountable for the health outcomes of a patient during disease care cycle. A medical condition is an inter-related set of patient medical circumstances that includes common co-occurring conditions and complications, and requires multiple specialties and services to best address the disease from the patient’s perspective. Care cycle for a medical condition like cancer may include: monitoring, diagnosing, preparing, intervening, recovering/rehabbing, and monitoring/managing. An IPU will have the capability and knowledge to provide the required care and the authority over the care process, which will incentivize the IPU to offer value for a patient. This type of healthcare delivery is in its developing phase in many countries including the U.S., Canada, Germany, Rwanda, Taiwan, Sweden.

In the healthcare delivery context, the healthcare purchaser (hereafter, she)

cannot provide the required treatments herself,¹ hence the presence of the healthcare provider (hereafter, he) becomes inevitable.² As a result, healthcare purchaser and provider enter into a contractual agreement in which the health care provider agrees to provide service to the covered population.

Given the current payment schemes, this chapter develops a principal-agent framework to analytically evaluate under what conditions *value-based healthcare delivery* can achieve its objectives of containing costs and improving quality. We will consider the case where the government is the healthcare purchaser and the healthcare delivery for a medical condition requires multiple providers with differing specialties and skills. When government is the purchaser, her objective is to improve social welfare. Specifically, the purchaser is concerned with who receives health services and whether they receive appropriate care to their condition. In our model, we will assume all patients have universal health insurance and therefore the price of healthcare services will not affect the utilization of health care. As a result, any individual would receive the necessary health care whenever they need it.

Since the government is the healthcare purchaser, we assume she is concerned with the *appropriateness* of care. Appropriateness of care could have many aspects, including receiving the right medical treatment and being treated in an understanding way [19]. In the literature, terms like intensity or quality have been used to capture the concept of appropriateness. In this chapter, we will refer to all these aspects of care as quality of service. Quality of service is defined as any aspect of service that benefits the patients whether during the process of treatment or after the treatment. Since the government is paying the providers from tax revenues, she is also attentive to the costs of care. To incorporate *appropriateness*

¹e.g. government or health insurer

²e.g. hospitals or physicians

and *cost-effectiveness* in our model we will define two dimensions for the care: quality and integration of care.

Both quality and integration of care will determine the outcomes of the care. Overprovision of services may reduce the benefits of care when unnecessary services or excessive treatments have harmful consequences [16]. Similarly, underprovision of services may reduce health benefits if the essential services are not provided [10]. Integrated care will result in optimized delivery, neither over-providing nor under-providing will occur. We assume when the care is integrated and the processes are streamlined, patients will benefit from better health outcomes. Integration of care could result in cost savings both by avoiding excessive harmful treatments and by preempting the avoidable complications. In other words, the integration of care and coordination among the providers will lead to enhanced health outcomes.

In the specialized healthcare delivery structure providers are organized based on their specialties and each provider might provide services to multiple different medical conditions that might require his expertise. In specialized healthcare delivery structure, providers will work individually. Working on their own may cause individual providers to overlook comorbidities and the interdependencies between interventions being provided somewhere else, which can increase the probability of avoidable complications and result in lower expected health benefits. In the IPU structure, efficiency gains may occur as a result of complementarities among the treatments. Moreover, the probability of complications will be diminished as a result of coordinated care.

There is supporting evidence showing that higher quality care drives down the long term costs [20]. The value-based healthcare delivery brings the outcomes and costs together, and reveals inefficiencies and opportunities to reallocate resource usage. The IPU has the authority to eliminate high cost activities which do not

lead to better health outcomes and identify activities with low costs corresponding to superior outcomes.

We will study how the current payment schemes, including fee-for-service, capitation, and performance-based payment, will direct the working mode among the providers. We will first determine the socially optimal delivery structure, i.e. the first-best solution, and then using the current payment schemes in the healthcare context we will determine under what conditions the objectives of value-based delivery will be met. Particularly, the results shed light on whether any of the payment schemes could fulfill the value-based delivery reform's objectives, encouraging cost efficiency and superior health outcomes at the same time.

Our results suggest that the performance-based payment system can achieve similar results to the first-best solution but, the applicability of performance-based payment system depends on the measurability and contractability of health outcomes. Payment systems like capitation can induce integration or cost reduction incentives, however, resulting in minimum quality provision. Low-powered contracts like fee-for-service can provide quality improvement incentives, nevertheless they are inefficient for the healthcare system. Capitation payment schemes can be modified to hold providers fully accountable for complication costs, which is also known as fundholding payment system [2]. When low quality provision results in higher expected costs of complication, capitation with full accountability payment system could outperform other contracts when rewards are not contractible.

To the best of our knowledge, this is the first study to analytically evaluate the consequences of current payment schemes on the healthcare delivery structures and evaluating their merit in fulfilling value-based delivery objectives.

The remainder of this chapter is organized as follows. We review the related literature in the next section and present the model in §2.3 . We characterize the first-best solution in §2.4 and study the effects of using fee-for-service, capitation, and performance-based contracts on the organizational structure in §2.5. All proofs appear in the Appendix.

2.2 Literature Review

This chapter builds upon the literature on contract theory, multi-dimensional task assignment, economics of organization, and payment schemes in the healthcare context.

There are two streams of research in contract theory. The first one focuses on hidden information, also known as adverse selection, where one party has better information than the other. An example of hidden information in healthcare delivery context is a case where the distribution of risk categories in a heterogeneous population is not observable by the healthcare purchaser. Hidden information studies in healthcare aim at finding incentive compatible contracts to reduce such inefficiencies [21, 22]. The second stream of research concentrates on the inefficiencies caused by hidden actions where one party's efforts are not verifiable by the other party.³ These situations happen when the healthcare purchaser cannot observe the provider's actions, treatment intensity, or the quality of the delivered care [7, 9, 22, 26, 27]. Monitoring the actions of providers is either too costly or not possible. When one party hires another party to take some action for her as her agent, there should be a contract among the entities that can mitigate

³In the contract theory literature, the term “moral hazard” has been used extensively to describe the problem where the efforts are not verifiable [23–25]. Here, however, we will use the hidden action term to denote the same problem.

difficulties cause by hidden action or hidden information. The contract design problem is known as *principal-agent problem* [24]. The problem of hidden action can be tackled by a contract that links the agent's performance to the output of his work. Inefficiencies can arise when the agent's performance measure is distorted, for example when it cannot incorporate time lags and interdependency.

In economics and game theory literature, when additional effort by one player increases the productivity and benefits of the higher efforts for the others, strategic complementarities exist [28]. Complementarities between diverse tasks have raised attention in the manufacturing literature as well [29, 30]. The literature on task assignment has shown the changing patterns in task allocations to employees in a variety of settings. Evidence exists both in intra-organization and inter-organization relationships. For example, the integration of marketing, billing, and repair departments into a single customer service department, and also engaging suppliers in the design of parts in the manufacturer-supplier relationships [29–31].

Principal-agent models have been extensively used to study the job design problems as well. One of the most relevant studies is the work by Holmstrom and Milgrom [32] in which they analyze the principal-agent model in a multi-task context. Their paper focuses on the multidimensional tasks and how the difficulty of measuring performance in all aspects can derive adverse affects in task accomplishment. Broad task assignments, i.e. assigning one agent the responsibility of broad range of tasks, can result in higher incentive costs. When complementarities are present, such incentive costs have been characterized as heterogeneity loss. Under the assumption that the performance of each task is measured separately, Holmstrom and Milgrom [32] also studied the possibility of task sharing. When tasks are divisible and independent and the performance measure can only assess one of the tasks, they have shown that one agent should perform the measurable

task, while the other agent should perform the tasks that are hard to measure. In healthcare delivery, each provider's tasks have multiple dimensions also collective efforts of the providers will determine the health outcomes for a patient.

In multi-task studies, the main assumption is that either several different tasks are being assigned to an agent, or the agent's tasks have several dimensions to it. Zhang [33] have analyzed the impact of complementarities in assignment of multiple tasks to agents. In his model, the principal chooses the optimal delegation of tasks to two identical agents, whom are capable of doing all the tasks on their own. When only the aggregate performance measures are available to the principal and when tasks are complementary, Zhang has shown that the broad task assignments results in more informative agent efforts and can overcome the heterogeneity loss.

When a better health outcome is what the purchaser is striving for, the system boundaries for the healthcare provision should change. Value-based healthcare delivery, defines a medical condition as the set of patient medical circumstances that includes common co-occurring conditions and complications, and requires multiple specialties and services to best address the disease from the patient's perspective [34]. Providers can either work individually, as it is in the current healthcare delivery, or they could team up to provide all necessary care for the medical condition. Using the notions of economies of scope [35], we assume providers could benefit from the economies of diversification because of complementarity of their skills and sharing the fixed costs of providing care to the patients [36].

Another important study is about the problem of hidden action within teams [37]. Hidden action exists when actions of an agent cannot be observed and contracted directly. In contrast to the single-agent models, even when there is no uncertainty in output level, hidden action may occur in teams. The reason is that, when the only available signal is aggregate output level, the agents who shirk

cannot be identified. Holmstrom [37] showed that non-cooperative behavior will result in an inefficient outcome when the output is fully shared among the agents. Considering a single period principal-agent problem, their model demonstrates the value of group incentives in eliminating the free-rider problem. However, implementation of group incentives requires the principal to enforce financial incentives, that is to impose penalties when output is wasted and provide bonuses when output is exceeded. Holmstrom expresses this function as the principal's primary role, and they call it breaking the budget-balance constraint [37].

In our model, we consider multiple players working toward a common goal of improving health outcomes. With the language of Holmstrom [37], we assume the providers will have a budget-breaker to impose the penalties among team members, eliminating the problem of double hidden action. This complies with the recommendations of value-based healthcare delivery, in which it is proposed there should be a care manager overseeing each patient's care process [20]. The distinguishing part of our study is that we are investigating the effects of current payment schemes on working mode among the providers, that is if current payment schemes to the team of providers can encourage both cost efficiency and quality improvement.

Several payment schemes have been proposed and implemented for the healthcare systems. Among these are fee-for-service, capitation, and pay-for-performance. For a comprehensive review of healthcare payment systems, including an empirical analysis, refer to [8, 38, 39]. We will briefly describe the literature on fee-for-service, capitation, and pay-for-performance payment schemes in this section.

Fee-for-service mechanism covers all costs borne by the provider plus a margin for each service they provide. Consequently, a fee-for-service system can encourage the overprovision of healthcare services [10]. In a capitation payment system,

providers are paid an up-front fixed amount for each patient enrolled under their care for a set of services for a period of time. Capitation payment mechanism encourages providers to keep their patients healthy, but yet it creates incentives for physicians to systematically select patients who are healthier and require less care in the future, known as cream skimming [2]. Further, capitation payments mainly cover primary care services and exclude specialty or hospital care. As a result, if a primary care physician refers a patient to a specialist or to a hospital, the primary care physician can keep the capitation fee and not need to provide the care. In summary, capitation payment could result in underprovision of care [11]. Neither fee-for-service nor capitation payment systems can incentivize the providers to improve health outcomes.

In response to quality issues, changes to the payment systems have been proposed to reward appropriate and high-quality care. This type of payment system is known as pay-for-performance. Proposals for implementing pay-for-performance payment systems vary from rewarding the providers for their processes (how things are done) to the patients' health outcomes (the effectiveness of treatments). Despite promising principles of pay-for-performance systems, their implementation faces several significant challenges [12]. Among these challenges is the problem of "multitasking," that is if the providers face several tasks and their resources are limited, then their effort will be allocated toward explicitly rewarded tasks. Tying remuneration to processes is administratively easy to implement, but might create unwanted results by encouraging the providers to concentrate on the processes that are incentivized under pay-for-performance and ignore the processes that are not. Many fundamental questions still are open. For example, what type of clinical conditions or healthcare services should be target of quality improvement incentives? To whom should such incentives be directed to: the patient, the

healthcare provider, the provider group, or hospital, or all of these parties? What type of measures should be rewarded: process of care, outcomes, or both? For a systematic review and evaluation of pay-for-performance systems see Petersen et. al. [40]. Additionally, pay-for-performance systems that pay for outcomes are challenging to implement in the current structure: when healthcare providers work individually, they cannot be held accountable for the health outcomes since each provider can argue the other providers involved in the well-being of a patient have not provided good quality of care [2].

In this chapter, we are not concerned with designing the optimal payment scheme, however we are interested to see how payment schemes can reinforce the objectives of value-based healthcare delivery.

2.3 The Model

We assume multiple providers are required to treat a patient with a specific medical condition. In the current healthcare delivery structure, the healthcare providers work individually in silos on each patient without coordinating the care. Lack of integration and coordination of care might result in suboptimal care provision by either underproviding or overproviding the services. Integrated care could result in better health outcomes through complementary skills, information, and also through the accumulation of experience. The value-based healthcare delivery reform intends to bring all the necessary service providers for a medical condition together to optimize the delivery of healthcare. The main tenets of this reform are based on integrating and optimizing the care for the patients with a specific condition to improve the health outcomes while containing the costs.

TABLE 2.1: Quality of service and integration efforts and the working outcomes

Quality \ Integration	High	Low
High	Duo	Sole ^q
Low	Sole ^e	None

We assume the health outcome, which is commonly observable, depends on the treatments each patient receives. Providers will choose the quality of service q and integration effort level e to be exerted on the patient. We consider two possible levels of quality and two possible levels of integration: high and low, denoted by H and L , respectively. As a result of this binary action sets in quality and integration aspects of care, four outcomes are possible, as depicted in table 2.1. If both high quality services and high integration efforts have been exerted, the outcome is referred to as *Duo*. If only quality of service or integration of care is at high level, the outcome is referred to as *Sole^q* or *Sole^e*, respectively. When both quality and integration are at low level, the outcome is referred to as *None*.

Note when providers integrate the care, they have the opportunity to focus on optimizing the total cost of the care cycle rather than minimizing the costs of discrete services.

The total monetary cost of treatments consists of fixed cost F and variable cost c . The variable cost for the providers depends on the quality of treatments q and integration efforts e . The variable cost increases with the quality of treatment and decreases with integration efforts. Therefore, the cost of healthcare delivery c^j depends on the working outcome $j = \{d, s^q, s^e, n\}$ representing *Duo*, *Sole^q*, *Sole^e*, and *None*, respectively. When the integration efforts are at high level, the working outcome is assumed to be efficient because the providers are focusing on streamlining the care and optimizing the delivery process, as a result $c^{s^e} < c^n$ and $c^d < c^{s^q}$. Furthermore, the high quality service costs more than low quality

service. As a result, $c^d > c^{s^e}$ and $c^{s^q} > c^n$. With the above assumptions, the ordering of costs for all the working modes is $c^{s^q} > c^d > c^n > c^{s^e}$.

Similarly, the fixed cost for the providers in working mode j is F^j . However, we will assume the fixed cost will only depend on the level of integration. When providers employ low level of integration, they are working individually and therefore, the fixed cost of providing care is the same for the providers at working mode *Sole^q* and *None*, $F^{s^q} = F^n$. We will denote the fixed cost of working at low integration level with F_s . Furthermore, the fixed cost of working in full integration is equal as well, $F^d = F^{s^e}$ and will be denoted by F_c . The fixed cost of integration might appear as hindrance if the cost of establishing the team of providers is greater than working individually, $F_c > F_s$. However, after the initial establishment of the team, the providers will have the opportunity of consolidating their resources. Eliminating the redundant administrative and scheduling resources might result in cost savings for all the providers involved in the integrated unit, $F_c \leq F_s$. As a result, if integration of care results in lower fixed cost for all the providers in comparison to them working individually $\Delta F = F_c - F_s \leq 0$. However, if integration is more costly, then $\Delta F = F_c - F_s > 0$.

The health benefit for the patients also depends on the working outcome j , where $j = \{d, s^q, s^e, n\}$ representing *Duo*, *Sole^q*, *Sole^e*, and *None*, respectively and will be denoted by b^j . We assume health benefits are in dollar units, representing the value of health outcomes to the healthcare purchaser.

In this chapter, we are not concerned with how to measure the benefits from provision of different services. The benefits of the healthcare service could encompass multiple dimensions including survival, degree of health achieved or retained, time to recovery, disutility of care process, and complications and recurrences. In the value-based healthcare delivery setting, health outcome improvement has been

defined as prevention of illness, early detection, right diagnosis, right treatment to the right patient, fewer avoidable complications, greater functionality, slower disease progression, and less care induced illnesses [5].

Research has shown that the excessive utilization of services does not produce better health outcomes [20]. Furthermore, fragmentation of services could cause diseconomies of scale, where inadequate volume of patients in a medical condition has hindered the accumulation of knowledge and achievement of excellence in the care [20]. As a result, we assume the health benefit increases with the quality of services and integration efforts. When providers work collaboratively and integrate the care, their skills are complementing each other and the care process is more effective. As a result, $b^d > b^{s^q}$ and $b^{s^e} > b^n$. Plus, patients will always prefer high quality service to low quality service, $b^{s^q} > b^n$ and $b^d > b^{s^e}$. Therefore, the ordering of benefit for the patients for all the working modes is $b^d > b^{s^q} > b^{s^e} > b^n$.

We assume perfect information about the patient and the costs. When efforts are verifiable by a third party, the parties can write a contract to maximize the total surplus. We refer to this situation as the first-best solution. When efforts are not contractible, the first-best contract is generally not attainable, resulting in the situation of hidden action.

2.4 The First-Best Solution

In this section, we study the first-best working mode for the team of providers. Consider the first-best solution as the benchmark solution, as if the all of providers efforts were verifiable at no cost. In this case, the purchaser would determine the working mode so as to maximize the total surplus. To determine the first-best

TABLE 2.2: Total surplus in each working mode

Quality \ Integration	High	Low
High	$b^d - c^d - F_c$	$b^{s^q} - c^{s^q} - F_s$
Low	$b^{s^e} - c^{s^e} - F_c$	$b^n - c^n - F_s$

working mode, we need to identify the objective function for the purchaser and providers. The payoff for both purchaser and provider is dependent on the working mode.

The healthcare purchaser transfers payment t to the providers according to a predetermined arrangement. As a result, healthcare purchaser's objective is,

$$b^j - t.$$

The healthcare providers would receive t in compensation while spending c^j and F^j for providing treatments in the working mode j ,

$$t - c^j - F^j.$$

As a result, the total surplus is,

$$b^j - c^j - F^j.$$

Recall, according to the cost and benefit structure, $c^{s^q} > c^d > c^n > c^{s^e}$ and $b^d > b^{s^q} > b^{s^e} > b^n$, we will determine which working mode would be the first-best in each situation. The next observation provides a full characterization of first-best working mode.

Observation 2.1. *When integration is effective i.e. $c^{s^q} > c^d > c^n > c^{s^e}$ and $b^d > b^{s^q} > b^{s^e} > b^n$,*

(i) whenever $\Delta F < 0$

(a) working mode Duo or Sole^e would always dominate and therefore, integration of care is always optimal.

(b) Duo will be the optimal working mode if $b^d - b^{s^e} > c^d - c^{s^e}$.

(c) if the marginal benefits of providing high quality treatment is smaller than its marginal costs, it is always optimal to work in Sole^e mode.

(ii) whenever $\Delta F \geq 0$

(a) integration would only be beneficial if health benefit exceeds the total cost, or when

$$b^d - b^{s^q} > c^d - c^{s^q} + \Delta F$$

and

$$b^{s^e} - b^n > c^{s^e} - c^n + \Delta F.$$

(b) if (a) holds, Duo will be the optimal working mode if $b^d - b^{s^e} > c^d - c^{s^e}$.

(c) the necessary conditions for integration of care are when

$$c^{s^q} - c^d > \Delta F$$

and

$$c^n - c^{s^e} > \Delta F$$

both hold.

2.5 Hidden Action

In this section, we analyze the following situation: the healthcare purchaser offers a contract to the healthcare providers to supply care to the patient population. The providers choose quality of service q and integration effort level e , which in turn affect the health benefits for the patients. Exerting efforts and providing high quality service is costly to the providers, and the purchaser has to compensate the providers for incurring these costs. We assume the efforts are not verifiable to the purchaser, thus not contractible, and the providers might deviate from the socially optimal solution in order to maximize their own payoff. Since efforts are not verifiable, we are facing the hidden action problem. In this chapter, we assume that the providers whom are working in a team can verify the other providers' efforts and therefore we will not face the multisided hidden action problem. Multisided hidden action arises when several agents take complementary hidden actions [25, 37].

Using similar assumptions to the standard hidden action problem, we assume all parties are symmetrically informed about costs, patient risk factors, and health outcomes [41]. The health outcomes are verifiable, hence contractible. The healthcare purchaser observes the health outcomes and will pay providers a transfer payment t according to a predetermined arrangement. Once the payment scheme is determined, the healthcare providers simultaneously take action. The actions are completed before any performance measures are realized. Therefore, the actions cannot be changed when the health outcomes have been observed. These assumptions are common assumptions in the context of principal-agent model [24].

Even though there may be other sources of incentives for the providers, such

as reputation, we will only consider monetary incentives in our model. As mentioned before, the health outcome depends on some parameters that are beyond the providers' control including compliance to the treatment, income level, education level, and family history. However, health outcomes critically depend on the collective level of effort providers exert on the patient. In other words, the quality of service as well as level of integration will determine the health benefits.

We will focus on the single period model. In particular, there is only one period in which providers take action. As a result, the effect of feedback and dynamics is not studied in this chapter. The contract will govern the one period game among the providers and purchaser.

We analyze the resulting working mode for the providers using different payment schemes in the healthcare contracting literature. Specifically, we will study the effects of fee- for-service, capitation, and performance-based payment schemes on the optimal working mode for the providers. Assuming the healthcare purchaser wants to maximize the value of health outcome, the purchaser's problem is

$$\max_j b^j - t, \quad (2.1)$$

where the money transfer to the providers is denoted by t . The provider is also concerned about his financial surplus either because it is a for-profit organization or because the surplus can be spent on improving the facilities. Therefore, the problem is subject to the providers' incentive compatibility constraint

$$\arg \max_j t - c^j - F^j \quad (2.2)$$

and their participation constraints:

$$t - c^j - F^j \geq w, \quad (2.3)$$

where c^j and F^j is respectively, the variable and fixed cost of working in mode $j = \{d, s^q, s^e, n\}$ and w is providers' reservation wage. Particularly, reservation wage is the amount of money providers could get if they worked in another healthcare delivery structure. For simplification, we will assume $w = 0$.

2.5.1 Fee-for-Service

One of the most commonly used payment schemes is the *cost-based* payment scheme in which the providers are fully reimbursed for all costs of medical services provided to the patient. The provider's reimbursement varies with the actual treatment decisions or the resource use [39]. Under this payment scheme, the providers have weaker incentives to reduce the costs or resource usage.

We will compare the payoffs for each working mode, given that the providers will be reimbursed for the cost of treatments. To control the costs the purchaser usually caps the fees at a certain level based on the treatment type. Furthermore, in fee-for-service payment system, the fees are usually set based on services provided and since the collaboration efforts are not verifiable, we will assume that the payments for providing high quality care is denoted by t^H and the payments for providing low quality care is denoted by t^L .

The providers' payoffs for each working mode, given the fee-for-service payment scheme, are given in table 2.3. Providers will choose their desired working mode based on the total payoff in each working mode.

TABLE 2.3: Providers' payoffs with fee-for-service payment scheme

Quality \ Integration	High	Low
High	$t^H - c^d - F_c$	$t^H - c^{s^q} - F_s$
Low	$t^L - c^{s^e} - F_c$	$t^L - c^n - F_s$

Observation 2.2 describes the optimal working mode for the providers, when they are reimbursed according to fee-for-service payment scheme and when integration is effective, $c^{s^q} > c^d > c^n > c^{s^e}$ and $b^d > b^{s^q} > b^{s^e} > b^n$.

Observation 2.2. *When providers are reimbursed according to fee-for-service payment scheme and $t^d = t^{s^q} = t^H$ and $t^{s^e} = t^n = t^L$*

(i) *If $\Delta F < 0$, it is always optimal for the providers to integrate the care and benefit from the cost savings they create.*

(ii) *If $\Delta F \geq 0$, integration is optimal when both*

$$c^{s^q} - c^d > \Delta F$$

and

$$c^n - c^{s^e} > \Delta F$$

hold.

(iii) *when either (i) or (ii) holds, working mode Duo is optimal if $t^H - t^L > c^d - c^{s^e}$.*

When $\Delta F \geq 0$ the conditions for integration using fee-for-service payment, are similar to the necessary conditions of integration in the first-best solution. But, with the fee-for-service payment system when $c^{s^q} - c^d < \Delta F$ and $c^n - c^{s^e} < \Delta F$, integration will not occur, even though it is still optimal to integrate the care in the first-best solution. Recall, from the first-best solution that the integration

would be optimal if

$$b^d - b^{s^q} > c^d - c^{s^q} + \Delta F$$

and

$$b^{s^e} - b^n > c^{s^e} - c^n + \Delta F$$

hold. With fee-for-service payment scheme, the providers could deviate or not deviate from the efficient levels of quality. Basically, the fee-for-service payment does not incentivize the providers provide efficient level of quality. Even though fee-for-service payment system could implement the first-best solution, given the conditions in observation 2.2, the cost savings from integration are not transferred to the purchaser. Even though it is not socially optimal, the purchaser still could gain from the higher expected benefits for the patients if the providers work in *Duo* mode.

2.5.2 Capitation

In order to provide cost-reduction incentives to healthcare providers, the purchaser might prefer a *capitation* payment scheme in which the providers receive a fixed payment per patient for a given time period. In return, the provider is required to provide care to the patients for a given period of time without any additional reimbursements. To incorporate capitation in our model, we assume providers receive a fixed fee t^C in each period for each patient. With capitation payment system, the providers are encouraged to keep patients as healthy as possible. Since the providers' income depends on how much care their patients need, the providers might choose the healthier patients whom require less care, also known as *cream skimming*. To eliminate this effect, we assume providers are dealing with a homogeneous patient population.

We will consider two scenarios on who is accountable for the complication costs. In the first scenario, which we call *partial accountability*, the purchaser reimburses the providers for any adverse health outcomes. However, in the second scenario, we assume the providers must provide care to the patient for a given period without any marginal reimbursement. We call the second scenario *full accountability*, which is also known as *fundholding* system in healthcare contracting literature [2]. In this case, the providers will be encouraged to prevent any avoidable complications. We will denote the expected costs of complications by v^j given working mode $j = \{d, s^q, s^e, n\}$. We assume if providers work in integration and provide high quality care, the expected cost of complication will be the least, we assume the following ordering for the expected cost of complication $v^d < v^{s^q} < v^{s^e} < v^n$.

2.5.2.1 Partial Accountability

First, let us assume the purchaser pays for any adverse health outcomes for the patients, including hospitalization. In this case the healthcare purchaser's objective is to maximize

$$b^j - t^C - v^j,$$

where t^C is the money transfer to the providers in capitation, and v^j is the expected costs of complications given working mode $j = \{d, s^q, s^e, n\}$. The problem is subject to the providers' incentive compatibility constraints

$$\arg \max_j t^C - c^j - F^j,$$

and the providers' participation constraints

$$t^C - c^j - F^j \geq w.$$

TABLE 2.4: Providers' payoffs with capitation payment scheme

Quality \ Integration	High	Low
High	$t^C - c^d - F_c$	$t^C - c^{s^q} - F_s$
Low	$t^C - c^{s^e} - F_c$	$t^C - c^n - F_s$

Recall, w is the providers' reservation wage if they worked in another healthcare delivery system. Again, for simplification we will set $w = 0$. The summary of providers' payoffs is given in table 2.4.

Observation 2.3. *When the purchaser is responsible to pay for any adverse health outcome expenses; the case of partial accountability*

- (i) *If $\Delta F < 0$, it is always optimal to work in Sole^e mode.*
- (ii) *However when $\Delta F > 0$*
 - (a) *if $c^{s^q} - c^d > \Delta F$, Sole^e is the dominating working mode.*
 - (b) *if $c^{s^q} - c^d < \Delta F$, None is the dominating working mode.*
 - (c) *if $c^{s^q} - c^d = \Delta F$, either Sole^e or None will dominate.*

When better health outcomes involve higher costs, the providers set the effort levels to the lowest level. As a result, capitation with *partial accountability* payment system cannot fulfill the objectives of value-based delivery.

2.5.2.2 Full Accountability

The second scenario is when the providers are fully accountable for health outcome. In this case, if any adverse health outcomes occur, it is the responsibility of the providers to treat the patients using the allocated money, t^C . As a result, the expected cost of complication for each patient in each period v^j will be paid out

TABLE 2.5: Providers' payoff with capitation payment scheme and full accountability for complications

Quality \ Integration	High	Low
High	$t^C - c^d - v^d - F_c$	$t^C - c^{s^q} - v^{s^q} - F_s$
Low	$t^C - c^{s^e} - v^{s^e} - F_c$	$t^C - c^n - v^n - F_s$

of t^C . Recall, v^j only depends on the working mode j . The healthcare purchaser's objective is to maximize

$$b^j - t^C$$

subject to the providers' incentive compatibility constraints

$$\arg \max_j t^C - c^j - v^j - F^j,$$

and their participation compatibility constraints

$$t^C - c^j - v^j - F^j \geq 0.$$

The summary of payoffs for the capitation payment with full accountability for the health outcomes is given in table 2.5. Observation 2.4 summarizes the working outcomes with capitation payment scheme and full accountability for the complications.

Observation 2.4. *When the providers are accountable for the complication costs:*

(i) *If $\Delta F < 0$, it is always optimal to work in Duo mode.*

(ii) *However when $\Delta F > 0$*

(a) *if $-(v^d - v^{s^q}) > c^d - c^{s^q}$, Duo will dominate all other working modes.*

(b) *otherwise, working mode Sole^q dominates.*

Observation 2.4 suggests that the providers will always work in collaboration offering high quality treatments if the fixed costs of collaboration is lower than fixed costs of individual service provision. The reason for this effect is because providers' income is linked to the benefits for the patients. Better health outcomes and benefits for the patients correspond to lower complication costs. When high quality service is offered in integration it will result in lower expected costs, and therefore working mode *Duo* is optimal. In contrast, when collaboration has more fixed costs than individual work the providers will only work in *Duo* mode if the total expected cost of care in *Sole^q* mode exceeds the total expected cost of care in *Duo*. When providers are responsible for the complication costs, the providers' compensation should increase to satisfy the providers' participation constraint, i.e. $t^C \geq c^d + v^d + F_C$. Even though the purchaser is incurring v^d , she can motivate the providers to work in *Duo* mode. Since $v^d < v^{s^q} < v^{s^e} < v^n$, the purchaser is paying the minimum cost of complication instead of v^{s^e} or v^n , which was the result of capitation with partial accountability. As a result, capitation with full accountability payment encourages the providers to work in *Duo* mode instead of working in *Sole^e* or *None* mode which is costlier to the purchaser. Expected complication costs in *Duo* mode, could be interpreted as unavoidable complication costs.

2.5.3 Performance-Based

Health outcomes are the result of collective efforts of all providers involved in the care process. Healthcare purchaser evaluates health outcome based on some performance measure. We assume this performance measure is a single aggregate measure that depends positively on each provider's efforts. Under the assumptions stated in §2.3 about the expected health outcome and costs, integration and quality

TABLE 2.6: Payoff for the providers with performance-based payment scheme in each working mode

Quality \ Integration	High	Low
High	$t_{PB}^d - c^d - F_c$	$t_{PB}^{s^q} - c^{s^q} - F_s$
Low	$t_{PB}^{s^e} - c^{s^e} - F_c$	$t_{PB}^n - c^n - F_s$

are complementing each other. That is, the providers are inclined to exert high quality treatment when they work in collaboration and the care is integrated.

One of the main insights from Holmstrom's study on moral hazard in teams is that the principal can provide optimal incentives to multiple agents involved in the process by acting as the budget-breaker and giving shares of aggregate output to the agents [37]. In this section, we will study the preferred working mode under the performance-based contracts. According to these type of contracts the providers will receive a fixed fee β and a share α of the output. In the case of healthcare delivery, providers will receive transfer payment $t_{PB}(b^j)$ dependent on the health outcome b^j . We denote this payment scheme with t_{PB}^j , and define it as

$$t_{PB}^j = \alpha b^j + \beta,$$

where b^j is the health outcome in working mode j . Since the expected health outcome depends on both the quality of care and the level of integration, expected payoff for the providers also depends on both dimensions of the care.

In the performance-based contract, the purchaser's objective is to maximize $b^j - t_{PB}^j$. However, each provider wants to maximize his payoff while satisfying the participation constraint. The payoffs for each of the working modes are listed in table 2.6. We are assuming that β is the same for all the working modes.

Observation 2.5. *When the providers are paid based on the health outcomes attained:*

(i) If $\Delta F < 0$, integration is always optimal.

(a) Duo is optimal only if $\alpha(b^d - b^{s^e}) > c^d - c^{s^e}$.

(b) Otherwise, $Sole^e$ will dominate all other working modes.

(ii) However, when $\Delta F > 0$:

(a) Integration is optimal if both

$$\alpha(b^d - b^{s^q}) > (c^d - c^{s^q}) + \Delta F$$

and

$$\alpha(b^{s^e} - b^n) > (c^{s^e} - c^n) + \Delta F$$

hold.

(b) if (a) holds, Duo is optimal only if $\alpha(b^d - b^{s^e}) > c^d - c^{s^e}$.

(c) The necessary conditions for integration are when

$$c^{s^q} - c^d > \Delta F$$

and

$$c^n - c^{s^e} > \Delta F$$

hold.

Working mode outcomes under performance-based contracts are in general similar to the first-best solution. However, the purchaser pays $\alpha b^j + \beta$ to the providers to induce similar results to the first-best solution. Rewarding or punishing the providers based on the health outcomes encourages them to exert high quality treatments in integration. This situation belongs to the class of noncooperative games named *supermodular games*. In this type of models, more activity

by some members raises the return to the increased levels of activity by others [28, 42]. The same situation applies here to the two dimensions of care, when care is integrated the payoffs will be higher if high quality care is provided as well. Even though performance-based payment system could create similar results to the first-best solution, its applicability depends on the contractability of health outcomes.

2.6 Conclusions

Currently, healthcare delivery is challenged by increasing expenditures and less than optimal health outcomes. Value-based healthcare delivery reform intends to improve health outcomes while containing the costs [5]. Michael Porter argues that organizational structure, outcome and cost measurement, and reimbursement systems should be modified to encourage value provision. In this study, we investigate the effect of current payment systems including fee-for-service, capitation, and performance-based on the objectives of value-based healthcare delivery. Specifically, we determined conditions under which the providers would be encouraged to work collaboratively to integrate the care and provide high quality treatments.

The results suggest that the performance-based payment scheme can achieve similar results to the first-best solution, but the applicability of this payment scheme depends on the measurability and contractability of health outcomes. Contracts like capitation can induce integration or cost reduction incentives, however results in minimum quality provision. Contracts like fee-for-service can support high quality treatments but may not result in lower costs. Thus, fee-for-service is not an efficient way of achieving high quality treatment. Capitation payment system can be modified to include the possibility of complications, When the

purchaser holds the providers fully accountable for the avoidable complications, providers can be motivated to provide high quality service and integrate the care. When low quality provision results in higher expected costs of complication, capitation with full accountability can outperform other contracts when the health outcomes are noncontractible.

Fixed cost of care affects the results critically. When integration results in cost savings for the providers $\Delta F < 0$, the providers could be incentivized to work in collaboration and to provide high quality care. However, if the fixed cost of providing care in all the working modes is the same, our notation of fixed cost could be interpreted as the nonmonetary cost of care. In other words, ΔF could be used to describe a situation where the providers enjoy working collaboratively, $\Delta F < 0$ or where the providers find collaboration ineffective, $\Delta F \geq 0$. With this interpretation, the current results of the model could be applied to draw insights on the non-monetary cost of care.

To the best of our knowledge, this is the first study to analytically evaluate the consequences of current payment schemes on the healthcare delivery structures and evaluating their merit in fulfilling value-based delivery objectives. We built a simple model to determine what is the resulting organizational structure given the current healthcare payment system.

This research can be extended in few directions. First, since health outcomes are only observed after care has been provided and some of outcomes will only be observable after a time lag, it is crucial to study the healthcare delivery system considering a dynamic programming approach. Currently, we have used a static model to evaluate the working mode outcomes, however this model could be improved to consider the dynamics of the relationship over the care cycle. Second, the reputation incentives could be incorporated in the model, when the purchaser

infer information about the benefits of future care from observing current health outcomes. The current model could be modified to include the behavioral effects. By adding non-monetary cost of integration or quality provision to the model, we can analyze how the effectiveness of teamwork or the care providers' altruistic motives affect the results.

2.A Appendix: Chapter 2 Proofs

2.A.1 The First-Best Solution

We will start the analysis when $\Delta F < 0$. Under the assumption that $c^d < c^{s^q}$ and $b^d > b^{s^q}$, the RHS of

$$b^d - b^{s^q} > c^d - c^{s^q} + \Delta F$$

is always negative and the above equation holds. In this case, the working mode *Duo* would always dominate *Sole^q*. With the same argument, since $c^{s^e} < c^n$ and $b^{s^e} > b^n$, if $\Delta F < 0$, *Sole^e* dominates *None* as well, i.e. $b^{s^e} - b^n > c^{s^e} - c^n + \Delta F$ will always hold.

Since the fixed costs of working in *Duo* and *Sole^e* are assumed to be equal, with $b^d > b^{s^e}$ and $c^d > c^{s^e}$, the providers would always prefer *Duo* if

$$b^d - b^{s^e} > c^d - c^{s^e}, \quad (2.4)$$

otherwise, if $b^d - b^{s^e} < c^d - c^{s^e}$, working mode *Sole^e* will be optimal.

Now we will explore the first-best working mode when the fixed cost of integration is higher than working individually, $F_v > F_s$ or $\Delta F > 0$. In this case,

providing integrated care would be beneficial if

$$b^d - b^{s^q} > (c^d - c^{s^q}) + \Delta F \quad (2.5)$$

and

$$b^{s^e} - b^n > (c^{s^e} - c^n) + \Delta F \quad (2.6)$$

hold. When both equations (2.5) and (2.6) hold, the high quality treatment would be preferred to the low quality treatment only if the marginal benefit of high quality treatment exceeds its marginal cost, or when

$$b^d - b^{s^e} > c^d - c^{s^e}. \quad (2.7)$$

Consequently, working mode *Duo* is optimal only if equations (2.5), (2.6), and (2.7) hold.

Since $b^d > b^{s^q}$, if the RHS of the equations (2.5) and (2.6) is negative, i.e.

$$c^d - c^{s^q} + \Delta F < 0$$

and

$$c^{s^e} - c^n + \Delta F < 0$$

both equations (2.5) and (2.6) will hold and integration is optimal. As a result, the necessary conditions for integration exist when

$$c^{s^q} - c^d > \Delta F$$

and

$$c^n - c^{s^e} > \Delta F.$$

2.A.2 Fee-for-Service

As with the first-best analysis, we will start by assuming that integration is effective and its fixed cost could be less than the fixed cost of providing care individually, $F_v < F_s$ or $\Delta F < 0$. First, we will compare the working modes *Duo* and *Sole^q*. As a result, if

$$t^H - c^d - F_v > t^H - c^{s^q} - F_s$$

$$c^{s^q} - c^d > \Delta F$$

Duo would dominate *Sole^q*. Assuming that the providers' integration efforts can result in cost reduction, $c^d < c^{s^q}$, the LHS is always negative and the above equation always holds. As a result, when $\Delta F < 0$ the collaborative work would always be optimal. The argument is similar for comparing integration in providing low quality service as well. Since $c^{s^e} < c^n$, when $\Delta F < 0$

$$c^n - c^{s^e} > \Delta F$$

will always hold. In summary, if integration efforts are efficient, the providers will always prefer to work in integration. This is the same condition that was required to achieve the first-best solution as well.

To determine if providers would prefer providing high quality treatment or low quality treatment, we need to compare the payoffs at *Duo* and *Solo^e* modes

$$t^H - c^d - F_c > t^L - c^{s^e} - F_c$$

Consequently, if

$$t^H - t^L > c^d - c^{s^e} \tag{2.8}$$

the providers would always prefer providing high quality care. Nonetheless, if $t^H - t^L = c^d - c^{s^e}$ the providers have no incentive to not deviate from the efficient level of quality, although no incentive to deviate either, which might not be socially optimal given that $b^d > b^{s^e}$.

The purchaser prefers high quality treatments if

$$b^d - t^H > b^{s^e} - t^L$$

$$b^d - b^{s^e} > t^H - t^L \quad (2.9)$$

The providers might prefer providing high quality care, equation (2.8), while the purchaser does not consider that as the optimal mode, when the expected benefit of high quality treatment is marginal, $b^d - b^{s^e} < t^H - t^L$. This is a confirmation to the criticism that the providers would prefer overproviding care even if it is not socially optimal.

If integration has higher fixed cost than working alone $F_v \geq F_s$ or $\Delta F > 0$, given the fee-for-service payment scheme the providers would prefer working individually if

$$c^{s^q} - c^d < \Delta F$$

and

$$c^n - c^{s^e} < \Delta F$$

hold, although this might not mean that it is socially optimal to work individually. In other words, the latter conditions would not necessarily translate to

$$b^d - b^{s^q} < (c^d - c^{s^q}) + \Delta F$$

and

$$b^{s^e} - b^n < (c^{s^e} - c^n) + \Delta F.$$

2.A.3 Capitation

We will first analyze the case where the purchaser is responsible to pay for any adverse health outcome expenses. The providers' payoffs are given in table 2.4. For the integration to be rewarding,

$$t^C - c^d - F_v > t^C - c^{s^q} - F_s$$

and

$$t^C - c^{s^e} - F_v > t^C - c^n - F_s$$

should hold. In other words,

$$c^{s^q} - c^d > \Delta F \tag{2.10}$$

and

$$c^n - c^{s^e} > \Delta F. \tag{2.11}$$

If the above equations hold, the providers will always prefer the low quality treatments.

In the case where $\Delta F < 0$, since $c^d < c^{s^q}$ and $c^{s^e} < c^n$, equations (2.10) and (2.11) will always hold, therefore integration is optimal. However, since $c^{s^e} < c^d$, the provider will always prefer the low quality treatments and working mode $Sole^e$ will dominate all other working modes.

In the case where $\Delta F > 0$, if equations (2.10) and (2.11) are valid, with the same arguments, *Sole^e* dominates other working modes. However, if

$$c^{s^q} - c^d < \Delta F$$

working mode *None* will dominate. If $c^{s^q} - c^d = \Delta F$ either *None* or *Sole^e* is the dominant work mode. In all the above cases, this type of capitation payment results in low quality treatments.

However, if the providers are responsible for the avoidable costs of complications, results will change. From the perspective of the providers, the integration is optimal if

$$(c^{s^q} - c^d) + (v^{s^q} - v^d) > \Delta F \quad (2.12)$$

and

$$(c^n - c^{s^e}) + (v^n - v^{s^e}) > \Delta F \quad (2.13)$$

hold. If $\Delta F < 0$, both equations (2.12) and (2.13) will always hold, and therefore it is always optimal to work in integration. In this case, to evaluate providers' choice between high and low quality treatments, we have to compare the payoffs in working modes *Duo* and *Sole^e*. If the payoff in working mode *Duo* is higher than that of *Sole^e*,

$$t^C - c^d - v^d - F_c > t^C - c^{s^e} - v^{s^e} - F_c$$

or

$$v^{s^e} - v^d > c^d - c^{s^e}, \quad (2.14)$$

the providers would prefer *Duo*. As a result, when $\Delta F < 0$, the providers would prefer high quality treatments to low quality treatments if equation (2.14) holds.

This means that the providers would prefer higher quality treatments, if the marginal expected cost of complications providing low quality treatments is greater than the marginal cost of providing high quality treatments.

However, when $\Delta F > 0$, the collaboration will be optimal if both equations (2.12) and (2.13) hold. In the case that collaboration is not preferred, i.e.

$$(c^{s^q} - c^d) + (v^{s^q} - v^d) < \Delta F \quad (2.15)$$

and

$$(c^n - c^{s^e}) + (v^n - v^{s^e}) < \Delta F, \quad (2.16)$$

the providers' choice between high quality treatments and low quality treatments will depend on the payoffs in working modes *Sole^q* and *None*. Providing high quality treatments individually, i.e. *Sole^q*, would be preferred if both equations (2.15) and (2.16) hold and also

$$v^n - v^{s^q} > c^{s^q} - c^n.$$

2.A.4 Performance-Based

For the providers to prefer integration, the following two conditions must hold.

$$\alpha b^d + \beta - c^d - F_c > \alpha b^{s^q} + \beta - c^{s^q} - F_s$$

and

$$\alpha b^{s^e} + \beta - c^{s^e} - F_c > \alpha b^n + \beta - c^n - F_s.$$

In summary, if

$$\alpha(b^d - b^{s^q}) + (c^{s^q} - c^d) > \Delta F \quad (2.17)$$

$$\alpha(b^{s^e} - b^n) + (c^n - c^{s^e}) > \Delta F \quad (2.18)$$

integration is optimal. Providing high quality treatments in collaboration would be the preferred option if

$$\alpha(b^d - b^{s^e}) > c^d - c^{s^e} \quad (2.19)$$

also holds.

If $\Delta F < 0$, the RHS of equations (2.17) and (2.18) is always negative. Given the cost and benefit structure described in section 3, the LHS of equations (2.17) and (2.18) is always positive and therefore both equations will always hold. As a result, if $\Delta F < 0$, collaboration is optimal. However, if $\Delta F > 0$, no working mode will be dominant in all circumstances. The working mode choice will depend on the expected payoff. If equations (2.17), (2.18), and (2.19) hold, working mode *Duo* will dominate all other choices.

Nevertheless, if neither equation (2.17) nor (2.18) holds, then any working mode could be dominant.

Since $b^d > b^{s^q}$, if

$$c^d - c^{s^q} + \Delta F < 0$$

and

$$c^{s^e} - c^n + \Delta F < 0$$

both equations (2.17) and (2.18) will hold and integration is optimal. As a result, the necessary conditions for integration are when

$$c^{s^q} - c^d > \Delta F$$

and

$$c^n - c^{s^e} > \Delta F.$$

Of course, the analysis is sensitive to the cost and benefit structure. If collaboration does not result in cost reduction for providers, i.e. $c^d < c^{s^q}$ and the expected health benefits are not any different than working individually, $b^d \leq b^{s^q}$, the collaboration might not dominate other working modes.

Chapter 3

Coordinating Contracts in the Value-Based Healthcare Delivery; Effect of Integration and Dynamic Incentives

3.1 Introduction

Achieving good health outcomes is the core objective of healthcare systems. Improving healthcare systems' performance depends on the healthcare provision, human and physical resources, financing, and setting and enforcing rules and strategic direction for all the entities involved [3]. Currently, Canada along with other developed countries are facing rising healthcare expenditures while they are challenged to improve the healthcare system's performance [1].

One of the reforms that addresses the current challenges in the healthcare delivery system is Michael Porter's *Value-based Healthcare Delivery* [5], which defines value as a patient's health outcome per dollar spent. Porter argues that value is the goal that can unite all stakeholders' interests, because when providers succeed in delivering higher value, everyone (including patients, healthcare providers, health plans, and governments) will win through better health outcomes for lower costs.

Historically, cost-effectiveness analyses and cost-utility analyses have been performed to examine the costs and outcomes of alternative medical strategies or programs. Even though there are similarities among cost-effectiveness analyses, cost-utility analyses, and value-based healthcare delivery, the first two are usually used to compare alternative programs with common health outcomes or to assess the consequences of expanding a program, but the latter intends to derive better health outcomes while minimizing the costs by requiring changes in different pillars of healthcare delivery, including organizational structure, outcome measurement, and reimbursement systems. Value-based healthcare delivery reform requires changes to the structure of the healthcare delivery so health outcomes could be defined around patient needs. Outcomes are multidimensional, and they are comprised of survival, extent of recovery or functionality restored, mistakes, complications, recovery time, patient's experience, and other aspects of a patient's health.

Michael Porter suggests that for specialty care, outcomes should be measured for each medical condition or set of interrelated patient medical circumstances, such as asthma, diabetes, or breast cancer [5]. A medical condition encompasses common complications, coexisting or co-occurring conditions. For primary and preventive care, outcomes should be measured for defined patient populations

with similar health circumstances, such as healthy adults, disabled elderly people, or adults with defined sets of chronic conditions. Furthermore, the outcomes that matter are the results of care over the disease cycle, rather than only results of an individual intervention or a single visit.

In the current healthcare delivery structure, providers can only measure what they can directly control in a particular intervention and what is easily measured rather than what matters for health outcomes. As a result, health outcomes are measured for providers, departments, or billing units rather than for the full cycle of care. If the healthcare delivery structure changes to a multidisciplinary group of providers for a medical condition, called *Integrated Practice Unit* (hereafter, IPU), then providers are able to jointly accept the responsibility for outcomes.

To incentivize providers to work in the proposed organizational structure and be accountable for health outcomes over the care cycle, payment systems should be modified. It is not usually feasible for the purchaser to determine in advance precise levels of all the aspects of care for every condition. It is also not possible to determine the delivery specifications have met ex post or not. The terminology used in literature is that the IPU's efforts are *unverifiable* to a third party. In this case, IPU's actions are hidden from the purchaser. The hidden action is also known as moral hazard [24]. Similar to Chalkley et al. [19] we use the term *contract* to refer to any arrangement for payment between the healthcare purchaser and the IPU. We assume the government is the healthcare purchaser with the objective of maximizing health outcomes while containing the costs. The contract that can *coordinate* the healthcare purchaser-IPU relationship will be called the "optimal contract" or "coordinating contract". In other words, coordinating contract is a contract that allow each party to optimize their objective function while maximizing the social welfare.

Several payment options have been proposed and implemented for healthcare systems over the past few decades. These payment systems affect physicians behavior, the quality of care received by patients, and healthcare costs. The current payment systems do not target quality of care and cost reducing efforts simultaneously. Fee-for-service payment could result in inefficiencies by providing financial incentives for the physicians to encourage overprovision of care [10]. Alternative payment system that can control the costs is capitation. However, with capitation payment scheme the provider's income depends on the assigned patient population's healthcare usage, capitation might result in less accessible care or less than efficient quality of care [2].

In this chapter, we study the coordinating contract between a healthcare purchaser and an IPU over the care cycle under dynamic hidden action. This research belongs to the rich literature on the continuous-time dynamic hidden action (moral hazard) problem. Our main contribution is to capture the health outcomes over the care cycle and link the payments to the history of health outcomes by using a continuous-time dynamic principal-agent model. We assume the IPU is protected by limited liability, but assumed to be risk-neutral. The IPU's efforts affect the health outcomes by modifying the probability of failures. High quality treatment is costly to the IPU and unverifiable to the purchaser.

The optimal payments to the IPU should be a function of the entire history of health outcomes. Using a continuous-time dynamic principal-agent model, we are able to summarize this complex history dependence by one state variable: *continuation value* of the IPU. Continuation value reflects the likelihood of future payments to the IPU and its evolution is the mirror of health outcome dynamics. As a result, continuation value could serve as IPU's track record.

The results suggest that the purchaser can motivate the IPU by promising a lump sum money transfer after a good performance record and also by threatening to reduce his continuation value after failures. The magnitude of the reduction in the IPU's continuation value is determined based on the incentive compatibility constraint. The more serious the hidden action problem is, the greater the reduction in the continuation value. With this mechanism, IPU's continuation value will be sensitive to the failures. Additionally, the healthcare purchaser's value function is concave in the IPU's continuation value and, consequently, punishing the IPU by decreasing his continuation value is costly for the purchaser. We show that it is socially optimal to set the reduction of the IPU's continuation value to the minimum level, consistent with the incentive compatibility constraint.

The remainder of this chapter is organized as follows. We review the related literature in the next section. We detail the model and formulate a principal-agent model to determine a set of incentive-compatible coordinating contracts to be offered to the IPU in §3.3. We characterize the optimal contract and explain its implementation in §3.4, and discuss future extensions and conclusions in §3.5.

3.2 Literature Review

This research connects the literature on healthcare contracting and continuous-time dynamic principal-agent models. The literature on the healthcare contracts have been reviewed in Chapter 2. Here we highlight important features and shortcomings of the current popular payment schemes including fee-for-service, capitation, and pay-for-performance.

Payment scheme, such as fee-for-service, is easy to implement and manage, but can be a source of inefficiency in the system since the scheme only rewards the volume of care and not the health outcomes associated with it. Empirical evidence suggest physicians that are paid by fee-for-service provide more consultation and order more diagnostic tests than the physicians that are not paid by fee-for-service [10].

While a capitation payment scheme encourages physicians to keep their patients healthy, it creates incentives for physicians to systematically select patients who are healthier and require less care in the future. Further, capitation payments mainly cover primary care services and exclude specialty or hospital care. As a result, if a primary care physician refers a patient to a specialist or to a hospital, the primary care physician can keep the capitation fee and not need to provide the care. Physicians may underreport patients' illness to them or even not reveal all possible treatments [43, 44]. This may lead to inefficiently low levels of care. None of these two payment schemes, i.e. fee-for-service and capitation, incentivize physicians to improve health outcomes, which is the what patients and policy-makers want. Not surprisingly the issues of poor health outcomes are ample in healthcare systems. Unacceptable health outcomes and increasing healthcare expenditures are pushing governments to examine quality and provide suggestions for its improvement [14, 16].

In response to quality issues, changes to the payment systems have been proposed to reward appropriate and high-quality care. This type of payment system is known as pay-for-performance. Proposals for implementing pay-for-performance payment systems vary from rewarding the providers for their processes (how things are done) to the patients' health outcomes (the effectiveness of treatments). Despite promising values of pay-for-performance systems, their implementation faces

several significant challenges [12]. Among these challenges is the problem of “multitasking,” that is if the providers face several tasks and their resources are limited, then their effort will be allocated toward explicitly rewarded tasks. Tying remuneration to processes is administratively easy to implement, but might create unwanted results by encouraging the providers to concentrate on the processes that are targeted and ignore the processes that are not. On the other hand, pay-for-performance payments tied to patients’ outcomes are not easy to implement. First, measuring health outcomes is not trivial. Second, in the current healthcare delivery structure, there is no way to isolate individual provider’s contribution to the patients’ health outcomes [2].

This research contributes to the body of healthcare contracting by designing a dynamic contract between the healthcare purchaser and team of providers whom are collectively responsible for the health outcomes of patients over the care cycle. Several authors have captured the importance of using dynamic contracts. Among these, Radner [45] shows it is possible to achieve efficiency in long-term contracts by aggregating outcomes over several periods. The aggregation would allow to form better statistics about the agent’s efforts, only if the agent becomes more patient. In another study, Fudenberg et al. [46] show in several settings, it is possible to use agent’s wealth as a proxy for agent’s performance history to implement the optimal contract. The insight from these studies suggest that a firm’s financial slack can summarize past performance. The firm’s management team is typically modeled as one agent. Our study similarly models the IPU as one agent in the principal-agent model.

Characterization of the optimal contracts in dynamic settings is a challenging task. Describing the state of the problem is complex. The agent’s compensation can be a function of entire performance history. Furthermore, the principal-agent

problem is composed of one dynamic optimization problem embedded in another. The principal is optimizing her objective, realizing that the agent is looking for the optimum dynamic effort strategy to maximize his objective as well. Our research belongs to the growing literature on dynamic hidden action (moral hazard) problems that employs recursive techniques.

Solving a discrete-time dynamic principal-agent model problem is a daunting task and requires several assumptions to derive tractable results. In the context of healthcare delivery, Fuloria et al. [27], find an *outcome-adjusted* payment that maximizes societal welfare. Such payment schemes can potentially make significant improvements, however the implementation of the resulted contract requires accurate information about the treatment technology, patient characteristics, and the provider preferences [27]. In that paper, Fuloria and colleagues assume the patient will be treated out of the system for any occurrences of complications. Moreover, to derive a tractable solution they make several assumptions such as exponential utility function and unrestricted access to a bank for the provider. In contrast, the method we use to characterize the optimal dynamic contract originates from the literature on continuous-time contracting, specially [47] and [48]. In continuous-time the solution can be characterized by ordinary differential equations, using optimal stochastic control. This method benefits from tractability, due to the differential equation that characterizes the optimal contract, and computing power. Continuous-time principal-agent model has been extensively studied in corporate finance applications, for examples see [49–52].

3.3 The Model

Health could be broadly defined as longevity and illness-free days in a given year [6]. Good health has positive social value for two reasons. First, health provides positive utility for patients. Second, better health increases the total amount of time available for market and nonmarket activities [6]. Health outcomes depend on many factors, including medical care. Grossman defines “The Human Capital Model” to demonstrate the importance of many inputs that go into the production of health along with medical care. He relates the output of health to choice variables like diet, exercise, medical care utilization, healthy habits, education, and also the medical care the patient receives from the providers [6].

In our model we will control for the patient risk factors. We assume the IPU is treating a homogeneous patient population. Furthermore, we are assuming the population benefits from a universal health insurance and if they need care there are no monetary obstacles for utilizing the health services. Therefore, we will assume the health outcomes critically depend on the appropriateness or quality of the care provided. Appropriateness of care can have many aspects, including receiving the right medical treatment and being treated in an understanding way [19]. In the literature, terms like intensity or quality have been used to capture the concept of appropriateness. In this chapter, we will refer to all these aspects of care as quality of service. Quality of service is defined as any aspect of service that benefits the patients whether during the process of treatment or after the treatment.

There are two players in this problem, the healthcare purchaser and the IPU. Recall that an IPU is a team of providers formed around medical conditions to provide all the necessary care and is accountable for the health outcomes of a

patient during the disease care cycle. A medical condition is an interrelated set of patient medical circumstances that includes common co-occurring conditions and complications, and requires multiple specialties and services to best address the disease from the patient's perspective. We will assume that an IPU has unique necessary skills to treat the patients. However, treatments are costly and the IPU has limited liability. By contrast, government as the healthcare purchaser has unlimited liability and is able to cover the costs. The two players are risk-neutral. Time is continuous and treatments are provided over the care cycle, T . The purchaser discounts the future payoffs at rate r and the IPU discounts the future payoffs at rate $\gamma > r$, i.e. IPU is less patient than the purchaser. This assumption rules out the possibility of indefinitely postponing the payments to the IPU. Without loss of generality, we normalize the fixed cost of forming an IPU to 0.

Health outcomes are the results of care in terms of patients' health over time. They are different than the care process designed to achieve the results and the biologic indicators that predict the results [5]. Health outcome has been defined as survival, prevention of illness, early detection, right diagnosis, right treatment to the right patient, fewer avoidable complications, greater functionality, slower disease progression, and less care induced illnesses [5].

Although we are aware of the multi-dimensionality of the health outcome, in this research we will focus on one of its dimensions to demonstrate the coordinating contract between the purchaser and the IPU that can improve value provision. Similar to Grossman [6], we will use the "number of illness-free days in a given year" as the indicator of health outcomes.

We assume the IPU is treating a homogenous patient population with similar risk factors and thus we can fairly assume the value of functional days in dollars

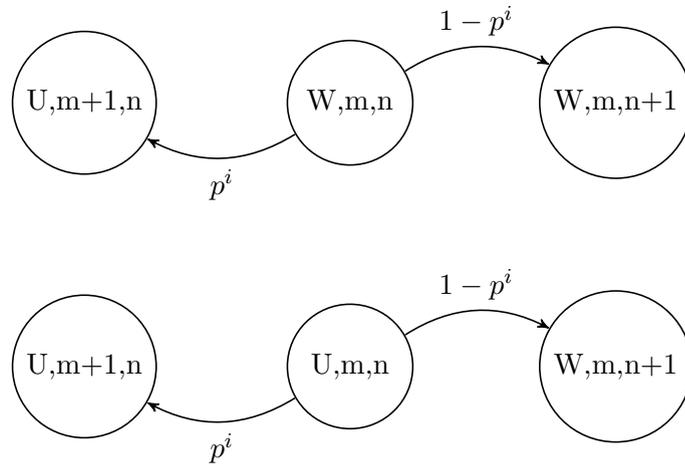


FIGURE 3.1: Problem state: patient status (W or U), number of failures (m), and number of successes (n); probability of failure is denoted by p^i

for each patient is μ , where $\mu > 0$ is a constant. We define the health status of a patient either as being “well” or “unwell”. The success of the IPU is defined as keeping the patient in the “well” state or bringing them back to the “well” state and the failure as being in the “unwell” state or transitioning back to the “unwell” health status.

The IPU can mitigate the risk of failures by choosing the quality of treatment, a_t . For simplicity, we will only consider two levels of quality, high and low, respectively denoted by a^H and a^L . When high quality care is provided, the probability of failure is denoted by p^H and when low quality treatment is provided the probability of failure is denoted by p^L . Assuming that attentive care reduces the probability of failure then $p^H < p^L$. The transition between problem states is shown in figure 3.1. The problem state is comprised of the patient health status (being “well” or “unwell” denoted by W or U , respectively), the number of failures is denoted by m , and the number of successes is denoted by n . The probability of failure and success for the IPU is denoted by p^i and $1 - p^i$, respectively where $i \in \{H, L\}$.

This construction helps us to connect the number of successes for the IPU to

the number of illness-free days for the patient. Since success for the IPU is the number of illness-free days for the patient, next we will reduce the problem state to the number of failures for the IPU.

The occurrence of failures is modeled as a point process $N = \{N_t\}_{t \geq 0}$, where for each t , N_t is the number of failures up to and including time t . The healthcare delivery failures impose two types of costs to the society. First, any complication and illness needs to be treated by the providers, whom will use valuable resources and capacity to serve the patients suffering from complications. Furthermore, there are costs tied to the lost working days. As a result, the number of complications can represent the inefficiencies in the healthcare system. We will assume the cost of health failure is C . Because of the limited liability, the IPU cannot be held responsible for costs exceeding his wealth. As a result, the purchaser incurs the complication costs and the net value of health outcome during the infinitesimal time interval $(t, t + dt]$: $\mu dt - CdN_t$.

We assume the IPU incurs a constant cost per unit of time, h . As a result, the total cost for treatments will be denoted with ha^H and ha^L for providing high quality and low quality treatments, respectively. The IPU can save $h(a^H - a^L)$ if he choose low treatments over high treatments.

Throughout this chapter we assume that

$$\mu - p^H C > 0 \tag{3.1}$$

and that

$$(p^L - p^H)C > h(a^H - a^L) \tag{3.2}$$

The left-hand side of (3.1) is the net expected health outcome value when IPU exerts high quality treatment. Condition (3.1) implies that a high quality treatment has positive net present value. We will focus on the case when the IPU's treatment choice is not verifiable but the health outcomes and costs are common information. Hence we are dealing with the hidden action problem. The left-hand side of (3.2) is the expected social cost of increased risk when IPU provides low quality treatment instead of high quality treatment. The right-hand side of (3.2) is the cost savings from providing low quality treatments. Condition (3.2) implies that in the absence of hidden action problem, it is socially optimal to require the IPU to provide high quality treatment.

The larger the cost savings from providing low quality treatment $h(a^H - a^L)$ is, the more attractive it is for the IPU to shirk. The lower $p^L - p^H$ is, the more strenuous it is to detect shirking. The contract between the healthcare purchaser and IPU is designed and agreed on at time 0. The IPU then chooses the treatment strategy process $A = \{a_t\}_{t \geq 0}$. We assume both purchaser and IPU can fully commit to a long-term contract. Our assumption about limited liability is in line with DeMarzo and Sannikov [49], Sannikov [47], and Biais et al. [51] where limited liability reduces the ability to punish the agent. This encourages the principal to replace punishments by actions. In this problem, we will assume the contract can be terminated as a result of steady adverse health outcomes. The stopping time or termination time will be represented by τ . We allow $\tau \leq T$, where T is the care cycle.

A contract specifies payments to the IPU and termination decision as functions of the history of past health outcomes. The cumulative money transfers to the IPU is nonnegative and increasing. The money transfers will be denoted by the process $S = \{s_t\}_{t \geq 0}$ and $s_t = 0$ for all $t > \tau$.

At anytime t prior to termination, the sequence of events during the infinitesimal time interval $(t, t + dt]$ can be described as follows:

1. The IPU takes her treatment strategy decision a_t^i , where $i \in \{H, L\}$.
2. With probability $p^i dt$, there is a failure, in which case $dN_t = 1$; otherwise $dN_t = 0$.
3. The IPU receives a nonnegative money transfer ds_t .
4. The treatments are either terminated or continued.

The payment and termination decisions are taken after observing the health outcomes. Formally, S is \mathcal{F}^N -adapted and τ is an \mathcal{F}^N -stopping time, where $\mathcal{F}^N = \{\mathcal{F}_t^N\}_{t \geq 0}$ is the filtration generated by N .¹ Informally, Martingale is a stochastic process defined on a probability space whose predicted value at any future time $u > t$ is the same as its present value at time t of prediction. The filtration \mathcal{F}_t^N contains all the information generated by N up to time t in an increasing sequence. A random variable S is called adapted to \mathcal{F}^N , if it “casually” depends on N [54]. This means that at each time t , S depends on the observation of the process N at time t . Lastly, τ is called an \mathcal{F}^N -stopping time if the decision to terminate the process or not depends on the information available from N at time t .

An effort process A will generate a unique probability distribution and we will use the expected payoffs for each player to demonstrate the dynamic principal-agent model. Next we will explain the objective function for the IPU and health-care purchaser.

¹For definitions of these concepts, for instance see [53] Chapter IV, Definitions 12, 49, and 61.

3.3.1 Objective Functions and the Contract Space

Given a contract $\Gamma = (S, \tau)$ and a treatment process A , the expected discounted payoff for the IPU is

$$\mathbb{E} \left[\int_0^\tau e^{-\gamma t} (ds_t - ha_t dt) \right] \quad (3.3)$$

while the expected discounted payoff of the purchaser is

$$\mathbb{E} \left[\int_0^\tau e^{-rt} (\mu dt - CdN_t - ds_t) \right] \quad (3.4)$$

Treatment strategy A is incentive compatible with respect to contract Γ if it maximizes the IPU's expected payoff (3.3). The healthcare purchaser's problem is to find a contract Γ and an incentive compatible treatment strategy A that maximizes expected discounted payoff (3.4), subject to fulfilling the IPU's required expected discounted payoff level. We will focus on the contracts Γ such that the present value of payments to the IPU is finite, that is

$$\mathbb{E} \left[\int_0^\tau e^{-\gamma t} ds_t \right] < \infty \quad (3.5)$$

Constraint (3.5) would assure the purchaser's expected discounted payoff is not infinitely negative.

3.3.2 Incentive Compatibility Constraint

Similar to the techniques introduced by Sannikov [47], we will employ martingale techniques to characterize the incentive compatibility constraint. We will similarly utilize the notion of *continuation value* for the IPU. Using the agent's continuation

value as a state variable is common technique in dynamic principal-agent models, see e.g. Spear and Srivastava [55] for an illustration.

The IPU will evaluate how the treatment strategy will affect her continuation value when taking a decision at time t . The IPU's continuation value is defined as

$$w_t(\Gamma, A) = \mathbb{E} \left[\int_t^\tau e^{-\gamma(u-t)} (ds_u - ha_u du) | \mathcal{F}_t^N \right] \quad (3.6)$$

Denote by $W(\Gamma, A) = \{w_t(\Gamma, A)\}_{t \geq 0}$ the IPU's continuation value process. Since $W(\Gamma, A)$ reflects whether there was a failure at time t , it is \mathcal{F}^N -adapted. To characterize the evolution of IPU's expected value, we will first consider his life time expected payoff, evaluated conditionally on the information available at time t , that is,

$$\begin{aligned} v_t(\Gamma, A) &= \mathbb{E} \left[\int_0^\tau e^{-\gamma(u)} (ds_u - ha_u du) | \mathcal{F}_t^N \right] \\ &= \int_0^t e^{-\gamma u} (ds_u - ha_u du) + e^{-\gamma t} w_t. \end{aligned} \quad (3.7)$$

Since $v_t(\Gamma, A)$ is the expectation of a random variable conditional on \mathcal{F}_t^N , the process $V(\Gamma, A) = \{v_t(\Gamma, A)\}_{t \geq 0}$ is an \mathcal{F}^N -martingale. Next, we will introduce a notation M_t^A that will represent the number of failures up to and including time t , minus its expectations.

$$M_t^A = N_t - \int_0^t p^i du \quad (3.8)$$

Since the occurrence of failures is modeled as a point process, according to a basic result from the theory of point processes M^A is an \mathcal{F}^N -martingale as well. The martingale representation theorem for point processes then implies the following lemma.

Lemma 3.1. *The martingale $v_t(\Gamma, A)$ satisfies*

$$v_t(\Gamma, A) = v_0(\Gamma, A) - \int_0^t e^{-\gamma u} \psi_u(\Gamma, A) dM_u^A \quad (3.9)$$

for all $t \geq 0$, for some \mathcal{F}^N -predictable process $\Psi = \{\psi_t(\Gamma, A)\}_{t \geq 0}$.

Lemma (3.1) along with (3.8) suggest that the lifetime expected value of the IPU evolves in response to the jumps of the process N . The term $-dM_t^A$ reflects the failures occurring at time t , the difference between the instantaneous probability of a failure $p^i dt$ and the instantaneous change in the total number of failures dN_t , which is equal to 0 or 1. As a result, $-e^{-\gamma t} \psi_t(\Gamma, A) dM_t^A$ can be interpreted as a function of past, in which $\psi_t(\Gamma, A)$ is the sensitivity of IPU's continuation value to the failures.

Lemma (3.1) reflects the fact that at any time t , the change in $v_t(\Gamma, A)$ is equal to $-e^{-\gamma t} \psi_t(\Gamma, A) dM_t^A$. Equations (3.7) and (3.9) imply that the IPU's continuation value evolves according to

$$dw_t(\Gamma, A) = (\gamma w_t(\Gamma, A) + h a_t^i) dt + \psi_t(\Gamma, A)(p^i dt - dN_t) - ds_t \quad (3.10)$$

for all $t \in [0, \tau)$. Equation (3.10) explains the expected instantaneous change in the IPU's continuation value. Since the parameter ψ_t is measuring the sensitivity of the IPU's continuation value to the failures, whenever the health outcomes features an unexpected failure dM_t , the IPU's continuation value changes by $\psi_t dM_t$. We can think of w_t as what the purchaser owes to the IPU. Using a similar analysis to Sannikov [47], we have

Proposition 3.1. *Given the contract $\Gamma = (S, \tau)$, a necessary and sufficient condition for the treatment strategy a^H to be incentive compatible is that*

$$\psi(\Gamma, A) \geq \ell, \quad (3.11)$$

for all $t \in [0, \tau)$, where $\ell = h \frac{a^H - a^L}{p^L - p^H}$.

Equation (3.10) shows that the IPU's continuation value will be instantaneously reduced by an amount $\psi_t(\Gamma, A)$ if there is an unanticipated failure. With this explanation, Proposition 3.1 states that to induce high quality treatments the reduction in the IPU's continuation value should be greater than the cost savings that the IPU can generate by shirking. Furthermore, because of the limited liability constraint, the IPU's continuation value must remain nonnegative. The continuation value of the IPU before observing the events at time t , $w_{t-}(\Gamma, A)$ should be greater than the loss in case of failure $\psi_t(\Gamma, A)$, therefore

$$w_{t-}(\Gamma, A) \geq \psi_t(\Gamma, A) \quad (3.12)$$

for all time $t \in [0, \tau)$. To simplify the notation, we drop the arguments Γ and A in the remainder of the chapter.

3.4 The Coordinating Contract

In this section, we characterize the optimal contract that induces high quality treatment, that is $A = a^H$ for all $t \in [0, \tau)$. The optimal contract maximizes the expected value for the purchaser from an incentive compatible contract to implement a high quality treatment strategy.

This section will offer more precise insights on how to induce high quality treatments that will result in better health outcomes at minimal cost. The contract that we derive can be described based on the continuation value of an IPU, reflecting future payment decisions. We will first provide the heuristic derivation of the purchaser's value function and the main properties of the optimal contract. Then we will show the formal derivation of the value function and optimal contract characteristics. All the proofs are provided in Appendix 3.A.

3.4.1 Properties of the Optimal Contract

Since the healthcare purchaser strives to maximize the value of health outcomes subject to the IPU's incentive compatibility constraint, formally we have

$$\begin{aligned}
 J(w) &= \max_{\{\psi_t, s_t, \tau\}} E \left[\int_0^\tau e^{-rt} (\mu - cdN_t - ds_t) | F_0 \right] \\
 dw_t &= (\gamma w_t + ha_t + \psi_t p^h) dt - ds_t - \psi_t dN_t \\
 \psi_t &\geq \ell
 \end{aligned} \tag{3.13}$$

We will first characterize the healthcare purchaser's continuation value $J(w)$ which is a function of the state of the problem, the IPU's continuation value w . Since the purchaser discounts the future payoffs at rate r , the expected flow of value at time t is given by $rJ(w)$. This should be equal to the sum of the expected instantaneous value of health outcomes and the expected rate of change in the principal's continuation value. The former is the expected health outcomes minus the expected payments to the IPU, which is

$$\mu - p^H C - ds_t. \tag{3.14}$$

To construct the expected rate of change in the purchaser's continuation value, we use the dynamics of the IPU's continuation value (3.10) and set $A = a^H$. The purchaser's continuation value satisfies the Hamilton-Jacobi-Bellman equation:

$$\begin{aligned} rJ(w) = & \sup_{s_t, \psi_t} (\mu - p^H C - ds_t) \\ & + [\gamma w_t^- + ha^H + p^H \psi_t - ds_t] J_w(w_t^-) \\ & - p^H [J(w_t^-) - J(w_t^- - \psi_t)], \end{aligned} \quad (3.15)$$

where the maximization is over the set of controls (s_t, ψ_t) that satisfies constraint (3.11). The first term arises since the purchaser is maximizing the current payoff, the second term corresponds to the drift of the IPU's continuation value, and the third term reflects the possibility of jumps in the purchaser's continuation value due to the failures. In this part we require J to be globally concave: proposition 3.2 and the proofs will formally establish this property for the purchaser's continuation value.

We can derive several properties of the optimal control using the Hamilton-Jacobi-Bellman equation (3.15). The concavity of J also implies that it is optimal to let ψ_t be as low as possible in (3.15). Consequently, including the incentive compatibility condition for the IPU (3.11) leads to the first property of the optimal contract.

Property 3.1. *The sensitivity to failures of the IPU's continuation value is given by*

$$\psi_t = \ell \quad (3.16)$$

Based on the concavity of the purchaser's continuation value, condition (3.16) reflects that it is optimal to expose the IPU to the minimal risk $\psi_t = \ell$.

Optimizing Hamilton-Jacobi-Bellman equation (3.15) with respect to ds_t yields to

$$J'(w) \geq -1, \quad (3.17)$$

where equality occurs at $ds_t > 0$. Intuitively, $J'(w)$ is the increase in the purchaser's continuation value due to a marginal increase in the IPU's continuation value, while the right-hand side of (3.17) is the marginal cost for the purchaser to make an immediate money transfer to the IPU. Thus it is optimal to delay the payments as long as the inequality (3.17) strictly holds. This reflects that the purchaser will benefit more from increases in w than transferring immediate money to the IPU. The concavity of $J(w)$ implies that condition (3.17) will hold when w_t is below a threshold. We will denote this threshold with w^p . The optimal contract satisfies the next property.

Property 3.2. *The payments to the IPU are made only if w_t exceeds the threshold w^p , which satisfies the following condition*

$$J'(w) = -1. \quad (3.18)$$

This threshold and the concavity of $J(w)$ suggest that the purchaser will provide incentives to the IPU contingent on the record of good performance. The optimal contract includes a lump sum payment of $w - w^p$, when $w > w^p$. Thus the money transfers are

$$ds_t = \max(w - w^p, 0). \quad (3.19)$$

To incentivize high quality treatment, the IPU's continuation value will be reduced by an amount that is proportional to the cost savings from shirking. To clarify this, suppose at the beginning of time t the continuation value of the IPU is

w_{t-} . If there is a failure at time t , the IPU's continuation value should be reduced from w_{t-} to $w_t = w_{t-} - \psi_t$. This results to the third property of the contract.

Property 3.3. *Whenever the continuation value w_t drops to zero, the contract will be terminated.*

Next, we will show that the above properties are consistent with the formal derivation of the optimal contract. First, we will establish there is a function J that fits the properties described above, then we will characterize the optimal contract.

Proposition 3.2. *The Hamiltonian-Jacobi-Bellman equation*

$$rJ(w) = \mu - p^H C + (\gamma w + p^H \ell + ha^H)J'(w) - p^H [J(w) - J(w - \ell)],$$

with the threshold w^p , which is endogenously determined according to

$$J'(w) = -1,$$

has a concave solution and equals the purchaser's optimal value function.

The concavity of $J(w)$ was the essential to characterizing the properties of the optimal contract. The next step is to show that the function formulated in Proposition 3.2 will lead to the maximal value for the purchaser and also implement the optimal contract. For this, we will fix the initial expected continuation value for the IPU to w_0 and consider $\{w_t\}_{t \geq 0}$ continuation value of the IPU and $\{s_t\}_{t \geq 0}$ cumulative money transfers up to and including time t , respectively, to be solutions to

$$w_t = w_0 + \int_0^{t^-} (\gamma w_u + ha^H + \ell p^H) du - \ell dN_u - ds_u, \quad (3.20)$$

$$s_t = (w_0 - w^p) \vee 0 + \int_0^{t^-} (\gamma w^p + ha^H + \ell p^H) \mathbf{1}_{\{w_u = w^p\}} du \quad (3.21)$$

for all $t \geq 0$, where w_0 is the initial expected value for the IPU and w^p is defined in Proposition 3.2.²

Proposition 3.3. *The optimal contract $\Gamma = (S, \tau)$ that motivates high quality treatments and delivers the IPU the initial expected discounted value w_0 satisfies properties (1-3). At any time t ,*

- i If $w > w^p$ a lump sum transfer of $w - w^p$ is paid to the IPU.*
- ii As long as $w_t = w^p$ and no failure occurs, money transfers to the IPU is equal to the increase in the continuation value $\gamma w^p + ha^H + \ell p^H$.*
- iii Occurrence of failure will reduce the IPU's continuation value by ℓ .*
- iv The IPU will be terminated when the continuation value is $w_t = 0$, which is stochastic and unpredictable.*

The features of the optimal contract are in line with the properties described before. To keep the social value at its maximum $J(w^p) + w^p$, lump sum money transfer will occur when $w_t > w^p$. The purchaser will transfer $\gamma w^p + ha^H + \ell p^H$ to the IPU as long as $w_t = w^p$ and no failure occurs. This term can be seen in equation (3.10) and is equal to the increase in the IPU's continuation value when $dN_t = 0$.

Using the results of the propositions, we will discuss one of the possible contract implementations. Recall that w can be considered as the amount of money the

²For each x and y , we denote by $x \vee y$ the maximum of x and y .

purchaser owes to the IPU. Starting at time 0, the purchaser and IPU agree on a contract based on the initial expected value for the IPU w_0 . If w_0 exceeds w^p the IPU will be paid the lump sum money transfer of $w_0 - w^p$. Since w^p is where the social value will be maximized, maximum of $J(w) + w$, the purchaser wants to keep the continuation value at w^p . Assuming w as the amount of credit the IPU has with the purchaser, the IPU can withdraw money from the credit line up to w . Good performance will increase w by $\gamma w + ha^H + \ell p^H$. Optimizing the care results in lower costs and better health outcomes, and thus increases IPU's continuation value. When continuation value exceeds w^p , lump sum money transfer will occur. As a result, the characterized optimal contract in this research incentivizes the IPU to invest in cost reducing efforts while providing high quality care to the patient. Furthermore, the possibility of punishment by decreasing the continuation value is another source of incentive for the IPU to provide high quality treatments to the patient. Any adverse health outcomes will decrease w by ℓ and also the contract will be terminated when the IPU maxes out the credit limit, when $w = 0$.

3.5 Conclusions

We studied the contracts that coordinate the healthcare purchaser-IPU relationship in the context of value-based healthcare delivery. Coordinating contracts allow the IPU to optimize its objective while maximizing the societal welfare. We considered the hidden action problem, in which the IPU's treatment strategy is unverifiable to the purchaser. To capture the dynamics of health outcomes over the care cycle we used a continuous-time dynamic principal-agent model to derive the optimal contract. This technique allows us to link the payments to the history of health outcomes.

Failure in healthcare delivery affects patients' health outcomes. Failures are expensive, not only because patients have to go through costly treatments but also they lose functional days. Accountability for avoidable complications is one of the ways value-based healthcare delivery is intending to reduce inefficiencies in the healthcare system. Our analysis suggests that, in order to prevent healthcare failures, compensation policies should be made contingent on the accumulated performance. We used the concept of *continuation value* previously used in the continuous-time principal-agent problem in several contexts. Continuation value can be interpreted as the IPU's track record. Accumulated good performance will result in positive money transfers, and failures will reduce the compensation.

Our research contributes to the literature of healthcare contracting in several ways. First, it closes the gap in designing a dynamic incentive contract in the healthcare delivery context. Current payment systems like fee-for-service reimburse the providers for discrete services and can encourage providers to overprovide services. However, to add value for the patients the providers should evaluate if the additional test or treatment can create better health outcomes or not. Alternative payment system like capitation can control the increasing healthcare expenditures, but might incentivize the providers to underprovide services or pick the healthy patients that will require less care. On the other hand, current healthcare delivery structure does not let the providers to leverage the wisdom other providers might have acquired on the patient's condition. The value-based healthcare delivery brings all the providers that determine the well-being of a patient with certain medical condition together in an IPU. Since the IPU's treatment strategy stochastically affects the health outcomes over time, we need a model that considers the health outcomes over the care cycle and evaluates the IPU based on that. We determined a payment arrangement that encourages the IPU to provide high quality

treatments to the patients with a given budget. Good performance record for the IPU will be reimbursed by a bonus payment. This is a significant contribution because most of the existing payment systems that link the payment to the performance, pay for fulfilling some targeted processes. Paying for processes might not result in better health outcomes. Additionally, pay-for-Performance systems that pay for outcomes are challenging to implement in the current structure. When healthcare providers work individually, they cannot be held accountable for the health outcomes. Each provider can argue the other providers involved in the well-being of a patient have not provided good quality of care.

Second, our research mathematically demonstrates what is the optimal payment system in the value-based delivery context. Michael porter argues that a bundled payment should coordinate the relationship between the healthcare purchaser-IPU relationship [20]. But, we find that other than the bundled payment, IPU should be compensated with a bonus when they achieve superior performance. Basically, the payment to the IPU should be adjusted based on the health outcomes.

Third, our way of characterizing the optimal contract can arguably result in straightforward implementation. The use of continuous-time principal-agent model helped us to model the problem with minimal assumptions and devised us with great technical tool to summarize IPU's track over the care cycle. Cost-reducing efforts and value-adding treatments increase the IPU's *continuation value*, which can eventually result in the bonus payments if they exceed a certain threshold for the continuation value. Thus, our proposed payment scheme can fulfill the tenets of value-based healthcare delivery by acting as the source of incentive for the IPU to improve the health outcomes and minimize the costs at the same time.

Our model can be extended to include the possibility of learning throughout time for the IPU. Learning could provide cost-reducing opportunities and thus efficiently providing care to the patients. The implications of learning in value-based healthcare delivery on the payment systems is interesting to study. Further, the model could be modified for specific diseases. Particularly, terminally ill patients have distinct health evolution.

Alternatively, optimal design of non-financial incentives among the IPUs could be studied. If different IPUs could be rated in comparison to their peers, they might behave differently. As a result, the healthcare purchaser could benefit from building in social comparisons or peer pressure into their mechanisms.

3.A Appendix: Chapter 3 Proofs

Proof of Lemma 3.1. Condition (3.9) is the predictable representation of martingale $v_t(\Gamma, A)$, following from Brémaud (Chapter III, Theorems T9 and T17) [54]. □

Proof of Proposition 3.1. Let v'_t represent the IPU's lifetime expected value, given the information available at time t , when he uses an alternative treatment strategy $A' = \{a'_t = a^L\}_{t \geq 0}$ until time t and then changing the treatments to $A = \{a_t = a^H\}_{t \geq 0}$, as a result:

$$v'_t(\Gamma, A') = \int_0^t e^{-\gamma u} (ds_u - h a'_u du) + e^{-\gamma t} w_t(\Gamma, A), \quad (3.22)$$

Similar to Sannikov [47] and Biais et. al [51], it is straightforward to show that if $\psi \geq \ell$, where $\ell = h \frac{a^H - a^L}{p^L - p^H}$, the drift of the process is going to be nonpositive for all

t and thus v_t^i is supermartingale for any alternative strategy A' . As a result, the strategy A is at least as good as the alternative strategy A' . \square

Proof of Proposition 3.2. First, we will establish the existence of a twice differentiable solution to the Hamiltonian-Jacobi-Bellman (HJB) equation (3.15). Assumption (3.1), implies that the purchaser wants to avoid health failures at the first best solution.

$$rJ(w) = \mu - p^H C + (\gamma w + ha^H)J'(w) + p^H[J(w) - J(w - \ell)] \quad (3.23)$$

Let's define function H as

$$H(w, u, s) = -\min (ru - \mu + p^H C - s(\gamma w + ha^H) - p^H[u(w) - u(w - \ell)])$$

therefore, the HJB is equivalent to

$$H(w, u, s) = 0 \quad (3.24)$$

By Berge's Maximum Theorem, $H(w, u, s)$ is jointly continuous in its parameters. This implies that for any slope m , the initial value problem with boundary conditions $J(0) = 0$ and $J'(w) = m$ has a continuous differentiable solution on its domain. Furthermore, since the HJB equation is the sufficient condition for optimality any candidate function that solves the HJB, is indeed optimal.

Denote by $J(W)$ the purchaser's highest value from a contract that provides the IPU the continuation value of W . Since the purchaser has the option of providing W to the IPU by paying a lump-sum transfer of $ds > 0$ and moving to the optimal

contract of $W - ds$, the following condition must hold

$$J(W) \geq J(W - ds) - ds \quad (3.25)$$

which implies that $J'(w) \geq -1$ for all w . Condition (3.25) along with $rJ(w) < \mu - p^H C - (\gamma w + p^H \ell + ha^H)$ imply that $J'' < 0$. Therefore, to the left of w^p with boundary condition $J'(w^p) = -1$ and $rJ(w^p) = \mu - p^H C - (\gamma w^p + p^H \ell + ha^H)$, function J is concave. \square

Proof of Proposition 3.3. The proposed contract generates continuation value that satisfies (3.10), with $A = a^H$ and $\psi_t = \ell$. Therefore, by Proposition 3.1 induces high quality treatments. From Proposition 3.2 and the boundary condition for $J(w)$, the threshold that purchaser's value is maximized is w^p . When $w > w^p$ because of the concavity of the $J(w)$ the purchaser is better off to pay a lump sum money of $w_t - w^p$ reduce the amount of money they owe to the IPU to w^p . From equations (3.20) and (3.21), it can be seen that when IPU is performing well the continuation value increases, which might result in lump sum money transfers to the IPU. As far as there is no failures the IPU's continuation value will not be altered and payment is equal to the increase in the continuation value, $s_t = \gamma w^p + ha^H + \ell p^H$, w will stay constant at w^p since $dw_t = 0$. \square

Chapter 4

Healthcare Providers in an Integrated Practice Unit: Modeling Partnerships

4.1 Introduction

A focused and disciplined team has been defined as the organizational unit that can tackle both performance targets and complex changing environment challenges [56]. Teams exist in almost all business and professional environments, including car manufacturing, professional sports, and military. The value of teams is even more evident when the performance requires multiple skills, judgments, and experiences [56]. Nonetheless, not all teams reach their goals. Successful teams have a common purpose, an agreement on performance goals, and a common working approach. In these teams, members develop complementary skills and hold themselves accountable for the achievements [57].

Interdisciplinary teams have been recognized as the success factor for delivering health services [14]. Improved teamwork and collaborative care have been shown to improve healthcare delivery in many aspects, including quality of care and patient safety. It has also been shown that the teamwork can significantly increase job satisfaction and reduce patient morbidity [58, 59]. Despite the promising tenets of team-based healthcare delivery, building and maintaining effective teamwork is challenging. The challenges include lack of common definition of team and teamwork, organizational factors affecting teamwork, and current policies and regulations. The successful implementation of team-based delivery in healthcare depends on the existence of a common purpose, good communication, coordination, performance goals, working approach, complementary skills, and responsibility.

The primary care reform and the integrated client care project (ICCP) are two recent collaborative programs that have been introduced in Canada in order to shift to teams of providers who are accountable for providing comprehensive services to their clients. Other countries have introduced similar programs. Likewise the value-based healthcare delivery, which is the basis of the aforementioned reforms in Canada, requires a collaborative relationship between providers involved in the Integrated Practice Unit (IPU).

In the value-based healthcare delivery, each provider has comparative specialties in understanding different aspects of the patient's problems and technical capabilities in providing the service. All engaged providers assist the others in understanding the disease, planning and executing the care. They also may share infrastructure for providing the integrated care. The health outcomes of the patients then depend on the attentive and persistence efforts of all healthcare providers involved. Each healthcare provider can observe the outcome of the care process, but cannot verify the actions taken by the other providers.

When multiple agents are involved in taking hidden actions, they may not pay thorough attention to the unobservable actions because the costs and benefits would be shared among all agents [37]. In the healthcare delivery context, the stochasticity of health outcomes aggravates the hidden action problem. Each provider knows he can declare the unsatisfactory outcome was simply another provider's fault, bad luck with disease progression, or patient's compliance. Value-based healthcare delivery requires the providers to work in teams while each provider is performing part of the interrelated treatment, hence, the health outcomes of the patient depend on all providers' quality of care and cooperation.

Ontario's current proposed plan for the implementation of ICCP includes one single contract for multiple services based on patients' need, i.e. bundled payment approach, where multiple providers are reimbursed by a single sum of money for the related services to a defined episode of care. As a result, instead of paying for discrete services, as in fee-for-service payment, the bundled payment encourages the providers to improve the quality and eliminate clinically ineffective or redundant services [60]. Even though it has been argued that a bundled payment may improve quality and reduce the costs simultaneously, its implementation faces challenges [60]. The primary concerns with the bundled payment are that it will be impossible to manage when independent providers are expected to collaborate and share the payment [60, 61]. One of the major problems is the determination and the distribution of payment among the providers in an IPU [61]. A contractual and collaborative arrangement among the providers and facilities is needed to determine how the payments should be distributed among the providers in an IPU [61]. The implementation of these arrangements would require significant structural changes within the participating entities. For example, hospitals would need to reimburse physicians for their services. As a result, the responsible entity

would have to set up contracting, billing, and reimbursement systems with other providers in the team, which may translate to an expansion of current technological capabilities as well [61].

We assume the healthcare purchaser offers a contract to the IPU, based on their overall performance over the care cycle of the disease. Therefore, here we will focus on the payment system among the healthcare providers within an IPU which will encourage providers to collaborate and coordinate on the care.

To model this problem, we assume that one of the involved healthcare providers, is the entity who is accountable for health outcomes to the healthcare purchaser and receives the payments from her, whom we will call healthcare leader. The healthcare leader is committing to a contract with the health purchaser and is facing the risk of stochastic health outcomes, and hence needs to design a contract which leads to optimal treatment decisions by the other healthcare providers. The healthcare leader will play the principal role, but is also involved in providing the treatment to patients. When the principal is also involved in performing tasks that affect the outcomes, the problem of hidden action exacerbates to the problem of double hidden action. In our problem, we will assume all parties are risk-neutral and willing to consider providing different treatments that might result in better health outcomes for the patients.

The value of the care depends heavily on how different related healthcare providers interact and provide the service, therefore cooperation and coordination is a vital factor and needs to be enhanced. Care provision to the patient is subject to patient risk factors, type of disease, and the extent to which the care pathway is defined. Because both healthcare leader and healthcare provider are involved in the treatment decisions, they are facing the problem of multi-sided

hidden action.¹ Multisided hidden action arises when several agents take complementary hidden actions [25, 37]. In an IPU, health provider's and leader's actions stochastically influence the health outcomes in the current period and the state of the system in the next period. We model this collaborative, iterative, and stochastic work as a finite-horizon stochastic game [62, 63], in which the state of the patient evolves according to a Markov Chain and the transition probabilities depend on the quality of treatments.

The work dynamics among the team members are affected by the choice of contract. Since the collective efforts of the providers will affect the patient's health outcome, it is impossible to infer the respective intensity or quality of care from the health outcomes. The problem in the healthcare context is more cumbersome since, health outcomes stochastically depends on the care provided and patient's risk factors and compliance to the care process. When efforts are unverifiable, finding the optimal contract is challenging [37]. It is important to determine which contracts should be adopted and how they affect the collaboration dynamics among the providers.

We will consider two cases. First, we will assume the health outcome is contractible. We study the work dynamics under reward-sharing contracts. When the outcomes of treatments are either affected by other factors or emerge late in the process, they are not contractible. Second, we will investigate the work dynamics when the health outcomes are noncontractible. In contract theory, informal agreements have proved to be beneficial when the outcome of a partnership is stochastic and non-contractible, as it might be the case in the healthcare providers' relationships in an IPU. We will derive the work dynamics under an informal agreement which makes payments contingent on non-contractible health outcomes.

¹In the literature, the term multisided moral hazard has been used to refer to the same problem. However, in this chapter, we will use multisided hidden action.

4.2 Literature Review

Classic agency theory has provided tools to examine the principal-agent problem. The principal hires an agent to complete the tasks that are too complicated or too costly to perform. The main concern is how the principal can motivate the agent to perform the tasks as the principal would prefer [23]. Because of the noise in the (production) process, the output is uncertain, and compensations based on the output exposes the agent to an undesirable risk.

In the standard principal-agent model, the principal moves first and selects the payment system to the agent which depends on the output. Next, the agent selects an effort level, which the principal cannot observe. The agent's effort level will determine the output stochastically. This is a problem of single-sided hidden action, where the principal's action can be identified with a monetary transfer rule. When two parties are in a partnership, as in our case, each of them exerts an unobservable effort for an interrelated process. The output level is stochastic and depends on the effort level of all players, resulting in a situation of double hidden action [37]. When the objective is to optimize the global system, the problem of hidden action can be alleviated through coordinating contracts. Contract theory has been applied to business-to-business settings, including cross-functional coordination [64] and supply chain coordination [65–68]. Nevertheless, most of these contracts are developed specifically for manufacturer-supplier or supplier-retailer paradigms to maximize profit through sales, assuming the output quality will not be altered. In contrast to supply contracts, which are based on quantities, the model for healthcare providers is focused on the quality of the service and its outcomes. Furthermore, the health outcomes over the care cycle is the true measure for the IPU's success. Considering that the healthcare provider

and leader will be engaged in the joint service, there needs to be substantial modeling changes in comparison to the traditional principal-agent framework.

Linking the payoffs in one period to the outcomes in previous periods can reduce the severity of hidden action problem and therefore improve risk sharing. Malcomson et al. [69] shows that efficient contracts under hidden action do not require long-term commitments. The possibility of severe punishments makes the short-term contracts as efficient as long-term contracts for both agent and principal. Fudenberg et al. [46] also studies long-term principal-agent relationships where the agent's consumption is not necessarily based on his compensation, and also his compensation is according to a verifiable outcome. They have shown that short-term contracts are as good as long-term contracts. The main assumption in their study is that the principal and the agent have the same information about available technologies, the agent's preferences over future outcomes, and the agent has an unlimited access to saving and borrowing at the same interest rate as the principal. Since, in general, healthcare providers have access to the capital provided by a third party, the banking assumption is defensible in the healthcare delivery context and has been used by Fuloria et al. to derive an outcome-adjusted payment system as well [27].

Bonatti et al. examines the hidden action (moral hazard) problem in teams over time where agents are collectively involved in a project whose duration and outcome are uncertain, and their individual efforts are unverifiable [70]. Bonatti et al. consider projects with two states, characterize the work dynamics with respect to time, and find that collaboration diminishes over time, but the project will continue until it succeeds. Rahmani et al. [71] characterizes the work dynamics for knowledge-intensive projects for which the rewards increases as the project proceeds. Similar to Rahmani et al., we will characterize the work dynamics with

respect to time and state. However, our research concentrates on two payment systems that (arguably) reduce the problem of hidden action, specifically reward-sharing and relational contracts.

When the agent's total contribution is not contractible, there are two alternatives: court-enforceable contracts based on alternative performance measures and relational contracts [72]. Relational contracts are the agreements which are based on outcomes that are observed by contracting parties *ex post*. Relational contracts allow the parties to use their knowledge of specific situation and also utilize new information as it becomes available. Relational contracts are self-enforceable if each party's reputation is sufficiently valuable that neither party wishes to break a promise, for a seminal study see [73].

The use of the combination of both formal and informal contracts is customary. Many companies use both piece rates and subjective bonuses [73, 74]. The simultaneous use of formal contracts based on objective performance measures and relational contracts based on discretionary bonuses based on total contribution has been studied by Baker et al. [74]. The use of a relational contract can reduce the distortionary incentives that a formal contract could create. The use of a formal contract can also reduce the size of the bonus to be offered by the principal if she would use the relational contract only. Thus a combination of formal and informal contracts might be beneficial for both parties. For overview of contract theory see [72].

Previous studies have concentrated on the static models for designing the optimal contracts for the collaborative relationships [75, 76]. The exception is the study by Plambeck et al. [77], in which they formalize the optimal contract using a dynamic principal-agent framework. Plambeck et al. consider two profit maximizing risk-neutral firms that engage in joint production in which outputs of the

system, payments, and future of the relationship are determined by the actions of both manufacturer and buyer. They demonstrate the existence of a simple relational contract when buyer and manufacturer are interacting over multiple periods.

In a static setting, Bhattacharyya et al. [78] show the second-best outcome can be obtained by a reward-sharing contract and Roels et al. [76] analyze the pricing and assignment of efforts in the presence of complementary effort levels. Roels et al. characterize the optimal contract and its dependence to the service environment characteristics, such as output elasticity, effort verification costs, output uncertainty, output measurability, and process improvement opportunities. Their model finds the optimal contract for two risk-neutral agents working together on a project with identified output [76]. Roels and colleagues find that simple contracts like fixed-fee and time-and-materials perform poorly in the context of static double hidden action.

The remainder of this chapter is organized as follows. We will detail the partnership model in §4.3. We derive the dynamics of collaboration when efforts are verifiable in §4.4, we call this the first-best solution. We study the effects of two payment schemes that are established to alleviate the double hidden action problem, namely reward-sharing contracts and relational contracts in §4.5.1 and 4.5.2, respectively. We present our conclusions in §4.6. All proofs appear in Appendix §4.A.

4.3 Partnership Modeling

In this section we present a model for dynamic healthcare delivery. We assume two players are engaged in treating the patient over multiple periods, healthcare provider and healthcare leader $i = \{p, \ell\}$, respectively. The healthcare leader has the role of *care manager*, who also provides care to the patients [5]. We will model the healthcare delivery by the IPU as a dynamic game over the finite disease cycle T . As in chapter 3, we define the patient health status as being “well” or “unwell,” and the success for the IPU as keeping the patient in the “well” status or bringing them back to the “well” status. Accordingly, the IPU fails when the patient remains in the “unwell” status or transitions back to the “unwell” health status. Figure 4.1 summarizes the health dynamics for a patient.

At each stage, player $i = \{p, \ell\}$ chooses his effort a_i from a finite action space A_i . In this chapter, we assume each provider can either “work” collaboratively or “not work” collaboratively, $a_i \in \{W, N\}$. As a result of this binary action,

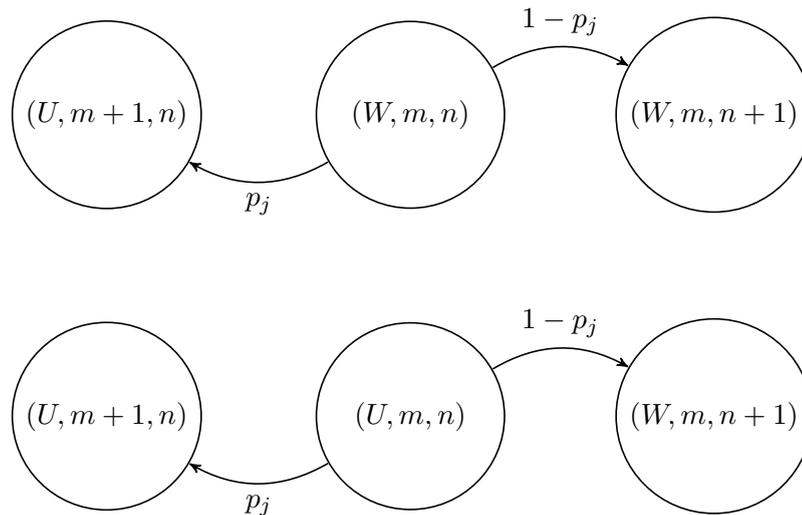


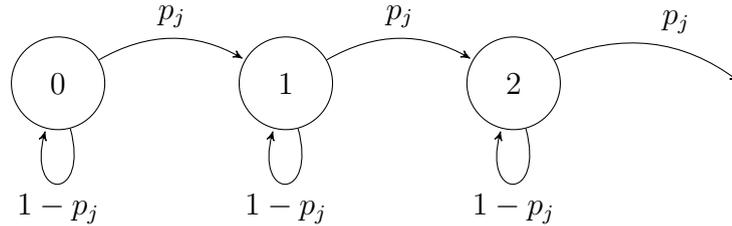
FIGURE 4.1: Problem state: patient status “well” or “unwell” (W or U), number of failures (m), and number of successes (n); probability of failure p_j

TABLE 4.1: Providers' efforts and working relationship outcomes

Provider \ Leader	Work W	Not Work N
Work W	High	Low^p
Not Work N	Low^ℓ	Stop

four working modes are possible *High*, Low^p , Low^ℓ , *Stop*. The working modes represent the outcome of the actions taken by the providers. When both providers work collaboratively, the outcome is “High” and will represent high quality treatment. Table 4.1 summarizes these working modes. Each working mode induces a probability distribution over the publicly observable outcome x .

As a result of the problem state dynamics explained in figure 4.1, the number of IPU failures can summarize the problem state. For each period $t \in T$, the state of the problem at the beginning of period t is denoted by x_t . When the IPU succeeds in healthcare delivery, the number of failures will not change. Therefore, the problem state at the beginning of period $t+1$ will be $x_{t+1} = x_t$. When the IPU fails, the number of failures increases in the next period by one unit $x_{t+1} = x_t + 1$. As mentioned before, health outcomes are stochastic and depend on the working mode. The probability of failure for each working mode is denoted by p_j , where $j = \{H, L^p, L^\ell\}$ representing *High*, Low^ℓ , Low^p . The state x_t evolves to state $x_t + 1$ with probability p_j or remains at state x_t with probability $1 - p_j$ as depicted in figure 4.2.

FIGURE 4.2: State dynamics in working mode $j = \{H, L^p, L^\ell\}$

We assume the probability of failure for modes Low^p and Low^ℓ is equal $p_{L^\ell} = p_{L^p}$ and will be noted by p_L . The probability of failure when both providers work collaboratively ($High$) is lower than individual efforts (Low^i), i.e. $p_H < p_L$. If both providers choose not to work in collaboration ($Stop$), they will terminate the treatments on the patient and the system will not evolve afterwards.

There are monetary and non-monetary costs associated with each working mode. Each player will incur costs according to their effort level and the resulting working mode. We denote the cost of working mode $High$ by $c_H > 0$ and the cost of working mode $low_{L^i}^i$ or $low_{L^{-i}}^i$ for player i , by $c_{L^i}^i$ and $c_{L^{-i}}^i$, respectively. We assume that working in collaboration with the other player requires more time and effort, the cost of working in $High$ mode for any player i , c_H^i is greater than working in $Low_{L^i}^i$ or $low_{L^{-i}}^i$, i.e. $c_H^i > \max\{c_{L^p}^\ell, c_{L^\ell}^\ell\}$ and $c_H^p > \max\{c_{L^p}^p, c_{L^\ell}^p\}$. In addition, we assume the player that is not collaborating will incur a lower cost than the one who is working individually $c_{L^\ell}^p < c_{L^\ell}^\ell$ and $c_{L^p}^\ell < c_{L^p}^p$. Since the healthcare leader is both providing service and overseeing the care process for the patient, $c_{L^\ell}^\ell \geq c_{L^p}^p$. We will assume that when the healthcare provider and leader decide to $Stop$ and terminate the treatments earlier than the end of the disease cycle $t < T$, they will incur the penalty of k^p and k^ℓ , respectively.

We assume perfect information about the patient and other player's costs. When efforts are contractible, the providers write a contract to maximize the total surplus. We refer to this situation as first-best solution. When efforts are not contractible, we will consider two situations. First, we will study the case where health outcomes are contractible and study the effect of payment schemes like performance-based and capitation. Next, we will consider the case where health outcomes are noncontractible and we will investigate the effects of using a relational contract among the providers.

4.4 The First-Best Solution

If the provider's and leader's efforts were verifiable, the first-best solution could be attained by writing a contract based on the rewards and costs. We will solve a finite-horizon dynamic programming problem to derive the optimal policy. We denote the first-best optimal policy in state x and time t by $\varepsilon_t^{fb}(x)$. Recall, x is the cumulative number of failures that the IPU has faced up until period t .

Denote by $V_t(x)$ the total discounted surplus in period t and state x , defined recursively as follows:

$$\begin{aligned} V_t(x) &= \max\{V_t(x|High), V_t(x|Low^p), V_t(x|Low^\ell), V_t(x|Stop)\} \quad t = 1, 2, \dots, T-1 \\ V_T(x) &= R(x) \end{aligned} \tag{4.1}$$

The value function at the end of disease cycle T depends on the number of failures during the care cycle. As the number of failures x increases, the reward $R(x)$ decreases. We assume the healthcare purchaser encourages providing quality care from the outset and preventing avoidable complications instead of curing the complications later in the disease cycle. As a result, the reward function $R(x)$ is decreasing convex.

With the assumption $c_{L^\ell}^\ell \geq c_{L^p}^p$, Low^p weekly dominates Low^ℓ . As a result, Low^ℓ is never optimal. In the first-best solution, we refer to Low^p as Low . Hence only three working modes are attainable in period t and state x , $\varepsilon_t^{fb}(x) = \{High, Low, Stop\}$. Time is discounted at a rate $\delta \leq 1$.

$$\begin{aligned} V_t(x|High) &= -c_H^p - c_H^\ell + \delta [p_H V_{t+1}(x+1) + (1-p_H)V_{t+1}(x)] \\ V_t(x|Low) &= -c_L^p - c_L^\ell + \delta [p_L V_{t+1}(x+1) + (1-p_L)V_{t+1}(x)] \\ V_t(x|Stop) &= -K \end{aligned} \tag{4.2}$$

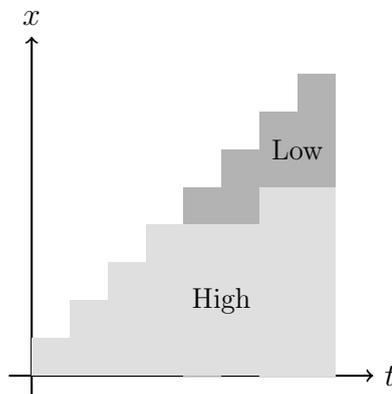


FIGURE 4.3: Dynamics of collaboration among the providers in an IPU; the first-best Solution

Throughout this chapter, we denote $\Delta V_t(x) = V_t(x+1) - V_t(x)$ and $\Delta R(x) = R(x+1) - R(x)$. Recall, reward function is decreasing with x thus, $\Delta R(x) \leq 0$.

We characterize the dynamics of the first-best solution in proposition 4.1. We show that there is a time-independent number of failures threshold above which it is optimal to work in *Low* mode. Comparing $V_t(x|High)$ and $V_t(x|Low)$ yields to the following optimal policy for any period t :

$$\begin{aligned} \varepsilon_t^{fb}(x) &= High \text{ if } \Delta V_{t+1} \leq \frac{(c_H^p + c_H^l) - (c_{LP}^p + c_{LP}^l)}{\delta(p_H - p_L)} \\ \varepsilon_t^{fb}(x) &= Low \text{ if } \Delta V_{t+1} \geq \frac{(c_H^p + c_H^l) - (c_{LP}^p + c_{LP}^l)}{\delta(p_H - p_L)} \end{aligned} \quad (4.3)$$

The dynamics are illustrated in figure 4.3. The horizontal axis denotes the time from the start of treatment $t = 0$ to the end of care cycle $t = T$ and the vertical axis shows the number of failures x_t . Because the number of failures increases at most by one unit in each period, the set of feasible states are below the 45-degree line, $x_t \leq t$. The proof consists of decomposing the space (t, x_t) into regions in which the value function has decreasing differences in (t, x_t) and then applying a standard induction argument.

We denote, for any t , the highest state where $\varepsilon_t^{fb}(x) = High$, as

$$x_{H,t}^{fb} = \left\{ \max\{x \in \mathbb{Z}^+ \mid V_{t+1}(x+1) - V_{t+1}(x) \leq \theta_{LH}^{fb}\} \right\} \quad (4.4)$$

where $\theta_{LH}^{fb} = \frac{(c_H^p + c_H^e) - (c_{LP}^p + c_{LP}^e)}{\delta(p_H - p_L)}$. For any $x < x_H^{fb}$, we will define $\tau(x) := \max\{t \in T \mid \Delta V_{t+1}(x) \leq \theta_{LH}^{fb}\}$, if there exists such a t that $\Delta V_{t+1}(x) \leq \theta_{LH}^{fb}$; otherwise, $\tau(x) = x - 1$.

Proposition 4.1. *There exists a threshold for number of failures x_H^{fb}*

- i. $\varepsilon_t^{fb}(x) = Low$ for all t if $x \geq x_H^{fb}$*
- ii. for $1 \leq x \leq x_H^{fb}$, there exists a time threshold $\tau(x)$, nondecreasing in x , such that $\varepsilon_t^{fb}(x) = High$ for all $\tau(x) < t < T$ and $\varepsilon_t^{fb}(x) = Low$ for all $x \leq t \leq \tau(x)$*

Proposition 4.1 identifies the phases in which the IPU members should work collaboratively. Working in *High* mode is optimal early in the disease cycle, and *Low* is optimal as failures increases. When failures increase the providers tend to work in *Low* mode longer.

4.5 Double-Sided Hidden Action

In the standard hidden action problem, the common assumption is that the principal is passive in the relationship and is not involved in the job being undertaken by the agent. The principal delegates all the decisions to the agent and designs an incentive contract based on observable outputs that are correlated with agent's hidden action. However, in many situations, the principal also make choices that

affects the outcomes. This is known as double-sided hidden action. In any IPU, the healthcare providers work in teams as well and the collective efforts of all providers over the care cycle determine the health outcomes of the patient. Recall, we consider the *care manager* as the principal in the relationship among the IPU members.

We will consider two scenarios. First, we will study the collaboration dynamics when the reward is contractible. We use the combination of reward-sharing contract with capitation. Next, we consider the case where the reward is not contractible. We use the combination of formal and informal contract to determine the IPU's collaboration dynamics.

4.5.1 Contractible Reward

Assuming the reward $R(x)$ is contractible, in this section we consider a reward-sharing contract among the providers. In static games, it has been shown that an optimal second-best contract can be implemented through a linear revenue- or profit-sharing contract [78]. We adopt a linear reward-sharing contract to study the collaboration dynamics among IPU members. Each player i will receive a share of α^i from the reward $R(x)$ that is, $s_p(x) = \alpha^p R(x)$ and $s_\ell(x) = R(x) - s_p(x) = \alpha^\ell R(x)$ with $\alpha^p + \alpha^\ell = 1$. In addition to the payment at the end of disease cycle $s_p(x)$, the healthcare provider will be paid an upfront fixed amount for each patient. In return to this fixed amount, the provider will provide all the care needed during the care cycle T . We will denote the upfront fixed payment at time $t = 0$ by F .

We model the collaboration process as a finite-horizon dynamic stochastic game [62]. We will denote the discounted equilibrium payoff-to-go of player i , $i \in \{p, \ell\}$

in state x in period t by $V_t^i(x)$. The healthcare provider's payoff given possible collaboration outcomes is as follows:

$$\begin{aligned} V_t^p(x|j) &= -c_j^p + \delta \mathbb{E}_j [V_{t+1}^p(x + \xi)] \text{ for } t = 1, 2, \dots, T - 1 \\ V_t^p(x|Stop) &= -k^p, \text{ for } t = 1, 2, \dots, T - 1 \\ V_T^p(x) &= \alpha^p R(x) \end{aligned} \quad (4.5)$$

where $j = \{High, Low^p, Low^\ell\}$ and $\mathbb{E}_j [V_{t+1}^p(x + \xi)] = p_j V_{t+1}^p(x + 1) + (1 - p_j) V_{t+1}^p(x)$. The healthcare leader's payoff $V_t^\ell(x)$ in state x and time period t is

$$\begin{aligned} V_t^\ell(x|j) &= -c_j^\ell + \delta \mathbb{E}_j [V_{t+1}^\ell(x + \xi)], \text{ for } t = 1, 2, \dots, T - 1 \\ V_t^\ell(x|Stop) &= -k^\ell, \text{ for } t = 1, 2, \dots, T - 1 \\ V_T^\ell(x) &= \alpha^\ell R(x) \end{aligned} \quad (4.6)$$

in which, $\mathbb{E}_j [V_{t+1}^\ell(x + \xi)] = p_j V_{t+1}^\ell(x + 1) + (1 - p_j) V_{t+1}^\ell(x)$.

$$\delta \mathbb{E}_j [V_{t+1}^p(x + \xi)] \geq 0 \quad (4.7)$$

$$\delta \mathbb{E}_j [V_{t+1}^\ell(x + \xi)] \geq 0 \quad (4.8)$$

constraints (4.7) and (4.8) ensure that both players prefer to continue the collaboration rather than stopping the treatments.

For this game, we will consider a feedback information structure [79], which means that during each period the players know exactly to which state the game has evolved and that information is fed back to their strategies and actions. We will focus on pure-strategy Markov-perfect equilibria [80]. Markov equilibria is the simplest form of behavior consistent with rationality and can capture the notion that “bygones are bygones” [80]. Similar to Marx and Matthews [81] and Rahmani [71], we will characterize the collaboration dynamics when a pure-strategy

equilibrium exists.

To be able to compare the working modes, we will define some thresholds which will allow us to compare the working modes to each other. Since the cost of working in Low^{-i} mode is lower than working in Low^i mode for player i , and the probability of failures are the same $p_{L^i} = p_{L^{-i}} = p_L$, working mode Low^{-i} dominates Low^i for player i . Therefore, we only need to compare $High$ and Low^{-i} for each player.

$$V_t^i(x|High) = -c_H^i + \delta \mathbb{E}_H[V_{t+1}^i(x + \xi)]$$

$$V_t^i(x|Low^{-i}) = -c_{L^{-i}}^i + \delta \mathbb{E}_L[V_{t+1}^i(x + \xi)]$$

where $\mathbb{E}_j[V_{t+1}^i(x + \xi)] = p_j V_{t+1}^i(x + 1) + (1 - p_j) V_{t+1}^i(x)$. We will define $\varepsilon_t^{RS}(x)$ as the selected pure-strategy equilibrium played in period t and state x under reward-sharing contract.

We will start from period $T - 1$ and define the following threshold for player $i \in \{p, \ell\}$:

$$\theta_{L^i H}^{RS} = \frac{c_H^i - c_{L^{-i}}^i}{\delta \alpha^i (p_H - p_L)} \quad (4.9)$$

In period $T - 1$, player i prefers Low^{-i} over $High$ if and only if $\Delta R(x) \geq \theta_{L^i H}^{RS}$. We define x_L^{RS} and x_H^{RS} as follows:

$$x_L^{RS} = \max\{x \in \mathbb{Z}^+ | \Delta R(x) \geq \max\{\theta_{L^p H}^{RS}, \theta_{L^\ell H}^{RS}\}\}. \quad (4.10)$$

$$x_H^{RS} = \max\{x \in \mathbb{Z}^+ | \Delta R(x) \leq \min\{\theta_{L^p H}^{RS}, \theta_{L^\ell H}^{RS}\}\} \quad (4.11)$$

By comparing working modes *Stop* and *High* and working modes *Stop* and *Lowⁱ* the following thresholds can be derived for $i \in \{p, \ell\}$:

$$\theta_{SL^i}^{RS} = \frac{c_{L^i}^i - k^i}{\delta \alpha^i p_L} - \frac{R(x)}{\delta p_L}$$

$$\theta_{SH}^{RS} = \frac{c_H^i - k^i}{\delta \alpha^i p_H} - \frac{R(x)}{\delta p_H}$$

At each time period t , the highest state for which collaboration is optimal is defined as $\bar{x}_{H,t}^{RS}$, in which

$$\bar{x}_{H,t}^{RS} \equiv \max\{x \leq \bar{x}_{H,t+1}^{RS} | \varepsilon_t^{RS}(x) = High\}.$$

For any $x < x_H^{RS}$, we will define $\tau_H^{RS} := \min\{t \in T | \varepsilon_t^{RS}(x) = High\}$, if there exists such a t that for both players $\Delta V_{t+1}^i(x) \leq \alpha^i \theta_{L^i H}^{RS}$; otherwise, $\tau_H^{RS}(x) = T$. For any $x < x_H^{RS}$, we will define $\tau_L^{RS}(x) := \max\{t \in T | \varepsilon_t^{RS}(x) = Low^\ell\}$, i.e. $\Delta V_{t+1}^p(x) \geq \alpha^p \theta_{L^p H}^{RS}$ if there exists such a t ; otherwise, $\tau_L^{RS}(x) = x - 1$.

For the *Stop* to dominate *High* or *Lowⁱ* in period $T - 1$ the following conditions should hold, respectively

$$\Delta R(x) \leq \theta_{SL^i}^{RS}$$

$$\Delta R(x) \leq \theta_{SH}^{RS}$$

Since $R(x + 1) - R(x) > -R(x)$ or $\Delta R(x) > -R(x)$, $\Delta R(x)$ will be greater than a smaller value as well, i.e. $\Delta R(x) > \frac{-R(x)}{\delta p_j}$. Assuming $k^i > c_j^i$, then $\Delta R(x)$ will never be smaller than $\theta_{SL^i}^{RS}$ and θ_{SH}^{RS} and therefore the cost stopping the treatments is much greater than the cost of the treatments. As a result, working mode *Stop* is dominated by *High* and *Lowⁱ* in all the periods.

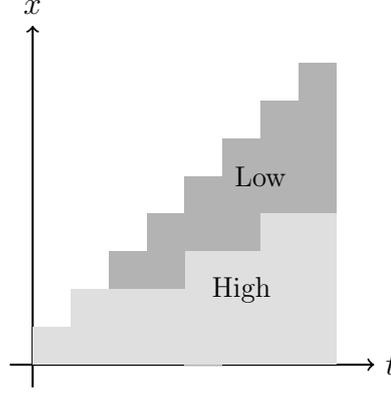


FIGURE 4.4: Dynamics of collaboration among providers in an IPU in a reward-sharing contract

We find that the collaboration dynamics under the reward-sharing contract are similar to the first-best solution. However, the players tend to start working individually at lower number of failures in comparison to the first-best solution. Proposition 4.2 characterizes the reward-sharing collaboration dynamics and figure 4.4 demonstrates that result.

Proposition 4.2. *There exists thresholds for the number of IPU failures x_H^{RS} and x_L^{RS} such that*

- i. for all $x \geq x_L^{RS}$, $\varepsilon_t^{RS}(x) = Low^\ell$.
- ii. for all $x_H^{RS} < x < x_L^{RS}$,
 - a. if $\alpha^p(c_H^\ell - c_{Lp}^\ell) < \alpha^\ell(c_H^p - c_{L\ell}^p)$, $\varepsilon_{T-1}^{RS}(x) = Low^\ell$.
 - b. if $\alpha^p(c_H^\ell - c_{Lp}^\ell) \geq \alpha^\ell(c_H^p - c_{L\ell}^p)$, $\varepsilon_{T-1}^{RS}(x) = Low^p$.
- iii. for all $x \leq x_H^{RS}$, there exists a time threshold $\tau_H^{RS}(x)$ nondecreasing with x , such that $\varepsilon_t^{RS}(x) = High$ for all $\tau_H^{RS}(x) < t < T$,
 - a. if $\alpha^p(c_H^\ell - c_{Lp}^\ell) < \alpha^\ell(c_H^p - c_{L\ell}^p)$, $\varepsilon_t^{RS}(x) = Low^\ell$ for all $x \leq t \leq \tau_H^{RS}(x)$.

- b. if $\alpha^p(c_H^\ell - c_{L^p}^\ell) \geq \alpha^\ell(c_H^p - c_{L^\ell}^p)$, there exists a time threshold $\tau_L^{RS}(x) < \tau_H^{RS}(x)$, such that $\varepsilon_t^{RS}(x) = Low^p$ for all $\tau_L^{RS}(x) \leq t \leq \tau_H^{RS}(x)$ and $\varepsilon_t^{RS}(x) = Low^\ell$ for all $x \leq t \leq \tau_L^{RS}(x)$.

Note that using the capitation payment alone cannot motivate the provider to work collaboratively with the leader. The reason for this is the fact that the provider might find providing less than efficient quality of care more beneficial. When the payments to the provider are connected to the overall performance of the IPU at the end of disease cycle, then the leader and provider will make decisions informatively based on the expected benefits of treatments for the patient and the expected costs of exerting higher efforts.

In this study, we combined the capitation payment with reward-sharing contracts. We find that such a payment system can invoke collaboration among the players. The results show that they create similar results to the first-best solution. Nevertheless, the providers start working in *Low* mode with smaller number of failures, $x_H^{RS} < x_H^{fb}$ in the reward-sharing contract. One of the difficulties in the implementation of such a contract is when multiple equilibria exist, i.e. when both Low^p and Low^ℓ is optimal. In deriving the results, we assumed the healthcare leader will determine the working mode when there are multiple equilibria. But, in practice this situation might result in conflicts and implementation challenges.

4.5.2 Noncontractible Reward

When it is difficult to assess and contract on the value of health outcomes, rewards are not contractible. Furthermore, the stochasticity of health outcomes could be exacerbated by patient's noncompliance to the treatment regimen. In this case, any of the players could assert that the failure of healthcare delivery is due to

other factors. In this part, we will investigate the effects of an alternative payment system that can incorporate the noncontractability of health outcomes. When the outcomes are not contractible, relational contracts have proved to alleviate the problem of hidden action [73, 77].

We study the effects of using a relational contract which is a combination of formal and informal contracts. Relational contracts were first introduced by Levin [73], where he found the optimal contract can take a form of simple stationary contract. Levin found that with the hidden action problem the self-enforced optimal contract is comprised of two levels of compensation. The principal sets a base payment independent of the agent's performance. The voluntary payment to the agent is "one-step:" a bonus if output exceeds a threshold. In another study Plambeck et al. [77] studied the relational contract in the context of infinitely repeated relationships in joint production. Plambeck and colleagues found that the ongoing relationship facilitates the self-enforcement of relational contracts. The optimal contract has a simple form and does not depend on the past history. Using the results of Plambeck et al. study, we will study the collaboration dynamics between the providers using a memoryless contract, where the agreed upon action depends on the current state and payments depend only on the observed transition. To discourage free riding, the optimal contract imposes termination to jointly punish the providers following undesirable outcomes.

The relational contract is comprised of a formal (court-enforced) contract and a discretionary payment. With the formal contract the healthcare leader pays a constant amount f at the beginning of each period to the healthcare provider, even if the health outcomes are suffering. The discretionary payment d_t depends on the performance at the end of period t . The discretionary payment may depend

on the public history at the end of period t and transition to the state x_{t+1} . We will assume the provider will receive a constant payment $f \geq c_H^p$ in each period.

The IPU's objective is to minimize the number of failures x . Hence, we will assume the discretionary payment to the provider $d_t(x)$ is decreasing with x , $d_t(x) > d_t(x+1)$ and increasing with t , $d_t(x) < d_{t+1}(x)$.

The healthcare provider will choose his effort level to maximize

$$\begin{aligned}
V_t^p(x|j) &= -c_j^p + f + \mathbb{E}_j [d_t(x + \xi) + \delta V_{t+1}^p(x + \xi)] \text{ for } t = 1, 2, \dots, T-1 \\
V_t^p(x|Stop) &= -k^p, \text{ for } t = 1, 2, \dots, T-1 \\
V_T^p(x) &= 0
\end{aligned} \tag{4.12}$$

The healthcare leader is responsible to pay f and d_t to the healthcare provider, whereas the rewards for the IPU's performance $R(x)$ will be collected by the healthcare leader at the end of disease cycle T .

$$\begin{aligned}
V_t^\ell(x|j) &= -c_j^\ell - f + \mathbb{E}_j [-d_t(x + \xi) + \delta V_{t+1}^\ell(x + \xi)], \text{ for } t = 1, 2, \dots, T-1 \\
V_t^\ell(x|Stop) &= -k^\ell, \text{ for } t = 1, 2, \dots, T-1 \\
V_T^\ell(x) &= R(x),
\end{aligned} \tag{4.13}$$

where

$$\begin{aligned}
\mathbb{E}_j [d_t(x + \xi) + \delta V_{t+1}^p(x + \xi)] &= p_j [\delta V_{t+1}^p(x + 1) + d_t(x + 1)] \\
&\quad + (1 - p_j) [\delta V_{t+1}^p(x) + d_t(x)]
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}_j [-d_t(x + \xi) + \delta V_{t+1}^\ell(x + \xi)] &= p_j [\delta V_{t+1}^\ell(x + 1) - d_t(x + 1)] \\
&\quad + (1 - p_j) [\delta V_{t+1}^\ell(x) - d_t(x)].
\end{aligned}$$

$$\mathbb{E}_j [d_t(x + \xi) + \delta V_{t+1}^p(x + \xi)] \geq 0 \quad (4.14)$$

$$\mathbb{E}_j [-d_t(x + \xi) + \delta V_{t+1}^\ell(x + \xi)] \geq 0 \quad (4.15)$$

constraints (4.14) and (4.15) capture the expected payoff for the healthcare leader and provider from continuing the partnership. Since *Stop* is the most credible punishment that can be imposed on a player that fails to execute the relational contract, constraints (4.14) and (4.15) are necessary conditions for the relational contract to be self-enforcing. If a contract is self-enforcing, then no party has incentive to deviate unilaterally.

The next proposition characterizes the equilibrium collaboration dynamics under relational contracts. Figure 4.5 illustrates the results.

We will first define the following threshold for the healthcare provider:

$$\theta_{LH}^{(R)p} = \frac{c_H^p - c_L^\ell}{p_H - p_L}$$

In period $T-1$, the provider prefers *Low* ^{ℓ} over *High* if and only if $\Delta d_{T-1}(x) \geq \theta_{LH}^{(R)p}$. Because, we assume $d_t(x)$ is decreasing with x , $\Delta d_t(x) < 0$. Further, we assume $d_t(x)$ is convex, hence $\Delta d_t(x)$ has increasing differences with x . As a result, there is a threshold $x_H^{(R)p}$ as defined below, where the provider would prefer working mode *High* for all $x \leq x_H^{(R)p}$ in period $T-1$.

$$x_H^{(R)p} = \max\{x \in \mathbb{Z}^+ | \Delta d_T(x) \leq \theta_{LH}^{(R)p}\} \quad (4.16)$$

At any $t < T-1$ the healthcare provider will choose *Low* ^{ℓ} over *High* if and only if

$$\delta \Delta V_{t+1}^p(x) + \Delta d_t(x) \geq \theta_{LH}^{(R)p}. \quad (4.17)$$

By comparing the value function for the healthcare leader, We will define $\theta_{LH}^{(R)\ell}$ as healthcare leader's threshold.

$$\theta_{LH}^{(R)\ell} = \frac{c_H^\ell - c_{L^p}^\ell}{p_H - p_L}$$

The healthcare leader will prefer *Low^p* over *High* at period $T - 1$, if and only if

$$\delta\Delta R(x) \geq \theta_{LH}^{(R)\ell}.$$

At any $t < T - 1$ the healthcare provider will choose *Low^p* over *High* if and only if

$$\delta\Delta V_{t+1}^\ell(x) - \Delta d_t(x) \geq \theta_{LH}^{(R)\ell} \quad (4.18)$$

We define

$$x_H^{(R)\ell} = \max\{x \in \mathbb{Z}^+ | \delta\Delta R(x) \leq \theta_{LH}^{(R)\ell}\} \quad (4.19)$$

for all $x \leq x_H^{(R)\ell}$, indicating that the healthcare leader prefers working mode *High* over *Low^p* at period $T - 1$.

Finally, we will define x_H^R and x_L^R as follows

$$x_H^R = \min\{x_H^{(R)p}, x_H^{(R)\ell}\}$$

and

$$x_L^R = \max\{x_H^{(R)p}, x_H^{(R)\ell}\}$$

For each state $x < x_L^R$, let us define

$$\bar{\tau}(x) := \{t \in T | \delta\Delta V_{t+1}^p(x) + \Delta d_t(x) \geq \theta_{LH}^{(R)p}\}$$

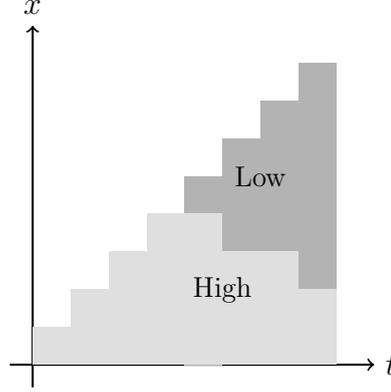


FIGURE 4.5: Dynamics of collaboration among providers in an IPU in a relational contract

$$\tau(x) := \{t \in T \mid \delta \Delta V_{t+1}^\ell(x) - \Delta d_t(x) \geq \theta_{LH}^{(R)^\ell}\}$$

if there exists such a t , otherwise $\tau(x) = x$ and $\bar{\tau}(x) = T$.

Proposition 4.3. *There exist thresholds for the number of failures x_H^R and x_L^R*

- i. $\varepsilon_t^R(x) = Low$ for all t and $x > x_L^R$.
- ii. for all $x < x_L^R$ there exists time thresholds $\tau(x)$ and $\bar{\tau}(x)$ such that $\varepsilon_t^R(x) = Low^p$ for $x \leq t < \tau(x)$, $\varepsilon_t^R(x) = High$ for $\tau(x) \leq t \leq \bar{\tau}(x)$, and $\varepsilon_t^R = Low^\ell$ for $\bar{\tau}(x) < t < T$.

The results from the relational contract suggest that the healthcare provider would prefer to work collaboratively sooner so that he can collect the rewards from keeping the patient healthy. However, the healthcare leader will increase his efforts as time passes and when the patient is in good health. The optimal policy for the IPU is to work in *High* mode earlier in the disease cycle. Comparing the results of relational contracts with the first-best solution indicates that the IPU members start working in *Low* mode sooner, $x_H^R \leq x_H^{fb}$. Comparison of the relational contract with reward-sharing contract suggests that there exists time

periods for which a relational contract could result in more collaborative work than the reward-sharing contract. Particularly, the results from the reward-sharing contract is similar to the first-best solution, but the collaboration might emerge later in the care process. With the relational contracts, discretionary payments at the end of each period can encourage the provider to work more earlier in the process while the reward-sharing contract stimulates collaboration at the same time for both the provider and the leader.

4.6 Conclusions

We studied how different contractual arrangements affect the collaboration dynamics among the IPU members in treating a patient over a finite time horizon. We identify that providers would prefer to work in *Low* mode when the number of failures have exceeded a threshold. We find that the IPU members prefer to work in *High* mode when the patient is showing good results to the high quality treatments. In this case, they want to keep the failures to a minimum and therefore, would provide high quality treatments. However, when the number of failures increases, the IPU members would prefer low quality treatments sooner.

When facing the double-sided hidden action problem, if the efforts are not contractible, the collaboration dynamics depend on the type of contract. We considered two types of hybrid contracts. First, we studied capitation with a reward-sharing contract. Second, we used the combination of formal and informal contracts to study how relational contracts can affect the collaboration dynamics. The reward-sharing contract resulted in similar dynamics to the first-best solution, but with less collaboration overall. Reward sharing contract performed very well,

nonetheless its implementation and applicability depends on the contractability of rewards.

We used relational contracts to evaluate the collaboration dynamics when the rewards are not be contractible. This type of contracts have a formal, court-enforced, component. In the IPU, the provider will be paid a fixed-amount no matter what the health outcomes are. Relational contracts also have a discretionary payment component. We considered a general form of discretionary payment, which is decreasing with the number of failures. We found that the provider is encouraged to work in collaboration sooner than the leader. The provider's incentive stems from the discretionary payment at the end of each period. Nonetheless, the healthcare leader and IPU's performance will be assessed at the end of disease cycle. The convolution of these preferences can result in more collaboration than the reward-sharing contract.

This work can be extended in several ways. First, the assumptions of the model can be generalized by considering continuous time models, constant rewards for sustaining patients' health, or stopping the collaboration without any costs. Second, the effect of learning among the providers could be studied by considering decreasing costs throughout time. Our model could be generalized to study the effect of having multiple providers within an IPU. More fundamentally, the IPU members might not behave rationally. Increasing or decreasing non-monetary costs of collaboration could also alter the collaboration dynamics.

The IPU members have to answer questions regarding the team management and resource allocation. It is valuable to study how the existence of a healthcare leader affects the dynamics. Specifically, what would happen to the dynamics of collaboration if the IPU members work without a leader? Would having no

leader affect how much resources the providers allocate to specific treatments? Investigating these questions can result in meaningful insights.

4.A Appendix: Chapter 4 Proofs

4.A.1 The First-Best Solution

We define the following threshold:

$$\theta_{LH}^{fb} = \frac{(c_H^p + c_H^\ell) - (c_{L^p}^p + c_{L^p}^\ell)}{\delta(p_H - p_L)} \quad (4.20)$$

Threshold θ_{LH}^{fb} along with equations (4.3) allow us to compare working modes *High* and *Low*. Working mode *Low* is preferred over working mode *High* in period $T - 1$ if and only if $R(x + 1) - R(x) \geq \theta_{LH}^{fb}$.

We define x_H^{fb}

$$x_H^{fb} = \min\{x \in \mathbb{Z}^+ | \Delta R(x) \geq \theta_{LH}^{fb}\} \quad (4.21)$$

We denote, for any t , the highest state where $\varepsilon_t^{fb}(x) = High$, as

$$x_{H,t}^{fb} = \left\{ \max\{x \in \mathbb{Z}^+ | V_{t+1}(x+1) - V_{t+1}(x) \leq \theta_{LH}^{fb}\} \right\}. \quad (4.22)$$

For any $x < x_H^{fb}$, we will define $\tau(x) := \max\{t \in T | \Delta V_{t+1}(x) \leq \theta_{LH}^{fb}\}$, if there exists such a t that $\Delta V_{t+1}(x) \leq \theta_{LH}^{fb}$; otherwise, $\tau(x) = x - 1$.

Throughout the proofs we will use the notation $\mathbb{E}_j[V_t(x + \xi)]$

$$\mathbb{E}_j[V_t(x + \xi)] = p_j V_t(x + 1) + (1 - p_j) V_t(x)$$

Lemma 4.1. *For any x , $V_t(x)$ is nondecreasing in t .*

Proof. We proof this lemma by induction on t . When $t = T - 1$, $V_T(x) = R(x) \geq V_{T-1}(x)$. To see this, note that $R(x)$ is decreasing with x , $R(x+1) \leq R(x)$. Hence, $\delta[p_j R(x+1) + (1-p_j)R(x)] \leq \delta R(x) \leq R(x)$. Since $V_{T-1}(x) = -c^j + \delta[p_j R(x+1) + (1-p_j)R(x)]$, $V_T(x) \geq V_{T-1}(x)$.

Fix $t < T - 1$ and suppose that $V_{t+1}(x) \geq V_t(x)$, $\forall x$. Then,

$$\begin{aligned} V_t(x) &= \max \{-K, -c_H + \delta \mathbb{E}_H[V_{t+1}(x + \xi)], -c_H + \delta \mathbb{E}_L[V_{t+1}(x + \xi)]\} \\ &\geq \max \{-K, -c_H + \delta \mathbb{E}_H[V_t(x + \xi)], -c_H + \delta \mathbb{E}_L[V_t(x + \xi)]\} = V_{t-1}(x) \end{aligned}$$

□

Lemma 4.2. $\varepsilon_t^{fb}(x) = Low$ for all t if and only if $x \geq x_H^{fb}$.

Proof. This proof uses lemma (4.1). Using equations (4.2), we obtain $\varepsilon_t^{fb}(x) = Low \Leftrightarrow \Delta R(x) \geq \theta_{LH}^{fb}$. Since $R(x)$ is decreasing and $\Delta R(x) \leq 0$ for all x , $\Delta R(x)$ is increasing with x . Further, $\theta_{LH}^{fb} < 0$ and does not vary with x , $\varepsilon_{T-1}^{fb}(x) = Low$ if $x \geq x_H^{fb}$.

Next, we show by induction on t that if $\varepsilon_{T-1}^{fb}(x) = Low$, then $\varepsilon_t^{fb}(x+k) = Low \forall k \geq 0$ and $t \leq T - 1$. When $t = T - 1$, $\varepsilon_{T-1}^{fb}(x+k) = Low \forall k \geq 0$. For any $t < T - 1$, suppose that $\varepsilon_{t+1}^{fb}(x+k) = Low \forall k \geq 0$, therefore $\Delta V_{t+2}(x+k) \geq \theta_{LH}^{fb}$. Applying the condition $\Delta V_{t+1}(x+k) = V_{t+1}(x+k+1) - V_{t+1}(x+k)$, $\Delta V_{t+1}(x+k) = -c_L + \delta \mathbb{E}_L[V_{t+2}(x+k+1+\xi) + c_L - \delta \mathbb{E}_L[V_{t+2}(x+k+\xi)]] = \delta p_L \Delta V_{t+2}(x+k+1) + \delta(1-p_L) \Delta V_{t+2}(x+k)$. As a result $\Delta V_{t+1}(x+k) \geq \delta p_L \theta_{LH}^{fb} + \delta(1-p_L) \theta_{LH}^{fb} \geq \delta \theta_{LH}^{fb}$. Since $\theta_{LH}^{fb} < 0$ for all x , $\delta \theta_{LH}^{fb} \geq \theta_{LH}^{fb}$. Thus, $\Delta V_{t+1}(x+k) \geq \theta_{LH}^{fb}$, i.e. $\varepsilon_t^{fb}(x) = Low$, completing the induction step. □

Lemma 4.3. *If $\varepsilon_{t+1}^{fb}(x+1) = \varepsilon_{t+1}^{fb}(x) = Low$, then $\varepsilon_t^{fb}(x) = Low$.*

Proof. Because If $\varepsilon_{t+1}^{fb}(x+1) = \varepsilon_{t+1}^{fb}(x) = Low$, $\Delta V_{t+2}(x+1) \geq \theta_{LH}^{fb}$ and $\Delta V_{t+2}(x) \geq \theta_{LH}^{fb}$. Moreover, since $\Delta V_{t+1}(x) = V_{t+1}(x+1) - V_{t+1}(x)$, $\Delta V_{t+1}(x) = -c_L + \delta \mathbb{E}_L[V_{t+2}(x+1+\xi) + c_L - \delta \mathbb{E}_L[V_{t+2}(x+\xi)]] = \delta p_L \Delta V_{t+2}(x+1) + \delta(1-p_L) \Delta V_{t+2}(x)$. Replacing $\Delta V_{t+2}(x+1) \geq \theta_{LH}^{fb}$ and $\Delta V_{t+2}(x) \geq \theta_{LH}^{fb}$, results in $\Delta V_{t+1}(x) \geq \delta p_L \theta_{LH}^{fb} + \delta(1-p_L) \theta_{LH}^{fb} \geq \delta \theta_{LH}^{fb}$. Since $\theta_{LH}^{fb} < 0$ for all x , $\delta \theta_{LH}^{fb} \geq \theta_{LH}^{fb}$. Thus, $\Delta V_{t+1}(x+k) \geq \theta_{LH}^{fb}$, i.e. $\varepsilon_t^{fb}(x) = Low$, completing the induction step. \square

Lemma 4.4. For any t , $x_{H,t}^{fb}$ defined in (4.22), is nondecreasing in t , i.e. $x_{H,t}^{fb} \leq x_{H,t+1}^{fb}$.

Proof. By definition, $\varepsilon_t^{fb}(x) \neq High$ for all $x > x_{H,t}^{fb}$, or $\varepsilon_t^{fb}(x) = Low$ for all $x > x_{H,t}^{fb}$. Suppose by contradiction that $x_{H,t}^{fb} > x_{H,t+1}^{fb}$ for some t . Then, there must exist some x , $x_{H,t}^{fb} \geq x > x_{H,t+1}^{fb}$, such that $\varepsilon_t^{fb}(x) = High$ and $\varepsilon_{t+1}^{fb}(x) = Low$. If $\varepsilon_{t+1}^{fb}(x+1) = Low$ by lemma 4.3, we should then have had $\varepsilon_t^{fb}(x) = Low$, a contradiction. \square

Lemma 4.5. For all t and all x , $V_t(x)$ has decreasing differences in (x, t) , i.e. $\Delta V_t(x) \geq \Delta V_{t+1}(x)$.

Proof. The proof uses lemmas 4.1 and 4.2. Consider period $T-1$ first. Since $R(x)$ is decreasing and convex, $\Delta R(x+1) \geq \Delta R(x)$ we have

$$\begin{aligned} \Delta V_{T-1}(x) &= -c_j + \delta \mathbb{E}_j[R(x+1+\xi)] + c_j - \delta \mathbb{E}_j[R(x+\xi)] \\ &= \delta p_j \Delta R(x+1) + \delta(1-p_j) \Delta R(x) \\ &\geq \delta p_j \Delta R(x) + \delta(1-p_j) \Delta R(x) \geq \Delta V_T(x) \end{aligned}$$

Next for any $t < T - 1$, suppose $\Delta V_{t+1}(x) \geq \Delta V_{t+2}(x)$.

$$\begin{aligned} \Delta V_t(x) &= -c_j + \delta \mathbb{E}_j[V_{t+1}(x+1+\xi)] + c_j - \delta \mathbb{E}_j[V_{t+1}(x+\xi)] \\ &= \delta p_j \Delta V_{t+1}(x+1) + \delta(1-p_j) \Delta V_{t+1}(x) \\ &\geq \delta p_j \Delta V_{t+2}(x+1) + \delta(1-p_j) \Delta V_{t+2}(x) \geq \Delta V_{t+1}(x) \end{aligned}$$

in which the inequality follows by induction hypothesis. This completes the induction step. \square

Lemma 4.6. *For all $x < x_L^{fb}$, there exists a time threshold $\tau(x)$, nondecreasing in x , such that $\varepsilon_t^{fb} = High$ for all $\tau(x) < t < T$ and $\varepsilon_t^{fb} = Low$ for all $x \leq t \leq \tau(x)$.*

Proof. Using lemma 4.5, $\Delta V_t(x) \geq \Delta V_{t+1}(x)$ for each x . By equations (4.3), if $\Delta V_t(x) \leq \theta_{LH}^{fb}$ the optimal working mode $\varepsilon_t^{fb} = High$. The decreasing differences for $V_t(x)$ assures there exists a time threshold $\tau(x)$ for which $\Delta V_t(x) = \theta_{LH}^{fb}$. Hence by lemma 4.5, $\Delta V_t(x) \leq \theta_{LH}^{fb}$ will hold for $\tau(x) < t < T$, i.e. $\varepsilon_t^{fb} = High$ for all $\tau(x) < t < T$. \square

Proof of Proposition 4.1. This proof uses lemmas 1-6 in this appendix.

- i. This is shown in lemma 4.2.
- ii. This is shown in lemma 4.4.
- iii. This is shown in lemma 4.6.

\square

4.A.2 Contractible Reward

Lemma 4.7. *For any x , $V_t^i(x)$ and $V_t^{-i}(x)$ are nondecreasing in t , which is $V_t^i(x) \leq V_{t+1}^i(x)$ and $V_t^{-i}(x) \leq V_{t+1}^{-i}(x)$.*

Proof. We proof this lemma by induction on t . When $t = T-1$, $V_T^i(x) = \alpha^i R(x) \geq V_{T-1}^i(x)$. To see this, note that $R(x)$ is decreasing with x , $R(x+1) \leq R(x)$. Hence, $V_{T-1}^i(x) = -c_j^i + \delta\alpha^i[p_j R(x+1) + (1-p_j)R(x)] \leq \delta\alpha^i R(x) \leq \alpha^i R(x) = V_T^i(x)$, $V_{T-1}^{-i}(x) \leq V_T^{-i}(x)$.

Fix $t < T-1$ and suppose that $V_{t+1}^i(x) \geq V_t^i(x)$, $\forall x$. Then,

$$\begin{aligned} V_t^i(x) &= -c_j^i + \delta\mathbb{E}_H[V_{t+1}^i(x+\xi)] \\ &\geq -c_j^i + \delta\mathbb{E}_H[V_t^i(x+\xi) = V_{t-1}^i(x)], \end{aligned}$$

completing the induction step. Similar deduction and proofs hold for $V_t^{-i}(x)$. \square

Lemma 4.8. $\varepsilon_t^{RS}(x) = Low$ for all t if and only if $x \geq x_L^{RS}$.

Proof. This proof uses lemma (4.7). Using equations (4.5), (4.6), and (4.5.1), we obtain $\varepsilon_t^{RS}(x) = Low^i \Leftrightarrow \Delta R(x) \geq \theta_{L^i H}^{RS}$. Since $R(x)$ is decreasing and $\Delta R(x) \leq 0$ for all x , $\Delta R(x)$ is increasing with x . Further, $\theta_{L^i H}^{RS} < 0$ and does not vary with x , $\varepsilon_{T-1}^{RS}(x) = Low^i$ if $x \geq x_L^{RS}$.

Next, we show by induction on t that if $\varepsilon_{T-1}^{RS}(x) = Low^i$, then $\varepsilon_t^{RS}(x+k) = Low^i \forall k \geq 0$ and $t \leq T-1$. When $t = T-1$, $\varepsilon_{T-1}^{RS}(x+k) = Low \forall k \geq 0$. For any $t < T-1$, suppose that $\varepsilon_{t+1}^{RS}(x+k) = Low \forall k \geq 0$, therefore $\Delta V_{t+2}^i(x+k) \geq \theta_{L^i H}^{RS}$. Applying the condition $\Delta V_{t+1}^i(x+k) = V_{t+1}^i(x+k+1) - V_{t+1}^i(x+k)$, $\Delta V_{t+1}^i(x+k) = -c_L^i + \delta\mathbb{E}_L[V_{t+2}^i(x+k+1+\xi)] + c_L^i - \delta\mathbb{E}_L[V_{t+2}^i(x+k+\xi)] = \delta p_L \Delta V_{t+2}^i(x+k+1) + \delta(1-p_L) \Delta V_{t+2}^i(x+k)$. As a result $\Delta V_{t+1}^i(x+k) \geq \delta p_L \theta_{L^i H}^{RS} + \delta(1-p_L) \theta_{L^i H}^{RS} \geq \delta \theta_{L^i H}^{RS}$. Since

$\theta_{L^i H}^{RS} < 0$ for all x , $\delta\theta_{LH}^{RS} \geq \theta_{LH}^{fb}$. Thus, $\Delta V_{t+1}^i(x+k) \geq \theta_{L^i H}^{RS}$, i.e. $\varepsilon_t^{RS}(x) = Low^i$, completing the induction step. \square

Lemma 4.9. *If $\varepsilon_{t+1}^{RS}(x+1) = Low^i$ and $\varepsilon_{t+1}^{RS}(x) = Low^i$ or, $\varepsilon_t^{RS}(x) = Low^i$.*

Proof. If $\varepsilon_{t+1}^{RS}(x+1) = \varepsilon_{t+1}^{RS}(x) = Low^i$, $\Delta V_{t+2}^i(x+1) \geq \alpha^i \theta_{L^i H}^{RS}$ and $\Delta V_{t+2}^i(x) \geq \alpha^i \theta_{L^i H}^{RS}$. Since $\Delta V_{t+1}^i(x) = V_{t+1}^i(x+1) - V_{t+1}^i(x)$, $\Delta V_{t+1}^i(x) = -c_j^i + \delta \mathbb{E}_L[V_{t+2}^i(x+1+\xi) + c_j^i - \delta \mathbb{E}_L[V_{t+2}^i(x+\xi)]] = \delta p_L \Delta V_{t+2}^i(x+1) + \delta(1-p_L) \delta V_{t+2}^i(x) \geq \alpha^i \delta p_L \theta_{L^i H}^{RS} + \alpha^i \delta(1-p_L) \theta_{L^i H}^{RS} \geq \alpha^i \theta_{L^i H}^{RS}$. Therefore, $\Delta V_{t+1}^i(x) \geq \alpha^i \theta_{L^i H}^{RS}$, i.e. $\varepsilon_t^{RS}(x) = Low^i$, completing the induction step. \square

We denote, for any t , the highest state where $\varepsilon_t^{RS}(x) = High$, as

$$x_{H,t}^{RS} = \max \{x \leq x_{H,t+1}^{RS} | V_{t+1}^i(x+1) - V_{t+1}^i(x) \leq \alpha^i \theta_{L^i H}^{RS}\} \quad (4.23)$$

Lemma 4.10. *For any t , $x_{H,t}^{RS}$ is nondecreasing in t , i.e. $x_{H,t}^{RS} \leq x_{H,t+1}^{RS}$.*

Proof. By definition, $\varepsilon_t^{RS}(x) \neq High$ for all $x > x_{H,t}^{RS}$, or $\varepsilon_t^{RS}(x) = Low^p$ or $\varepsilon_t^{RS}(x) = Low^\ell$ for all $x > x_{H,t}^{RS}$. Suppose by contradiction that $x_{H,t}^{RS} > x_{H,t+1}^{RS}$ for some t . Then, there must exist some x , $x_{H,t}^{RS} \geq x > x_{H,t+1}^{RS}$, such that $\varepsilon_t^{RS}(x) = High$ and $\varepsilon_{t+1}^{RS}(x) = Low$. If $\varepsilon_{t+1}^{RS}(x+1) = Low$ and $\varepsilon_{t+1}^{RS}(x) = Low$ by lemma 4.9, we should then have had $\varepsilon_t^{RS}(x) = Low$, a contradiction. \square

We denote, for any t , the lowest state where $\varepsilon_t^{RS}(x) = Low^\ell$, as

$$x_{L,t}^{RS} = \min \{x \leq x_{L,t+1}^{RS} | V_{t+1}^p(x+1) - V_{t+1}^p(x) \leq \alpha^p \theta_{L^p H}^{RS}\} \quad (4.24)$$

Lemma 4.11. *For any t , $x_{L,t}^{RS}$ is nondecreasing in t , i.e. $x_{L,t}^{RS} \leq x_{L,t+1}^{RS}$.*

Proof. By definition, $\varepsilon_t^{RS}(x) \neq High$ for all $x > x_{H,t}^{RS}$, or $\varepsilon_t^{RS}(x) = Low^p$ or $\varepsilon_t^{RS}(x) = Low^\ell$ for all $x > x_{H,t}^{RS}$. If $x_{L,t}^{RS} = x_{H,t}^{RS}$, by lemma 4.10 the proof is complete.

Next, suppose, $x_{L,t}^{RS} \neq x_{H,t}^{RS}$, i.e. $\min\{\theta_{L^p H}^{RS}, \theta_{L^\ell H}^{RS}\} = \theta_{L^\ell H}^{RS}$. by contradiction that $x_{L,t}^{fb} > x_{L,t+1}^{fb}$ for some t . Then, there must exist some x , $x_{L,t}^{RS} \geq x > x_{L,t+1}^{RS}$, such that $\varepsilon_t^{RS}(x) = Low^\ell$ and $\varepsilon_{t+1}^{RS}(x) \neq Low^\ell$. If $\varepsilon_{t+1}^{RS}(x+1) = Low^p$ and $\varepsilon_{t+1}^{RS}(x) = Low^p$ by lemma 4.9, we should then have had $\varepsilon_t^{RS}(x) = Low^p$, a contradiction. \square

Lemma 4.12. $\Delta V_t^i(x)$ has nonincreasing differences in (x, t) , i.e. $\Delta V_t^i(x) \geq \Delta V_{t+1}^i$.

Proof. The proof uses lemma 4.7. Consider period $T - 1$ first. Since $R(x)$ is decreasing and convex, $\alpha^i \Delta R(x+1) \geq \alpha^i \Delta R(x)$ we have

$$\begin{aligned} \Delta V_{T-1}^i(x) &= -c_j^i + \alpha^i \delta \mathbb{E}_j[R(x+1+\xi)] + c_j - \alpha^i \delta \mathbb{E}_j[R(x+\xi)] \\ &= \alpha^i \delta [p_j \Delta R(x+1) + (1-p_j) \Delta R(x)] \\ &\geq \alpha^i \delta [p_j \Delta R(x) + (1-p_j) \Delta R(x)] \geq \Delta V_T^i(x) \end{aligned}$$

Next for any $t < T - 1$, suppose $\Delta V_{t+1}^i(x) \geq \Delta V_{t+2}^i(x)$.

$$\begin{aligned} \Delta V_t^i(x) &= -c_j + \alpha^i \delta [\mathbb{E}_j[V_{t+1}(x+1+\xi)] + c_j - \mathbb{E}_j[V_{t+1}(x+\xi)]] \\ &= \alpha^i \delta [p_j \Delta V_{t+1}^i(x+1) + (1-p_j) \Delta V_{t+1}^i(x)] \\ &\geq \alpha^i \delta [p_j \Delta V_{t+2}^i(x+1) + \delta(1-p_j) \Delta V_{t+2}^i(x)] \geq \Delta V_{t+1}^i(x) \end{aligned}$$

in which the inequality follows by induction hypothesis. This completes the induction step. \square

Lemma 4.13. $\varepsilon_t^{RS}(x) \neq High$ for $x < x_H^{RS}$ for all t . If $\alpha^p(c_H^\ell - c_{L^p}^\ell) < \alpha^\ell(c_H^p - c_{L^\ell}^p)$, $\varepsilon_{T-1}^{RS}(x) = Low^\ell$.

Proof. When $\alpha^p(c_H^\ell - c_{L^p}^\ell) < \alpha^\ell(c_H^p - c_{L^\ell}^p)$, $\theta_{L^\ell H}^{RS} > \theta_{L^p H}^{RS}$, i.e. $\min\{\theta_{L^p H}^{RS}, \theta_{L^\ell H}^{RS}\} = \theta_{L^p H}^{RS}$. According to the decision rules, if $\Delta R(x) \geq \theta_{L^p H}^{RS}$, $\varepsilon_t^{RS}(x) = Low^\ell$, for all $x \geq x_H^{RS}$. Since $R(x)$ is decreasing and convex in x , $\Delta R(x)$ has increasing differences with x and there exists a threshold x_L^{RS} for which $\Delta R(x) \geq \max\{\theta_{L^p H}^{RS}, \theta_{L^\ell H}^{RS}\}$, for all $x \geq x_L^{RS}$, providers are indifferent between Low^p and Low^ℓ . When multiple working modes are possible, we assume the healthcare leader will determine which mode will be selected. Specifically, we assume the healthcare leader will choose Low^ℓ . $\theta_{L^\ell H}^{RS} x \geq x_H^{RS}$, $\varepsilon_t^{RS}(x) = Low^\ell$. We will assume the healthcare leader will prefer Low^ℓ when multiple equilibria exists. \square

Lemma 4.14. $\varepsilon_t^{RS}(x) \neq High$ for $x \geq x_H^{RS}$ for all t , if $\alpha^p(c_H^\ell - c_{L^p}^\ell) \geq \alpha^\ell(c_H^p - c_{L^\ell}^p)$, $\varepsilon_{T-1}^{RS}(x) = Low^p$ for all $x_H^{RS} < x < x_L^{RS}$.

Proof. When $\alpha^p(c_H^\ell - c_{L^p}^\ell) \geq \alpha^\ell(c_H^p - c_{L^\ell}^p)$, $\min\{\theta_{L^p H}^{RS}, \theta_{L^\ell H}^{RS}\} = \theta_{L^\ell H}^{RS}$. Therefore, $\varepsilon_{T-1}^{RS} = Low^p$ for $x_H^{RS} \leq x \leq x_L^{RS}$ and $\varepsilon_{T-1}^{RS} = Low^\ell$ and for $x > x_L^{RS}$, $\varepsilon_t^{RS} = Low^p$. \square

For any $x < x_H^{RS}$, we will define $\tau_H^{RS} := \min\{t \in T | \varepsilon_t^{RS}(x) = High\}$, if there exists such a t that for both players $\Delta V_{t+1}^i(x) \leq \alpha^i \theta_{L^i H}^{RS}$; otherwise, $\tau_H^{RS}(x) = T$.

For any $x < x_H^{RS}$, we will define $\tau_L^{RS}(x) := \max\{t \in T | \varepsilon_t^{RS}(x) = Low^\ell\}$, i.e. $\Delta V_{t+1}^p(x) \geq \alpha^p \theta_{L^p H}^{RS}$ if there exists such a t ; otherwise, $\tau_L^{RS}(x) = x - 1$.

Lemma 4.15. for all $x \leq x_H^{RS}$, there exists a time threshold $\tau_H^{RS}(x)$ nondecreasing in x , such that $\varepsilon_t^{RS}(x) = High$ for all $\tau_H^{RS}(x) < t < T$.

Proof. Using lemma 4.12, which describes the nonincreasing differences for $\Delta V_t^i(x)$, i.e. $\Delta V_t^i(x) \geq \Delta V_{t+1}^i(x)$ for each x . By comparing the value functions, if $\Delta V_t^i(x) \leq \alpha^i \theta_{L^i H}^{RS}$ for both players $i \in \{p, \ell\}$ the optimal working mode $\varepsilon_t^{RS} = High$. The

nonincreasing differences for $\Delta V_t^i(x)$ assures there exists a time threshold $\tau_H^{RS}(x)$ for which $\Delta V_t^i(x) = \alpha^i \theta_{L^i H}^{RS}$. Hence by lemma 4.12, $\Delta V_t^i(x) \leq \alpha^i \theta_{L^i H}^{RS}$ will hold for $\tau_H^i(x) < t < T$, i.e. $\varepsilon_t^{RS} = High$ for all $\tau_H^{RS}(x) < t < T$. \square

Lemma 4.16. *if $\alpha^p(c_H^\ell - c_{L^p}^\ell) \geq \alpha^\ell(c_H^p - c_{L^\ell}^p)$, there exists a time threshold $\tau_L^{RS}(x) < \tau_H^{RS}(x)$, such that $\varepsilon_t^{RS}(x) = Low^p$ for all $\tau_L^{RS}(x) \leq t \leq \tau_H^{RS}(x)$ and $\varepsilon_t^{RS}(x) = Low^\ell$ for all $x \leq t \leq \tau_L^{RS}(x)$.*

Proof. When $\alpha^p(c_H^\ell - c_{L^p}^\ell) \geq \alpha^\ell(c_H^p - c_{L^\ell}^p)$, $\min\{\theta_{L^p H}^{RS}, \theta_{L^\ell H}^{RS}\} = \theta_{L^\ell H}^{RS}$. Therefore, $\varepsilon_{T-1}^{RS}(x) = Low^\ell$ for all $x > x_L^{RS}$ and $\varepsilon_{T-1}^{RS}(x) = Low^p$ for all $x_H^{RS} \leq x \leq x_L^{RS}$. Using lemma 4.19 and $\Delta V_t^i(x) \geq \Delta V_{t+1}^i(x)$ for each x , and by definition for $\tau_L^{RS}(x)$ and by lemma 4.9 if there exist a time threshold for which $\Delta V_{t+1}^p(x) \geq \alpha^p \theta_{L^p H}^{RS}$, $\varepsilon_t^{RS}(x) = Low^\ell$ for all $t \leq \tau_L^{RS}(x)$. By lemmas 4.9, 4.11, and 4.15 and definitions for $\tau_H^{RS}(x)$ and $\tau_L^{RS}(x)$, $\varepsilon_t^{RS}(x) = High$ for all $\tau_H^{RS}(x) < t < T$ and $\varepsilon_t^{RS}(x) = Low^p$ for $\tau_L^{RS} \leq t \leq \tau_H^{RS}$. \square

Proof of Proposition 4.2.

- i. This proof uses lemma 4.8, since the definition for $x > x_L^{RS}$ assures that $\Delta R(x) \geq \theta_{L^p H}^{RS}$ and $\Delta R(x) \geq \theta_{L^\ell H}^{RS}$. There are multiple equilibria, however, we assume the healthcare leader is going to make the call in this situation when both Low^p and Low^ℓ working modes are possible. The healthcare leader will be the one who exerts high effort and therefore, outcome is Low^ℓ .
- ii. for all $x_H^{RS} < x < x_L^{RS}$
 - a. the working outcome follows from equations (4.10) and (4.11) and is shown in lemma 4.13.

- b. the working outcome follows from equations (4.10) and (4.11) and is shown in lemma 4.14.
- iii. This is shown in lemma 4.10.
- iv. This is shown in lemma 4.15.
- a. if $\alpha^p(c_H^\ell - c_{L^p}^\ell) < \alpha^\ell(c_H^p - c_{L^\ell}^p)$, $\varepsilon_{T-1}^{RS}(x) = Low^\ell$ for all $x \geq x_H^{RS}$. The details are shown in lemma 4.13.
- b. This is the result of lemmas 4.13, 4.15, and 4.16.

□

4.A.3 Noncontractible Reward

To determine the working mode, the provider will compare his value function in each of the working modes, *High*, *Low^ℓ*, *Low^p*, and *Stop*. At any $t < T - 1$ the healthcare provider will choose *Low^ℓ* over *High* if and only if

$$\delta \Delta V_{t+1}^p(x) + \Delta d_t(x) \geq \theta_{LH}^{(R)p} \quad (4.25)$$

Since $c_{L^p}^p > c_{L^\ell}^p$, $\theta_{LH}^{(R)p} \geq \frac{c_H^p - c_{L^p}^p}{(p_H - p_L)}$. Therefore, if the working mode *Low^ℓ* dominates *High*, working mode *Low^p* will dominate *High* as well. As a result, we do not need to compare working modes *Low^p* and *High* for the provider.

This means that the expected change in the provider's benefits if they choose *Low^ℓ* over *High*, $\Delta d_{T-1}(x)(p_H - p_L)$, will not exceed their instant cost savings $c_H^p - c_{L^\ell}^p$, i.e. $\Delta d_{T-1}(x)(p_H - p_L) \leq c_H^p - c_{L^\ell}^p$. To see this, note that $p_H - p_L$ is negative.

Next, we will compare the value function for the leader. At any $t < T - 1$ the healthcare provider will choose *Low*^p over *High* if and only if

$$\delta \Delta V_{t+1}^\ell(x) - \Delta d_t(x) \geq \theta_{LH}^{(R)^\ell} \quad (4.26)$$

Since $d_t(x)$ is decreasing and convex, $\Delta d_t(x)$ has increasing differences with x . As a result, $-d_t(x)$ is increasing and concave

Lemma 4.17. *For all $x < x_H^R$, $\varepsilon_{T-1}^R(x) = High$.*

Proof. For any $x < x_H^R$, by definition and by increasing differences of $\Delta R(x)$ and $\Delta d_t(x)$ with x we have $\Delta R(x) \leq \Delta R(x_H^R) \leq \theta_{LH}^{(R)^\ell}$, i.e. $V_{T-1}^\ell(x|High) > V_{T-1}^\ell(x|Low^p)$. Finally, $\Delta d_{T-1}(x) \leq \Delta d_{T-1}(x_H^R - 1) \leq \theta_{LH}^{(R)^p}$, $V_{T-1}^p(x|High) > V_{T-1}^p(x|Low^\ell)$. \square

Lemma 4.18. *For all t and all x , $V_t^p(x|Low^p) > V_t^p(x|Stop)$.*

Proof. $V_t^p(x|Stop) = -k^p$, since $\mathbb{E}_j [d_t(x + \xi) + \delta V_{t+1}^p(x + \xi)] \geq 0$ and $f \geq c_j^p$, the value function for working mode *Low*^p is always nonnegative and thus $V_t^p(x|Low^p) \geq V_t^p(x|Stop)$, i.e. $\varepsilon_t^R(x) \neq Stop$. \square

Since the healthcare provider prefers to exert high effort individually, the partnership will never stop.

Lemma 4.19. *For any x , $\Delta V_t^p(x)$ has nondecreasing differences in t , i.e. $\Delta V_t^p(x) \leq \Delta V_{t+1}^p(x)$.*

Proof. The proof proceeds by induction on t . We will show the proof for $\varepsilon_t^R(x) = \varepsilon_t^R(x + 1) = High$. The same proof holds when $\varepsilon_t^R(x) = \varepsilon_t^R(x + 1) = Low^i$ and when $\varepsilon_t^R(x) = High$, $\varepsilon_t^R(x + 1) = Low^i$. Start from period $T - 1$, $\Delta V_{T-1}^p(x) =$

$-c_H^p + f + \mathbb{E}_H[d_{T-1}(x+1+\xi) + \delta V_T^p(x+1+\xi)] + c_H^p - f - \mathbb{E}_H[d_{T-1}(x+\xi) + \delta V_T^p(x+\xi)] = p_H \Delta d_{T-1}(x+1) + (1-p_H) \Delta d_{T-1}(x)$. Since both $\Delta d_{T-1}(x+1) \leq 0$ and $\Delta d_{T-1}(x) \leq 0$, $\Delta V_{T-1}^p(x) \leq \Delta V_T^p(x) = 0$.

Next, for any $t < T-1$, $\Delta V_t^p(x) = -c_H^p + f + \mathbb{E}_H[d_t(x+1+\xi) + \delta V_{t+1}^p(x+1+\xi)] + c_H^p - f - \mathbb{E}_H[d_t(x+\xi) + \delta V_{t+1}^p(x+\xi)] = p_H \Delta d_t(x+1) + (1-p_H) \Delta d_t(x) + \delta p_H \Delta V_{t+1}^p(x+1) + \delta(1-p_H) \Delta V_{t+1}^p(x)$. Suppose $\Delta V_{t+1}^p(x) \leq \Delta V_{t+2}^p(x)$ and $\Delta d_t(x) \leq \Delta d_{t+1}(x)$. Hence, $\Delta V_t^p(x) = p_H \Delta d_t(x+1) + (1-p_H) \Delta d_t(x) + \delta p_H \Delta V_{t+1}^p(x+1) + \delta(1-p_H) \Delta V_{t+1}^p(x) \leq p_H \Delta d_{t+1}(x+1) + (1-p_H) \Delta d_{t+1}(x) + \delta p_H \Delta V_{t+2}^p(x+1) + \delta(1-p_H) \Delta V_{t+2}^p(x) = \Delta V_{t+1}^p(x)$. $\Delta V_t^p(x) \leq \Delta V_{t+1}^p(x)$. This completes the induction. \square

Lemma 4.20. *For any x , $\Delta V_t^\ell(x)$ has nonincreasing differences in t , i.e. $\Delta V_t^\ell(x) \geq \Delta V_{t+1}^\ell(x)$.*

Proof. The proof proceeds by induction on t . We will show the proof for $\varepsilon_t^R(x) = \varepsilon_t^R(x+1) = High$. The same proof holds when $\varepsilon_t^R(x) = \varepsilon_t^R(x+1) = Low^i$ and when $\varepsilon_t^R(x) = High$, $\varepsilon_t^R(x+1) = Low^i$. Start from period $T-1$, $\Delta V_{T-1}^\ell(x) = -c_H^\ell - f + \mathbb{E}_H[-d_{T-1}(x+1+\xi) + \delta V_T^\ell(x+1+\xi)] + c_H^\ell + f - \mathbb{E}_H[-d_{T-1}(x+\xi) + \delta V_T^\ell(x+\xi)] = -p_H \Delta d_{T-1}(x+1) - (1-p_H) \Delta d_{T-1}(x) + p_H R(x+1) + (1-p_H) R(x)$. Since both $-\Delta d_{T-1}(x+1) \geq 0$ and $-\Delta d_{T-1}(x) \geq 0$, $\Delta V_{T-1}^\ell(x) \geq \Delta V_T^\ell(x) = R(x)$.

Next, for any $t < T-1$, $\Delta V_t^\ell(x) = -c_H^\ell - f + \mathbb{E}_H[-d_t(x+1+\xi) + \delta V_{t+1}^\ell(x+1+\xi)] + c_H^\ell + f - \mathbb{E}_H[-d_t(x+\xi) + \delta V_{t+1}^\ell(x+\xi)] = -p_H \Delta d_t(x+1) - (1-p_H) \Delta d_t(x) + p_H \Delta V_{t+1}^\ell(x+1) + (1-p_H) \Delta V_{t+1}^\ell(x)$. Suppose $\Delta V_{t+1}^\ell(x) \geq \Delta V_{t+2}^\ell(x)$ and $-\Delta d_t(x) \geq -\Delta d_{t+1}(x)$. Hence, $\Delta V_t^\ell(x) = -p_H \Delta d_t(x+1) - (1-p_H) \Delta d_t(x) + p_H \Delta V_{t+1}^\ell(x+1) + (1-p_H) \Delta V_{t+1}^\ell(x) \geq -p_H \Delta d_{t+1}(x+1) - (1-p_H) \Delta d_{t+1}(x) + p_H \Delta V_{t+2}^\ell(x+1) + (1-p_H) \Delta V_{t+2}^\ell(x) = \Delta V_{t+1}^\ell(x)$. $\Delta V_t^\ell(x) \geq \Delta V_{t+1}^\ell(x)$. This completes the induction. \square

Lemma 4.21. *If $\varepsilon_{t+1}^R(x+1) = \varepsilon_{t+1}^R(x) = Low^p$, then $\varepsilon_t^R(x) = Low^p$.*

Proof. $\varepsilon_t^R(x) = Low^p$ if (i) $V_t^\ell(x|low^p) > V_t^\ell(x|High)$, i.e. $\delta\Delta V_{t+1}^\ell(x) - \Delta d_t(x) \geq \theta_{LH}^{(R)\ell}$ and (ii) $V_t^p(x|Low^p) > V_t^p(x|Stop)$,

(i) when $\varepsilon_{t+1}^R(x+1) = \varepsilon_{t+1}^R(x) = Low^p$, $\delta\Delta V_{t+2}^\ell(x+1) - d_{t+1}(x+1) \geq \theta_{LH}^{(R)\ell}$ and $\Delta V_{t+2}^\ell(x) - d_{t+1}(x) \geq \theta_{LH}^{(R)\ell}$. Since $\Delta V_{t+1}^\ell(x) = V_{t+1}^\ell(x+1) - V_{t+1}^\ell(x)$, $\Delta V_{t+1}^\ell(x) = -c_{L^p}^\ell - f + \mathbb{E}_L[-d_{t+1}(x+1+\xi) + \delta V_{t+2}^\ell(x+1+\xi) + c_{L^p}^\ell + f - \mathbb{E}_L[-d_{t+1}(x+\xi) + \delta V_{t+2}^\ell(x+\xi)]] = p_L[-\Delta d_{t+1}(x+1) + \delta\Delta V_{t+2}^\ell(x+1)] + (1-p_L)[-\Delta d_{t+1}(x) + \delta\Delta V_{t+2}^\ell(x)] \geq p_L\theta_{LH}^{(R)\ell} + (1-p_L)\theta_{LH}^{(R)\ell} = \theta_{LH}^{(R)\ell}$. Hence, $-\Delta d_t(x) + \delta\Delta V_{t+1}^\ell(x) \geq \Delta V_{t+1}^\ell(x) \geq \theta_{LH}^{(R)\ell}$.

(ii) by lemma 4.18, for all t and x , $V_t^p(x|Low^p) > V_t^p(x|Stop)$.

□

We denote, for any t , the highest state where $\varepsilon_t^{(R)\ell}(x) = High$, as

$$x_{H,t}^{(R)\ell} = \max \left\{ x \in \mathbb{Z}^+ \mid \delta\Delta V_{t+1}^\ell(x) - \Delta d_t(x) \leq \theta_{LH}^{(R)\ell} \right\} \quad (4.27)$$

Lemma 4.22. For any t , $x_{H,t}^{(R)\ell}$ is nondecreasing in t , i.e. $x_{H,t}^{(R)\ell} \leq x_{H,t+1}^{(R)\ell}$.

Proof. By definition, $\varepsilon_t^{(R)\ell}(x) \neq High$ for all $x > x_{H,t}^{(R)\ell}$. Suppose by contradiction that $x_{H,t}^{(R)\ell} > x_{H,t+1}^{(R)\ell}$ for some t . Then, there must exist some x , $x_{H,t}^{(R)\ell} \geq x > x_{H,t+1}^{(R)\ell}$, such that $\varepsilon_t^{(R)\ell}(x) = High$ and $\varepsilon_{t+1}^{(R)\ell}(x) = Low$. If $\varepsilon_{t+1}^{(R)\ell}(x+1) = Low$ and $\varepsilon_{t+1}^{(R)\ell}(x) = Low$ by lemma 4.21, we should then have had $\varepsilon_t^{(R)\ell}(x) = Low$, a contradiction. □

Lemma 4.23. If $\varepsilon_{t+1}^{(R)p}(x+1) = \varepsilon_{t+1}^{(R)p}(x) = High$, then $\varepsilon_t^{(R)p}(x) = High$.

Proof. $\varepsilon_t^R(x) = High$ if $V_t^p(x|High) > V_t^p(x|low^\ell)$, i.e. $\delta\Delta V_{t+1}^p(x) + \Delta d_t(x) \leq \theta_{LH}^{(R)p}$. When $\varepsilon_{t+1}^R(x+1) = \varepsilon_{t+1}^R(x) = High$, $\delta\Delta V_{t+2}^p(x+1) + \Delta d_{t+1}(x) \leq \theta_{LH}^{(R)p}$

and $\delta\Delta V_{t+2}^p(x) + \Delta d_{t+1}(x) \leq \theta_{LH}^{(R)p}$. Since $\Delta V_{t+1}^p(x) = V_{t+1}^p(x+1) - V_{t+1}^p(x)$, $\Delta V_{t+1}^p(x) = -c_H^p + f + \mathbb{E}_H[d_{t+1}(x+1+\xi) + \delta V_{t+2}(x+1+\xi) + c_H^p - f - \mathbb{E}_H[d_{t+1}(x+\xi) + \delta V_{t+2}(x+\xi)]] = p_H[\Delta d_{t+1}(x+1) + \delta\Delta V_{t+2}(x+1)] + (1-p_H)[\Delta d_{t+1}(x) + \delta\Delta V_{t+2}(x)] \leq p_H\theta_{LH}^{(R)p} + (1-p_H)\theta_{LH}^{(R)p} = \theta_{LH}^{(R)p}$. As a result, $\Delta V_{t+1}^p(x) \leq \theta_{LH}^{(R)p}$. Since $\Delta d_t(x) < 0$, $\delta\Delta V_{t+1}^p(x) + \Delta d_t(x) \leq \theta_{LH}^{(R)p}$, i.e. $\varepsilon_t^{(R)p}(x) = High$. \square

We denote, for any t , the highest state where $\varepsilon_t^{(R)p}(x) = High$, as

$$x_{H,t}^{(R)p} = \max \left\{ x \in \mathbb{Z}^+ \mid \delta\Delta V_{t+1}^p(x) + \Delta d_t(x) \leq \theta_{LH}^{(R)p} \right\} \quad (4.28)$$

Lemma 4.24. *For any t , $x_{H,t}^{(R)p}$ is nonincreasing in t , i.e. $x_{H,t}^{(R)p} \geq x_{H,t+1}^{(R)p}$.*

Proof. Since $\Delta V_t^p(x)$ is nondecreasing by lemma 4.19 and by definition, $\varepsilon_t^{(R)p}(x) \neq High$ for all $x > x_{H,t}^{(R)p}$. Suppose by contradiction that $x_{H,t}^{(R)p} < x_{H,t+1}^{(R)p}$ for some t . Then, there must exist some x , $x_{H,t}^{(R)\ell} < x \leq x_{H,t+1}^{(R)\ell}$, such that $\varepsilon_t^{(R)\ell}(x) = Low$. By decreasing differences of $\Delta V_{t+1}^p(x) + \Delta d_t(x)$ in t , $\varepsilon_{t+1}^{(R)\ell}(x) = Low$. Which by definition means that $\varepsilon_{t+1}^{(R)\ell}(x+1) = Low$, a contradiction. \square

For the sake of mathematical brevity, we will denote the optimal policy in state x and time t for the provider in the relational contract with $\varepsilon_t^{(R)p}(x)$ and the optimal policy in state x and time t for the healthcare leader with $\varepsilon_t^{(R)\ell}(x)$.

Lemma 4.25. *For all t and all $x < x_H^{(R)p}$, $\varepsilon_t^{(R)p}(x) = High$.*

Proof. By lemma 4.19, $\Delta V_t^p(x)$ is nondecreasing in t . Furthermore, $\Delta d_{t+1} \geq \Delta d_t(x)$. As a result, $\delta\Delta V_t^p(x) + \Delta d_{t-1}(x) \leq \delta\Delta V_{t+1}^p(x) + \Delta d_t(x)$. Since $\delta\Delta V_{t+1}^p(x) + \Delta d_t(x) \leq \theta_{LH}^{(R)p}$ holds for period t , it will hold for period $t-1$ as well, i.e. $\delta\Delta V_t^p(x) + \Delta d_{t-1}(x) \leq \theta_{LH}^{(R)p}$. \square

Lemma 4.26. For all $x > x_L^R$, $\varepsilon_t^R(x) = Low^i$.

Proof. By definition for x_L^R , $\varepsilon_{T-1}^R(x) = Low^i$ for all $x > x_L^R$. The proof for $t < T-1$ follows from lemma 4.21 and the definition for x_L^R . \square

For each state $x < x_L^R$, let us define

$$\bar{\tau}(x) := \{t \in T \mid \delta\Delta V_{t+1}^p(x) + \Delta d_t(x) \geq \theta_{LH}^{(R)p}\}$$

$$\underline{\tau}(x) := \{t \in T \mid \delta\Delta V_{t+1}^\ell(x) - \Delta d_t(x) \geq \theta_{LH}^{(R)\ell}\}$$

if there exists such a t , otherwise $\underline{\tau}(x) = x$ and $\bar{\tau}(x) = T$

Lemma 4.27. For all $x < x_L^R$, there exists a time threshold $\underline{\tau}(x)$, nondecreasing in x , such that $\varepsilon_t^{(R)\ell}(x) = Low^p$ for all $x \leq t \leq \underline{\tau}(x)$ and $\varepsilon_t^{(R)\ell}(x) = High$ for $\underline{\tau}(x) < t < T$.

Proof. This proof uses lemmas 4.20, 4.21, and 4.22. Recall, the highest state where the healthcare leader prefers working mode *High* is $x_{H,t}^{(R)\ell}$ and is nondecreasing in t , i.e. $x_{H,t}^{(R)\ell} \leq x_{H,t+1}^{(R)\ell}$. As a result, if there exists such a time threshold $\underline{\tau}(x)$, it is optimal for the healthcare leader to work in *High* mode for periods $t > \underline{\tau}(x)$. Fix $x > x_{H,t}^{(R)\ell}$, by definition we have $\varepsilon_t^R(x) \neq High$. By lemma 4.21 and nonincreasing differences in $\Delta V_t^\ell(x)$, if $\varepsilon_t^R(x) = Low^p$, then $\varepsilon_{t-1}^R(x) = Low^p$. Hence, for all $t \leq \underline{\tau}(x)$, $\varepsilon_t^R(x) = Low^p$. Furthermore, since $x_{H,t}^{(R)\ell}$ is nondecreasing in t , $\underline{\tau}(x)$ is nondecreasing in x as well. \square

Lemma 4.28. For all $x < x_L^R$, there exists a time threshold $\bar{\tau}(x)$, nonincreasing in x , such that $\varepsilon_t^{(R)p}(x) = High$ for all $x \leq t \leq \bar{\tau}(x)$ and $\varepsilon_t^{(R)p}(x) = Low^\ell$ for $\bar{\tau}(x) < t < T$.

Proof. This proof uses lemmas 4.19, 4.23, and 4.24. By definition, the highest state the healthcare provider prefers working mode *High* is $x_{H,t}^{(R)^p}$. Since $\Delta V_{t+1}^p(x) + \Delta d_t(x)$ has nondecreasing differences in t , if there exists such a time threshold $\bar{\tau}(x)$, it is optimal for the healthcare provider to work in *Low* ^{ℓ} mode for $t > \bar{\tau}(x)$. Furthermore, by lemma 4.23 if $\varepsilon_{t+1}^{(R)^p}(x) = \text{High}$ then $\varepsilon_t^{(R)^p}(x) = \text{High}$. Hence, for all $t \leq \bar{\tau}(x)$, $\varepsilon^{(R)^p} = \text{High}$. Because $x_{H,t}^{(R)^p}$ is nonincreasing in t , i.e. $x_{H,t}^{(R)^p} \geq x_{H,t+1}^{(R)^p}$, $\bar{\tau}(x)$ will be nonincreasing in x . \square

Lemma 4.29. *For all $x < x_L^R$, there exists time periods $\underline{\tau}(x)$ and $\bar{\tau}(x)$ such that $\varepsilon_t^R(x) = \text{Low}^p$ for $t < \underline{\tau}(x)$, $\varepsilon_t^R(x) = \text{High}$ for $\underline{\tau}(x) \leq t \leq \bar{\tau}(x)$, and $\varepsilon_t^R(x) = \text{Low}^\ell$ for $\bar{\tau}(x) < t < T$.*

Proof. This lemma follows from lemmas 4.27 and 4.28. \square

Chapter 5

Conclusions

Chapter 2 studied the effects of current payment mechanisms including fee-for-service, capitation, and performance-based on the objectives of value-based health-care delivery reform. Specifically, conditions under which the providers would be encouraged to simultaneously integrate the care and provide high quality treatments was determined. The results suggest that a performance-based payment scheme can achieve similar results to the first-best solution. The providers will be encouraged to provide high-quality, integrated care if the health benefits of care outweighs the total costs of care. Since the performance-based payment system is based on benefits of the health outcomes, its applicability depends on the measurability and contractability of health outcomes. Contracts like capitation can induce integration or cost reduction incentives; however, can also result in minimum quality provision. Contracts like fee-for-service can support high quality treatments but may not result in lower costs. As a result, fee-for-service is not an efficient way of achieving high quality treatments. The capitation payment system can be modified to include the possibility of complications. When the purchaser holds the providers accountable for the avoidable complications, providers can be

motivated to provide high quality service and integrate the care. When low quality provision results in higher expected costs of complication, capitation with full accountability can outperform other contracts. The second chapter analytically evaluates the consequences of current payment schemes on the healthcare delivery structures and evaluating their merit in fulfilling value-based delivery objectives.

Chapter 3 examined the coordinating contracts between the healthcare purchaser and the IPU in the context of value-based healthcare delivery. Coordinating contracts allow the IPU to optimize its objective while maximizing the societal welfare. The problem of hidden action, in which the IPU's treatment strategy is unverifiable to the purchaser, was considered.

Since the IPU's treatment strategy stochastically affects the health outcomes over time, the optimal contract should consider the health outcomes over the care cycle and evaluate the IPU based on that. To capture the dynamics of health outcomes over the care cycle a continuous-time dynamic principal-agent model was used to derive the optimal contract. The analysis suggests that, in order to prevent healthcare failures, compensation policies should be contingent on the accumulated performance. The concept of *continuation value* was devised to summarize the IPU's track record. Continuation value includes all the information from IPU's performance and can then be used to determine the payments to the IPU. The resulting compensation scheme includes a certain credit limit for each patient plus bonus payments for accumulated good performance. Good performance results in positive money transfers, and failures reduces the compensation.

Results of chapter 3 contributes to the literature of healthcare contracting in several ways. First and foremost, it closes the gap in designing a dynamic incentive contract in the healthcare delivery context. Current payment systems like fee-for-service reimburse the providers for discrete services and can encourage providers

to overprovide services. To add value for the patients, the providers should evaluate if the additional test or treatment can create better health outcomes and can result in better health in future. A payment system like capitation can control the increasing healthcare expenditures, but might incentivize the providers to underprovide services or select the healthy patients that will require less care. Characterization of the optimal contract in this research summarizes the IPU's performance with one single variable, which in turn enables the implementation of the optimal contract. This is a significant contribution because most of the existing payment systems that link the payment to the performance pay for fulfilling some targeted processes. Paying for processes might not result in better health outcomes. The payment scheme introduced in the third chapter pays for the health outcome while holding the IPU accountable for the health outcomes.

Second, this research mathematically demonstrates the optimal payment system in the value-based delivery context. Michael Porter argues that a bundled payment should coordinate the healthcare purchaser-IPU relationship [34]. Bundled payment reimburses providers with a fixed fee for delivering all the services required to deliver a complete cycle of patient care for a specific clinical condition. Nonetheless, the results of this research suggest that other than the bundled payment, the IPU should be compensated with a bonus when they achieve superior performance. Basically, the payment to the IPU should be adjusted based on the health outcomes.

Third, the technique of characterization of the optimal contract in this research can arguably result in a straightforward implementation. The use of a continuous-time principal-agent model helped us to characterize the problem with minimal assumptions and devised us with a great technical tool to summarize the IPU's track during the care cycle. Cost-reducing efforts and value-adding treatments

increase the IPU's *continuation value*, which can eventually result in bonus payments if they exceed a certain threshold for the continuation value. Thus, the proposed payment scheme can fulfill the tenets of value-based healthcare delivery by acting as the source of incentive for the IPU to improve the health outcomes and minimize the costs at the same time.

Chapter 4 studied how different contractual arrangements affect the collaboration dynamics among the IPU members in treating a patient over a finite time horizon. The analysis shows it is optimal for the providers to work in collaboration when patient is not suffering from consecutive complications. The more successful the IPU has been in sustaining the health for a patient, the more likely they will collaborate in the future periods. When successful, IPU members want to keep the failures to a minimum and therefore would provide high quality treatments, whereas when the number of failures increases, the IPU members would prefer low quality treatments sooner. When the efforts are not contractible, the collaboration dynamics depend on the type of contract. Two types of hybrid contracts were studied in this chapter; first, a capitation with a reward-sharing contract and second, a combination of formal and informal contracts. A capitation payment, in hand with a reward-sharing contract, creates similar results to the first-best solution, but, to implement this type of contract the rewards should be contractible. The relational contracts was used to evaluate the collaboration dynamics when rewards are noncontractible. This type of contracts have a formal, court-enforced, component. In the IPU, the provider will be paid a fixed-amount no matter what the health outcomes are. Relational contracts also have a discretionary payment component. A general form of discretionary payment was considered in this research, which is decreasing with the number of failures. The results demonstrate that the provider is encouraged to work in collaboration sooner than the leader.

The provider's incentive stems from the discretionary payment at the end of each period. Nevertheless, the healthcare leader and IPU's performance will be assessed at the end of disease cycle. The results show that in some periods relational contracts can achieve better results than the reward-sharing contracts. This study is the first to characterize the collaboration dynamics among the providers in the value-based healthcare delivery.

Bibliography

- [1] CIHI. National health expenditure trends, 1975 to 2013. Technical Report 978-1-77109-225-8, Canadian Institute for Health Information, 2013.
- [2] Pierre Thomas Léger, Canadian Health Services Research Foundation, et al. *Physician payment mechanisms: an overview of policy options for Canada*. Canadian Health Services Research Foundation, 2011.
- [3] World Health Organization. The world health report 2000, health systems: Improving performance. Technical report, World Health Organization, 2000. URL <http://www.who.int/whr/2000/en/index.html>.
- [4] Michael E. Porter and Elizabeth Olmsted Teisberg. *Redefining Health Care; Creating Value-Based Competition on Results*. Harvard Business School Publishing, 2006.
- [5] Michael E. Porter. What is value in health care? *New England Journal of Medicine*, 363(26):2477–2481, 2010.
- [6] Michael Grossman. The human capital model. In A. J. Culyer and J. P. Newhouse, editors, *Handbook of Health Economics*, volume 1 of *Handbook of Health Economics*, chapter 7, pages 347–408. Elsevier, 2000.
- [7] Ching-To Albert. Ma. Health care payment systems: Cost and quality incentives. *Journal of Economics & Management Strategy*, 3(1):93–112, 1994.

-
- [8] Joseph P. Newhouse. Reimbursing health plans and health providers: Efficiency in production versus selection. *Journal of Economic Literature*, 34(3): 1236–1263, 1996.
- [9] Ching-to Albert Ma and Thomas G McGuire. Optimal Health Insurance and Provider Payment. *American Economic Review*, 87(4):685–704, September 1997.
- [10] T Gosden, F Forland, IS Kristiansen, M Sutton, B Leese, Amf Giuffrida, M Sergison, and L Pedersen. Capitation, salary, fee-for-service and mixed systems of payment: effects on the behaviour of primary care physicians. *Cochrane Database Syst Rev*, 3(3), 2000.
- [11] Pierre Thomas Léger. Quality control mechanisms under capitation payment for medical services. *Canadian Journal of Economics*, 33(2):564–586, 2000. ISSN 1540-5982.
- [12] Meredith B. Rosenthal, Rushika Fernandopulle, Song HyunSook Ryu, and Bruce Landon. Paying for quality: Providers’ incentives for quality improvement. *Health Affairs*, 23(2):127–41, 2004.
- [13] François Sainfort, John Blake, Diwakar Gupta, and Ronald L Rardin. Operations research for health care delivery systems. *WTEC Panel Report*, 2005.
- [14] CIHI. Health care cost drivers: The facts. Technical Report 978-1-55465-985-2, Canadian Institute for Health Information, 2013.
- [15] Health Council of Canada. Progress timeline 2003-2013: Highlights of health care reform. Technical report, Health Council of Canada, 2014.

-
- [16] Committee on Quality of Health Care in America and Institute of Medicine. Crossing the quality chasm: A new health system for the 21st century. Technical Report 9780309072809, The National Academies Press, 2001.
- [17] A. Mark Fendrick and Michael E. Chernew. Value-based insurance design: Aligning incentives to bridge the divide between quality improvement and cost containment. *The American Journal of Managed Care*, 12(Special Issue), 2006.
- [18] Michael Pignone, Dawn Bucholtz, and Russell Harris. Patient preferences for colon cancer screening. *Journal of general internal medicine*, 14(7):432–437, 1999.
- [19] Martin Chalkley and James M. Malcomson. Government purchasing of health services. In A. J. Culyer and J. P. Newhouse, editors, *Handbook of Health Economics*, volume 1 of *Handbook of Health Economics*, chapter 15, pages 847–890. Elsevier, 2000. URL <http://ideas.repec.org/h/eee/heachp/1-15.html>.
- [20] Michael E. Porter and Clemens Guth. *Redefining German Health Care; Moving to a Value-Based System*. Springer-Gabler, 1st edition, 2012.
- [21] Robin Boadway, Maurice Marchand, and Motohiro Sato. An optimal contract approach to hospital financing. *Journal of Health Economics*, 23(1):85–110, 2004.
- [22] William Jack. Purchasing health care services from providers with unknown altruism. *Journal of Health Economics*, 24(1):73–93, 2005.

-
- [23] Sanford J Grossman and Oliver D Hart. An analysis of the principal-agent problem. *Econometrica: Journal of the Econometric Society*, pages 7–45, 1983.
- [24] Andreu Mas-Colell, Michael D. Whinston, and Jerry R. Green. *Microeconomic Theory*. Oxford University Press, 1995.
- [25] Patrick Bolton and Mathias Dewatripont. *Contract Theory*, volume 1. The MIT Press, 1 edition, 2005.
- [26] Martin Chalkley and James M. Malcomson. Contracting for health services when patient demand does not reflect quality. *Journal of Health Economics*, 17(1):1–19, 1998.
- [27] Prashant C. Fuloria and Stefanos A. Zenios. Outcomes-adjusted reimbursement in a health-care delivery system. *Management Science*, 47(6):735–751, 2001.
- [28] Jonathan Levin. Supermodular games. *Lectures Notes, Department of Economics, Stanford University*, 2006.
- [29] Paul Milgrom and John Roberts. The economics of modern manufacturing: Technology, strategy, and organization. *The American Economic Review*, 80(3):511–528, 1990.
- [30] Paul Milgrom and John Roberts. Complementarities and fit strategy, structure, and organizational change in manufacturing. *Journal of Accounting and Economics*, 19(2):179–208, 1995.
- [31] Hideshi Itoh. Job design, delegation and cooperation: A principal-agent analysis. *European Economic Review*, 38(3-4):691–700, 1994.

-
- [32] Bengt Holmstrom and Paul Milgrom. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization*, 7:24–52, 1991.
- [33] Li Zhang. Complementarity, task assignment, and incentives. *Journal of Management Accounting Research*, 15:225+, 2003. 225 Article.
- [34] Michael E. Porter and Thomas H. Lee. The strategy that will fix health care. *Harvard Business Review*, 91(10):50–70, 2013.
- [35] Jean-Paul Chavas and Kwansoo Kim. Economies of diversification: A generalization and decomposition of economies of scope. *International Journal of Production Economics*, 126(2):229 – 235, 2010.
- [36] Diego Prior. Technical efficiency and scope economies in hospitals. *Applied Economics*, 28(10):1295–1301, 1996.
- [37] Bengt Holmstrom. Moral hazard in teams. *The Bell Journal of Economics*, pages 324–340, 1982.
- [38] Jeremiah E. Hurley. *Health Economics*. McGraw-Hill Ryerson, 2010.
- [39] Mark McClellan. Hospital reimbursement incentives: An empirical analysis. *Journal of Economics & Management Strategy*, 6(1):91–128, 1997.
- [40] Laura A. Petersen, LeChauncy D. Woodard, Tracy Urech, Christina Daw, and Supicha Sookanan. Does pay-for-performance improve the quality of health care? *Annals of Internal Medicine*, 145(4):265–272, 2006.
- [41] Bengt Hölmstrom. Moral hazard and observability. *The Bell Journal of Economics*, pages 74–91, 1979.

-
- [42] Paul Milgrom and John Roberts. Rationalizability, learning, and equilibrium in games with strategic complementarities. *Econometrica*, 58(6):1255–1277, 1990.
- [43] Randall P Ellis and Thomas G McGuire. Provider behavior under prospective reimbursement: Cost sharing and supply. *Journal of health economics*, 5(2):129–151, 1986.
- [44] Randall P Ellis and Thomas G McGuire. Optimal payment systems for health services. *Journal of health economics*, 9(4):375–396, 1990.
- [45] Roy Radner. Repeated principal-agent games with discounting. *Econometrica: Journal of the Econometric Society*, pages 1173–1198, 1985.
- [46] Drew Fudenberg, Bengt Holmstrom, and Paul Milgrom. Short-term contracts and long-term agency relationships. *Journal of economic theory*, 51(1):1–31, 1990.
- [47] Yuliy Sannikov. A continuous-time version of the principal-agent problem. *The Review of Economic Studies*, 75(3):957–984, 2008.
- [48] Heinz Schaettler and Jaeyoung Sung. The first-order approach to the continuous-time principal-agent problem with exponential utility. *Journal of Economic Theory*, 61(2):331–371, 1993.
- [49] Peter M DeMarzo and Yuliy Sannikov. Optimal security design and dynamic capital structure in a continuous-time agency model. *The Journal of Finance*, 61(6):2681–2724, 2006.
- [50] Bruno Biais, Thomas Mariotti, Guillaume Plantin, and Jean-Charles Rochet. Dynamic security design: Convergence to continuous time and asset pricing implications. *The Review of Economic Studies*, 74(2):345–390, 2007.

-
- [51] Bruno Biais, Thomas Mariotti, Jean-Charles Rochet, and Stéphane Vilenneuve. Large risks, limited liability, and dynamic moral hazard. *Econometrica*, 78(1):73–118, 2010.
- [52] Peter M DeMarzo, Michael J Fishman, Zhiguo He, and Neng Wang. Dynamic agency and the q theory of investment. *The Journal of Finance*, 67(6):2295–2340, 2012.
- [53] Claude Dellacherie and Paul-André Meyer. *Probabilities and potential*, volume 29. North-Holland Publishing Co., Amsterdam, 1978.
- [54] Pierre Brémaud. *Point processes and Queues, Martingale Dynamics*, volume 30. Springer, 1981.
- [55] Stephen E Spear and Sanjay Srivastava. On repeated moral hazard with discounting. *The Review of Economic Studies*, 54(4):599–617, 1987.
- [56] Jon R Katzenbach. *The wisdom of teams: Creating the high-performance organization*. Harvard Business Press, 1993.
- [57] G.Ross Baker Ivy Oandasan. Analysis of bundled payment. Technical report, Canadian Health Services Research Foundation, 2006.
- [58] Health Council of Canada. Health care renewal in canada: Accelerating change. Technical report, Health Council of Canada, 2005.
- [59] Carol Borrill, Michael West, David Shapiro, and Anne Rees. Team working and effectiveness in health care. *British Journal of Healthcare Management*, 6(8):364–371, 2000.
- [60] Peter S. Hussey, M. Susan Ridgely, and Meredith B. Rosenthal. The prometheus bundled payment experiment: Slow start shows problems in implementing new payment models. *Health Affairs*, 30(11):2116–24, 2011.

-
- [61] RAND Policy. Analysis of bundled payment. Technical report, RAND, 2010.
- [62] Lloyd S Shapley. Stochastic games. *Proceedings of the National Academy of Sciences of the United States of America*, 39(10):1095, 1953.
- [63] Matthew J Sobel et al. Noncooperative stochastic games. *The Annals of Mathematical Statistics*, 42(6):1930–1935, 1971.
- [64] Evan L Porteus and Seungjin Whang. On manufacturing/marketing incentives. *Management Science*, 37(9):1166–1181, 1991.
- [65] Charles J. Corbett and Gregory A. DeCroix. Shared-savings contracts for indirect materials in supply chains: Channel profits and environmental impacts. *Management Science*, 47(7):pp. 881–893, 2001.
- [66] Gerard P. Cachon. Supply chain coordination with contracts, 2003.
- [67] Charles J Corbett, Gregory A DeCroix, and Albert Y Ha. Optimal shared-savings contracts in supply chains: Linear contracts and double moral hazard. *European Journal of Operational Research*, 163(3):653–667, 2005.
- [68] Martin Albrecht. *Supply chain coordination mechanisms*. Springer, 2010.
- [69] James M. Malcomson and Frans Spinnewyn. The multiperiod principal-agent problem. *The Review of Economic Studies*, 55(3):391–407, 1988.
- [70] Alessandro Bonatti and Johannes Horner. Collaborating. *American Economic Review*, 101(2):632–63, 2011.
- [71] Morvarid Rahmani, Guillaume Roels, and Uday S Karmarkar. Contracting and work dynamics in collaborative projects. *Available at SSRN 2285245*, 2013.

-
- [72] Robert Gibbons. Incentives between firms (and within). *Management Science*, 51(1):2–17, 2005.
- [73] J. Levin. Relational incentive contracts. *The American Economic Review*, 93(3):835–857, 2003. doi:10.1257/000282803322157115.
- [74] George Baker, Robert Gibbons, and Kevin J. Murphy. Subjective performance measures in optimal incentive contracts. *The Quarterly Journal of Economics*, 109(4):1125–1156, 1994.
- [75] Son Ku Kim and Susheng Wang. Linear contracts and the double moral-hazard. *Journal of Economic Theory*, 82(2):342–378, 1998. doi:10.1006/jeth.1998.2439.
- [76] Guillaume Roels, Uday S Karmarkar, and Scott Carr. Contracting for collaborative services. *Management Science*, 56(5):849–863, 2010.
- [77] Erica L. Plambeck and Terry A. Taylor. Partnership in a dynamic production system with unobservable actions and noncontractible output. *Management Science*, 52(10):1509–1527, 2006.
- [78] Sugato Bhattacharyya and Francine Lafontaine. Double-Sided Moral Hazard and the Nature of Share Contracts. *RAND Journal of Economics*, 26(4):761–781, Winter 1995.
- [79] Tamer Basar, Geert Jan Olsder, GJ Clsder, T Basar, T Baser, and Geert Jan Olsder. *Dynamic noncooperative game theory*, volume 200. SIAM, 1995.
- [80] Eric Maskin and Jean Tirole. Markov perfect equilibrium: I. observable actions. *Journal of Economic Theory*, 100(2):191–219, 2001.
- [81] Leslie M Marx and Steven A Matthews. Dynamic voluntary contribution to a public project. *The Review of Economic Studies*, 67(2):327–358, 2000.